sensors

MDPI

Article

An Enhanced Joint Hilbert Embedding-Based Metric to Support Mocap Data Classification with Preserved Interpretability

Cristian Kaori Valencia-Marin [1,*], Juan Diego Pulgarin-Giraldo [2], Luisa Fernanda Velasquez-Martinez [3], Andres Marino Alvarez-Meza [3] and German Castellanos-Dominguez [3]

[1] Faculty of Engineering, Universidad Tecnológica de Pereira, Pereira 660003, Colombia
[2] G-Bio Research Group, Automatic and Electronic Department, Universidad Autónoma de Occidente, Cali 760030, Colombia; jdpulgarin@uao.edu.co
[3] Signal Processing and Recognition Group, Universidad Nacional de Colombia sede Manizales, Manizales 170001, Colombia; lfvelasquezm@unal.edu.co (L.F.V.-M.); amalvarezme@unal.edu.co (A.M.A.-M.); cgcastellanosd@unal.edu.co (G.C.-D.)
* Correspondence: ckvalencia@utp.edu.co

Abstract: Motion capture (Mocap) data are widely used as time series to study human movement. Indeed, animation movies, video games, and biomechanical systems for rehabilitation are significant applications related to Mocap data. However, classifying multi-channel time series from Mocap requires coding the intrinsic dependencies (even nonlinear relationships) between human body joints. Furthermore, the same human action may have variations because the individual alters their movement and therefore the inter/intraclass variability. Here, we introduce an enhanced Hilbert embedding-based approach from a cross-covariance operator, termed EHECCO, to map the input Mocap time series to a tensor space built from both 3D skeletal joints and a principal component analysis-based projection. Obtained results demonstrate how EHECCO represents and discriminates joint probability distributions as kernel-based evaluation of input time series within a tensor reproducing kernel Hilbert space (RKHS). Our approach achieves competitive classification results for style/subject and action recognition tasks on well-known publicly available databases. Moreover, EHECCO favors the interpretation of relevant anthropometric variables correlated with players' expertise and acted movement on a Tennis-Mocap database (also publicly available with this work). Thereby, our EHECCO-based framework provides a unified representation (through the tensor RKHS) of the Mocap time series to compute linear correlations between a coded metric from joint distributions and player properties, i.e., age, body measurements, and sport movement (action class).

Keywords: Hilbert embedding; joint distribution; time series; classification; Mocap data

Citation: Valencia-Marin, C.K.; Pulgarin-Giraldo, J.D.; Velasquez-Martinez, L.F.; Alvarez-Meza, A.M.; Castellanos-Dominguez, G. An Enhanced Joint Hilbert Embedding-Based Metric to Support Mocap Data Classification with Preserved Interpretability. *Sensors* **2021**, *21*, 4443. https://doi.org/10.3390/s21134443

Academic Editor: Angelo Maria Sabatini

Received: 4 May 2021
Accepted: 28 May 2021
Published: 29 June 2021

1. Introduction

Time series classification is a real-world problem that frequently deals with vast quantities of numerical measurements acquired at regular time intervals, having applications in fields such as share markets, biomedicine, intelligent sensor networks, and dynamic objects, among others [1–4]. Thus, in the case of moving objects, a contour of a static object can be transformed into a time series representation to favor image-based object recognition tasks [5–7]. Moreover, when classifying time series, one of the essential tasks is recognizing human actions. Most applications focused on the recognition of human activities are based on the construction of 3D skeletons composed of the human body joints extracted from computer vision systems using traditional video cameras (Microsoft Kinect and similar devices) [8]. However, these systems suffer from optical phenomena that affect their precision, such as changes in lighting and occlusions [9]. Then, to improve human pose tracking, there is considerable interest in techniques that avoid using a video camera—for example, WiFi human sensing [10] and radio-frequency identification (RFID)

tags [11]. On the other hand, there are alternative methodologies based on holographic interferometry [12,13] that are remarkably robust to deformations and allow the skeletons of subjects to be adequately represented.

Regarding the motion capture (Mocap)-based human action analysis, different applications involve the classification of Mocap datasets, such as animation movies and video games [14], biomechanics systems for rehabilitation [15], and translation of sign languages [16], among others. However, Mocap data pose some issues for classifying human activities from time series. First, there is a need to code the time series dependencies (relationships between Mocap joints) to highlight discriminative patterns [17]. Second, a performance of a particular activity may have variations, which can be the results of individuals' alteration of expression, posture, motion, and perspective effects [18]. In addition, the same sequence can be executed in different ways (styles) by distinct subjects [19]. Third, the Mocap data trajectories, obtained from 3D skeletal representations, are coded on high-dimensional spaces holding non-stationary dynamics [20].

In the literature, two main approaches are used to deal with time series representation and classification tasks: model-based (MB) and distance-based (DB) methods [2]. MB allows coding the temporal dependencies between time series from a set of parameters associated with a given stochastic or deterministic model. Some of the relevant examples include the hidden Markov models (HMMs) [21], the adaptive filters (AFs) [22,23], the Gaussian processes (GPs) [24,25], and deep networks [26,27]. HMM represents the input data from a sequence of hidden states that encode temporal dependencies among samples; nevertheless, an appropriate choice of the model's topology/architecture is required, e.g., the covariance matrix shape and the number of hidden states [28]. In the case of AFs, they allow recursive learning of the time series, giving prominence to the most relevant data samples [29]. However, the quantization size and the error tolerance must be tuned appropriately, which can be problematic for 3D skeletal-based samples [30]. Regarding the GP-based methods, a Bayesian representation of time series is carried out. Although GPs are considered nonparametric models, their training is often computationally expensive when calculating the posterior distribution [25]. Recently, deep learning methods have been used for Mocap data classification [26,27,31]. Even though the classification performance is reasonable, exhaustive training is required, the overfitting issue arises for small databases, and the provided algorithms often lack straightforward interpretability [32].

Now, DB approaches reside in the construction of a dissimilarity space from the input time series, which are later used to train a classifier, e.g., a K-nearest neighbors [33,34]. In general, the Euclidean distance (ED) is the most straightforward DB approach. Nonetheless, ED can only be applied to discriminate time series of the same length [35]. Therefore, the dynamic time warping (DTW) dissimilarity appears as an extension of the ED, also known as 2-norm-based distance, to compare series of different lengths [36]. The DTW is quite well known for discriminating time series as it can be seen as a generalization of the ED exclusively for this kind of data [37]. Nevertheless, DTW requires crucial hyperparameter (warping percentage) tuning, and l2-based approaches tend to fail when coding nonlinear patterns [36]. In turn, reproducing kernel Hilbert space (RKHS)-based approaches have been proposed to highlight nonlinear data relationships [38]. Furthermore, Hilbert embedding-based dissimilarities have been introduced in the literature as a generalization of traditional kernel methods, mapping the input data probability distribution as a vector/operator in RKHS. The latter favors the estimation of dissimilarity-based measures within high dimensional spaces [39]. Of note, the Lie group representation approach is commonly applied on skeletal action recognition tasks [40–42]. However, Lie group-based methods suffer from temporal misalignment, which tends to deteriorate the classification accuracy [31]. To solve this problem, the DTW is coupled with the Lie group; nonetheless, the computational time is increased, and a two-step algorithm typically performs worse than an end-to-end learning strategy [31].

In this paper, an enhanced Hilbert embedding-based framework is proposed as a DB approach to support Mocap data classification. In this sense, a novel metric is introduced

to map joint probability distributions, from two different input spaces, in a tensor RKHS through the cross-covariance operator [43,44]. Our approach, termed enhanced Hilbert embedding from cross-covariance operator (EHECCO), allows comparing input data from sample-based kernel evaluations, circumventing the direct estimation of probability functions. The latter helps in the analysis of multi-view instances in pattern recognition tasks, i.e., classification from data fusion [45]. Then, we aim to code temporal information from sequentiality data to support further classification stages regarding human action recognition (HAR). The most significant contributions in this work can be summarized as follows: (i) a novel analytical expression for calculating an RKHS-based dissimilarity to discriminate between joint probability distributions; (ii) a representation strategy for the extraction and processing of skeletons from Mocap videos, which allows finding the most relevant and discriminating movement patterns; and (iii) a recognition framework of human activities and style based on EHECCO, which allows anthropometric analysis and proper interpretation of the results obtained. Indeed, our EHECCO-based framework for HAR facilitates the computation of linear correlations between the coded metric, player properties (age, body measurements, among others), and human action classes. Of note, EHECCO can deal with different time series lengths, preserving the most relevant frames (human poses) when comparing the Mocap time series. Our method is a crucial improvement compared with conventional human movement analysis approaches, which employ alienation angles, linear velocities, and angular velocities as factors to be evaluated [46]. The approach is tested on both public (for action and style recognition) and our own (for action recognition and anthropometric analysis) Mocap datasets. Results obtained are competitive in terms of the achieved classification accuracy with the benefit of Mocap data interpretability.

The remainder of this paper is organized as follows: Section 2 describes the mathematical background. Section 3 shows the experimental set-up. Section 4 presents the results and discussion. Finally, the conclusions appear in Section 5.

2. Methods

In this section, we provide the mathematical background concerning our Hilbert embedding-based metric. First, the well-known marginal embedding approach is briefly described. Then, we present our joint embedding proposal to build a metric in a tensor RKHS from joint distributions. Our approach seeks to exploit two main issues: (i) joint distribution-based modeling from two different input spaces, and (ii) non-linear sample mapping to code relevant data dependencies from joint distributions circumventing the direct estimation of probability functions. The latter would be helpful to deal with multi-channel time series, which is the basis of our experimental set-up concerning HAR from Mocap data.

2.1. Marginal Embedding-Based Metric in RKHS

Let $\mathcal{P}_\mathcal{X}$ be the space of all marginal probability distributions on $\mathcal{X}$. Moreover, let X be a random variable with distribution $\mathbb{P}_X \in \mathcal{P}_\mathcal{X}$. A marginal embedding $\mu^X_\mathcal{H} \in \mathcal{H}$ can be defined as [47]:

$$\mu^X_\mathcal{H} = \mathbb{E}_x[\varphi(x)] = \int_X \varphi(x) d\mathbb{P}_X, \tag{1}$$

where $x \in X$ is a given sample and $\mathcal{H}$ is a reproducing kernel Hilbert space (RKHS) holding the nonlinear mapping $\varphi : \mathcal{X} \to \mathcal{H}$. $\mathbb{E}[\cdot]$ stands for the expectation operator. Furthermore, let Z be another random variable with distribution $\mathbb{P}_Z \in \mathcal{P}_\mathcal{X}$ and marginal embedding $\mu^Z_\mathcal{H} \in \mathcal{H}$. Then, a distance metric $\mathrm{d} : \mathcal{P}_\mathcal{X} \times \mathcal{P}_\mathcal{X} \to \mathbb{R}^+$ between probability distributions can be defined in $\mathcal{H}$ from the marginal embeddings $\mu^X_\mathcal{H}$ and $\mu^Z_\mathcal{H}$ as:

$$\mathrm{d}^2(\mathbb{P}_X, \mathbb{P}_Z) = \left\| \mu^X_\mathcal{H} - \mu^Z_\mathcal{H} \right\|^2_\mathcal{H}, \tag{2}$$

where $\|\cdot\|_{\mathcal{H}}$ stands for the norm operator in $\mathcal{H}$. Founded on the kernel trick property $\kappa_\varphi(x,x') = \langle \varphi(x), \varphi(x') \rangle_{\mathcal{H}}$, being $\kappa_\phi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ a positive semi-definite characteristic kernel function [39], the metric in Equation (2) can be rewritten as [48]:

$$d^2(\mathbb{P}_X, \mathbb{P}_Z) = \mathbb{E}_{x,x'}[\kappa_\varphi(x,x')] + \mathbb{E}_{z,z'}[\kappa_\varphi(z,z')] - 2\mathbb{E}_{x,z}[\kappa_\varphi(x,z)], \tag{3}$$

with $x, x' \in X$ and $z, z' \in Z$.

The expression in Equation (3) is an analytical metric function in RKHS for probability distributions [49]. In fact, the well-known maximum mean discrepancy (MMD) distance arises from Equation (3) to extend traditional kernel methods for estimating probability functions [22,50]. Namely, let $\{x_n \in \mathcal{X}\}_{n=1}^N$ and $\{y_m \in \mathcal{X}\}_{m=1}^M$ be a pair of sets holding N and M samples, respectively. Moreover, let us assume that the probability distributions $\mathbb{P}_X$ and $\mathbb{P}_Z$ admit density functions $p(x)$ and $p(y)$. Then, after fixing the empiric-based estimators $\hat{p}(x) = \frac{1}{N}\sum_{n=1}^N \delta(x - x_n)$ and $\hat{p}(y) = \frac{1}{M}\sum_{m=1}^M \delta(y - y_m)$, $\delta(\cdot) \in \{0,1\}$ stands for the delta function, and using a Gaussian characteristic kernel $\kappa_\sigma(x_n, y_m) = \exp(-\|x_n - y_m\|_2^2/2\sigma^2)$, $\sigma \in \mathbb{R}^+$ is a similarity bandwidth, the MMD estimator is given by [51]:

$$\hat{d}^2_{\text{MMD}}(\mathbb{P}_X, \mathbb{P}_Z) = \frac{1}{N^2}\mathbf{1}_N^\top \mathbf{K}^{x,x}\mathbf{1}_N + \frac{1}{M^2}\mathbf{1}_M^\top \mathbf{K}^{y,y}\mathbf{1}_M - \frac{2}{NM}\mathbf{1}_N^\top \mathbf{K}^{x,y}\mathbf{1}_M, \tag{4}$$

where $\mathbf{K}^{x,x} \in \mathbb{R}^{N\times N}$, $\mathbf{K}^{y,y} \in \mathbb{R}^{M\times M}$, and $\mathbf{K}^{x,y} \in \mathbb{R}^{N\times M}$ are kernel matrices computed from $\kappa_{\sqrt{2}\sigma}(\cdot,\cdot)$. $\mathbf{1}_N$ and $\mathbf{1}_M$ are all one column vectors of size N and M, respectively.

2.2. Enhanced Hilbert Embedding from Cross-Covariance Operator (EHECCO)

Though MMD in Equation (4) allows comparing samples without any assumption over probability distributions, it only codes the marginal information when performing the distance-based representation. Therefore, dealing with complex data relationships—for example, Mocap time series classification for HAR—will benefit from representing the instances on different RHKS to code contrasting properties of the samples. Then, a joint distribution-based metric can be developed.

Let us consider another pair of random variables $Y, L \in \mathcal{Y}$ with distributions $\mathbb{P}_Y, \mathbb{P}_L \in \mathcal{P}_\mathcal{Y}$, where $\mathcal{P}_\mathcal{Y}$ is the space of all marginal distributions on $\mathcal{Y}$; further, let $y \in Y$ and $l \in L$ be samples from the aforementioned random variables. Our enhanced Hilbert embedding from cross-covariance operator (EHECCO) allows computing a metric between the joint distributions $\mathbb{P}_{X,Y}, \mathbb{P}_{Z,L} \in \mathcal{P}_{\mathcal{X},\mathcal{Y}}$, where $\mathcal{P}_{\mathcal{X},\mathcal{Y}}$ is the space of all joint probability distributions defined on the Cartesian product $\mathcal{X} \times \mathcal{Y}$. Following the metric in Equation (2), the RKHS-based distance $\mathrm{d_J} : (\mathcal{P}_{\mathcal{X},\mathcal{Y}} \times \mathcal{P}_{\mathcal{X},\mathcal{Y}}) \times (\mathcal{P}_{\mathcal{X},\mathcal{Y}} \times \mathcal{P}_{\mathcal{X},\mathcal{Y}}) \to \mathbb{R}^+$ between joint probability distributions yields:

$$\mathrm{d_J}^2(\mathbb{P}_{X,Y}, \mathbb{P}_{Z,L}) = \left\| \mu^{X,Y}_{\mathcal{H}\otimes\mathcal{G}} - \mu^{Z,L}_{\mathcal{H}\otimes\mathcal{G}} \right\|^2_{\mathcal{H}\otimes\mathcal{G}}, \tag{5}$$

where the Hilbert embeddings $\mu^{X,Y}_{\mathcal{H}\otimes\mathcal{G}}, \mu^{Z,L}_{\mathcal{H}\otimes\mathcal{G}} \in \mathcal{H}\otimes\mathcal{G}$, being $\mathcal{H}\otimes\mathcal{G}$ a tensor space, can be defined as the following cross-covariance operators [48]:

$$\mu^{X,Y}_{\mathcal{H}\otimes\mathcal{G}} = \mathbb{E}_{X,Y}[\varphi(x)\otimes\phi(y)], \tag{6}$$

$$\mu^{Z,L}_{\mathcal{H}\otimes\mathcal{G}} = \mathbb{E}_{Z,L}[\varphi(z)\otimes\phi(l)], \tag{7}$$

where $\varphi(x), \varphi(z) \in \mathcal{H}$, $\phi(y), \phi(l) \in \mathcal{G}$ are nonlinear mappings to the RKHS $\mathcal{H}$ and $\mathcal{G}$, following the positive semi-definite characteristic kernels: $\kappa_\varphi(x,x') = \langle \varphi(x), \varphi(x') \rangle_{\mathcal{H}}$, $\forall x, x' \in X$ and $\kappa_\phi(y,y') = \langle \phi(y), \phi(y') \rangle_{\mathcal{G}}, \forall y, y' \in Y$ [49]. The latter is accomplished too for samples of the random variables Z and L, respectively.

Furthermore, let us assume that $\mathbb{P}_{X,Y}$ and $\mathbb{P}_{Z,L}$ admit density functions $p(x,y)$ and $p(z,l)$, respectively; then, $d\mathbb{P}_{X,Y} = p(x,y)dxy$ and $d\mathbb{P}_{Z,L} = p(z,l)dzl$. We can rewrite Equation (5) as follows [52]:

$$\begin{aligned} \mathrm{d_J}^2(\mathbb{P}_{X,Y},\mathbb{P}_{Z,L}) = &\int_{\mathcal{X}\times\mathcal{Y}}\int_{\mathcal{X}\times\mathcal{Y}} \kappa_\varphi(x,x')\kappa_\phi(y,y')p(x,y)p(x',y')dxydx'y' \\ &+ \int_{\mathcal{X}\times\mathcal{Y}}\int_{\mathcal{X}\times\mathcal{Y}} \kappa_\varphi(z,z')\kappa_\phi(l,l')p(z,l)p(z',l')dzldz'l' \\ &- 2\int_{\mathcal{X}\times\mathcal{Y}}\int_{\mathcal{X}\times\mathcal{Y}} \kappa_\varphi(x,z)\kappa_\phi(y,l)p(x,y)p(z,l)dxydzl. \end{aligned} \tag{8}$$

Of note, the metric presented in Equations (5) and (8) (see Figure 1 for a schematic illustration) favors the extraction of relevant patterns from joint distributions as vector-based mappings in RKHS. Indeed, Hilbert embedding-based feature representations allow mapping marginal, conditional, and joint distributions into feature spaces using kernels, comparing and manipulating these distributions via feature space operations [44]. Our proposal is a direct extension of the conventional marginal embedding approach presented in Equation (2) towards a metric between joint distribution (see Theorem 1 in [48]). Moreover, it is well known in the machine learning literature that kernel-based methods favor highlighting nonlinear dependencies from input samples by mapping them to high-dimensional, possibly infinite, Hilbert space, revealing discriminative data patterns [53].

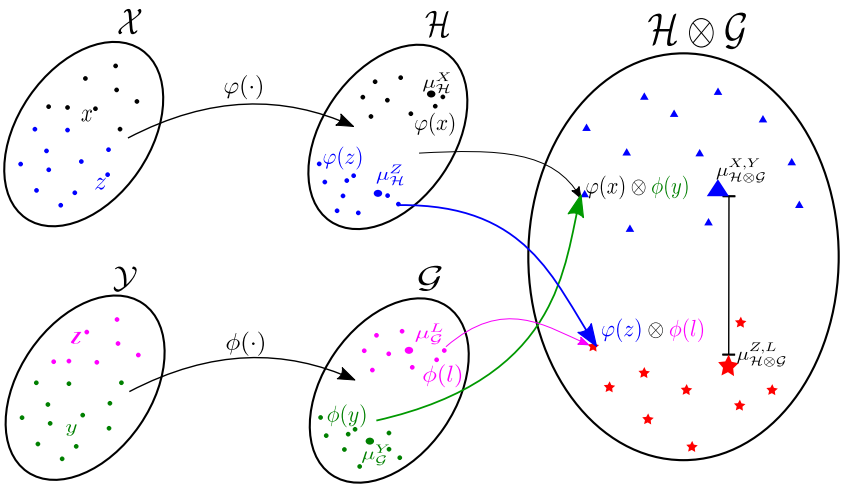

Figure 1. Schematic illustration of our EHECCO-based metric. Input spaces $\mathcal{X}$ and $\mathcal{Y}$ are mapped to RKHSs $\mathcal{H}$ and $\mathcal{G}$, respectively. Then, the tensor space $\mathcal{H}\otimes\mathcal{G}$ is built using a cross-covariance operator strategy.

For concrete testing, let $\{\mathbf{x}_n \in \mathbb{R}^V, \mathbf{y}_n \in \mathbb{R}^Q\}_{n=1}^N$ and $\{\mathbf{z}_m \in \mathbb{R}^V, \boldsymbol{l}_m \in \mathbb{R}^Q\}_{m=1}^M$ be a pair of input sets (time series coded into two different spaces), and our matrix-based estimator in Equation (8) yields:

$$\hat{\mathrm{d}}^2_{\mathrm{J}}(\mathbb{P}_{X,Y},\mathbb{P}_{Z,L}) = \boldsymbol{\alpha}_{x,y}^\top\left(\mathbf{K}_\varphi^{x,x}\circ\mathbf{K}_\phi^{y,y}\right)\boldsymbol{\alpha}_{x,y} + \boldsymbol{\alpha}_{z,l}^\top\left(\mathbf{K}_\varphi^{z,z}\circ\mathbf{K}_\phi^{l,l}\right)\boldsymbol{\alpha}_{z,l} - 2\boldsymbol{\alpha}_{x,y}^\top\left(\mathbf{K}_\varphi^{x,z}\circ\mathbf{K}_\phi^{y,l}\right)\boldsymbol{\alpha}_{z,l}, \tag{9}$$

where the kernel matrices $\mathbf{K}_\varphi^{x,x}, \mathbf{K}_\phi^{y,y} \in^{N\times N}$, $\mathbf{K}_\varphi^{z,z}, \mathbf{K}_\phi^{l,l} \in^{M\times M}$, and $\mathbf{K}_\phi^{x,z}, \mathbf{K}_\varphi^{y,l} \in^{N\times M}$ are computed based on the kernel functions $\kappa_\varphi(\cdot,\cdot)$ and $\kappa_\phi(\cdot,\cdot)$. The operator $\circ$ stands for the Hadamard product. Moreover, the probability column vectors $\boldsymbol{\alpha}_{x,y} \in [0,1]^N$ and $\boldsymbol{\alpha}_{z,l} \in [0,1]^M$ hold the joint probability estimators $\hat{p}(\mathbf{x}_n,\mathbf{y}_n)$ and $\hat{p}(\mathbf{z}_m,\boldsymbol{l}_m)$, respectively.

It is worth mentioning that our EHECCO estimator in Equation (9) provides a data-driven metric in the tensor space $\mathcal{H}\otimes\mathcal{G}$ to compare the joint distributions $\mathbb{P}_{X,Y}$ and $\mathbb{P}_{Z,L}$ as kernel-based operations of input vectors. Remarkably, it can benefit further classification

stages by extracting discriminative features from high-dimensional feature spaces through our kernel-based approach.

In short, our EHECCO-based metric seeks to exploit two main issues: (i) joint distribution-based time series modeling from two different input spaces, and (ii) nonlinear data mapping to code relevant sample dependencies from joint distributions, circumventing the direct estimation of probability functions. Regarding the classification of multi-channel time series, i.e., HAR based on Mocap records, spatio-temporal relationships can be highlighted from the joint space (tensor RKHS), favoring data discrimination. Moreover, as our EHECCO-based metric can deal with different time series lengths, the most relevant frames (human poses) can be preserved when comparing time series. The latter is a crucial improvement compared with conventional human movement analysis approaches, which employ alienation angles, linear velocities, and angular velocities as factors to be evaluated [46].

3. Experimental Setup

Our EHECCO metric in Equation (9) is used to construct a HAR framework from Mocap videos. Thereby, we aim to demonstrate the discriminative capability and interpretability benefits of our joint distribution-based embedding approach to deal with multi-channel time series related to human movement. Then, the experimental design of our EHECCO-based framework can be summarized in the following stages:

- 3D joint normalization. A 3D joint representation is extracted from each Mocap record followed by a hip-based normalization [27].
- Codebook generation. A codebook of Mocap frames is built to gather the most representative movement poses. Then, a set of Nc clusters is computed using the well-known spectral clustering algorithm [54], from a vector-based concatenation of the 3D joints. The radial basis function is used as similarity, fixing the bandwidth as the median of the input Euclidean distances.
- Joint and latent space-based representations. To code relevant patterns from provided codebooks, both the input joints and their latent space are considered to build a Mocap video input set: $\{\mathbf{x}_n \in \mathbb{R}^V, \mathbf{y}_n \in \mathbb{R}^Q\}_{n=1}^{N_c}$. Here, the well-known principal component analysis (PCA) algorithm is employed to compute a latent space coding the most relevant orthonormal basis concerning the preserved input channels' variability [55]. In fact, for concrete testing, three principal components are considered ($Q = 3$). According to our experiments, three components preserve at least 75% of the input data variability. Note that the V value equals the number of Mocap joints times three (3D skeleton).
- EHECCO-based dissimilarity representation and classification. Given a a pair of Mocap video sets: $\{\mathbf{x}_n \in \mathbb{R}^V, \mathbf{y}_n \in \mathbb{R}^Q\}_{n=1}^{N_c}$, $\{\mathbf{z}_m \in \mathbb{R}^V, \boldsymbol{l}_m \in \mathbb{R}^Q\}_{m=1}^{N_c}$, our EHECCO-based distance measure in Equation (9) is computed. In turn, a dissimilarity matrix $\mathbf{D} \in \mathbb{R}^{\Lambda \times \Lambda}$ is calculated as EHECCO-based pairwise Mocap video comparisons (Λ stands for the number of processed Mocap videos). For the tested databases, the probability vectors are fixed as $\boldsymbol{\alpha}_{x,y}, \boldsymbol{\alpha}_{z,l} \sim U[0, N_c]$, being $U[0, N_c]$ the uniform distribution. Since the Gaussian kernel is preferred in pattern classification because of its universal approximating ability and mathematical tractability [56], $\kappa_\varphi(\cdot,\cdot)$ and $\kappa_\phi(\cdot,\cdot)$ are fixed as Gaussians. Each kernel bandwidth is searched within the range $\{0.5\sigma_0, \sigma_0, 2\sigma_0, 5\sigma_0, 10\sigma_0\}$ concerning the final classification performance. $\sigma_0 \in \mathbb{R}^+$ equals the median of input Euclidean distances in accordance with each studied space $\mathcal{X}$ (input Mocap joints) or $\mathcal{Y}$ (PCA-based latent projection). Finally, a support vector machine (SVM) classifier is trained on the EHECCO's distance matrix. A radial basis function (nonlinear mapping) is set for the SVM, and the penalty and precision hyper-parameters are settled from the grids $\{1, 10, 100, 1000, 10,000\}$ and $\{0.01, 0.1, 1, 100, 1000\}$, respectively, concerning the classification performance. In addition, 2D data projection is also provided from the EHECOO metric for visualization purposes.

Figure 2 also summarizes the provided EHECCO-based flowchart for Mocap data classification.

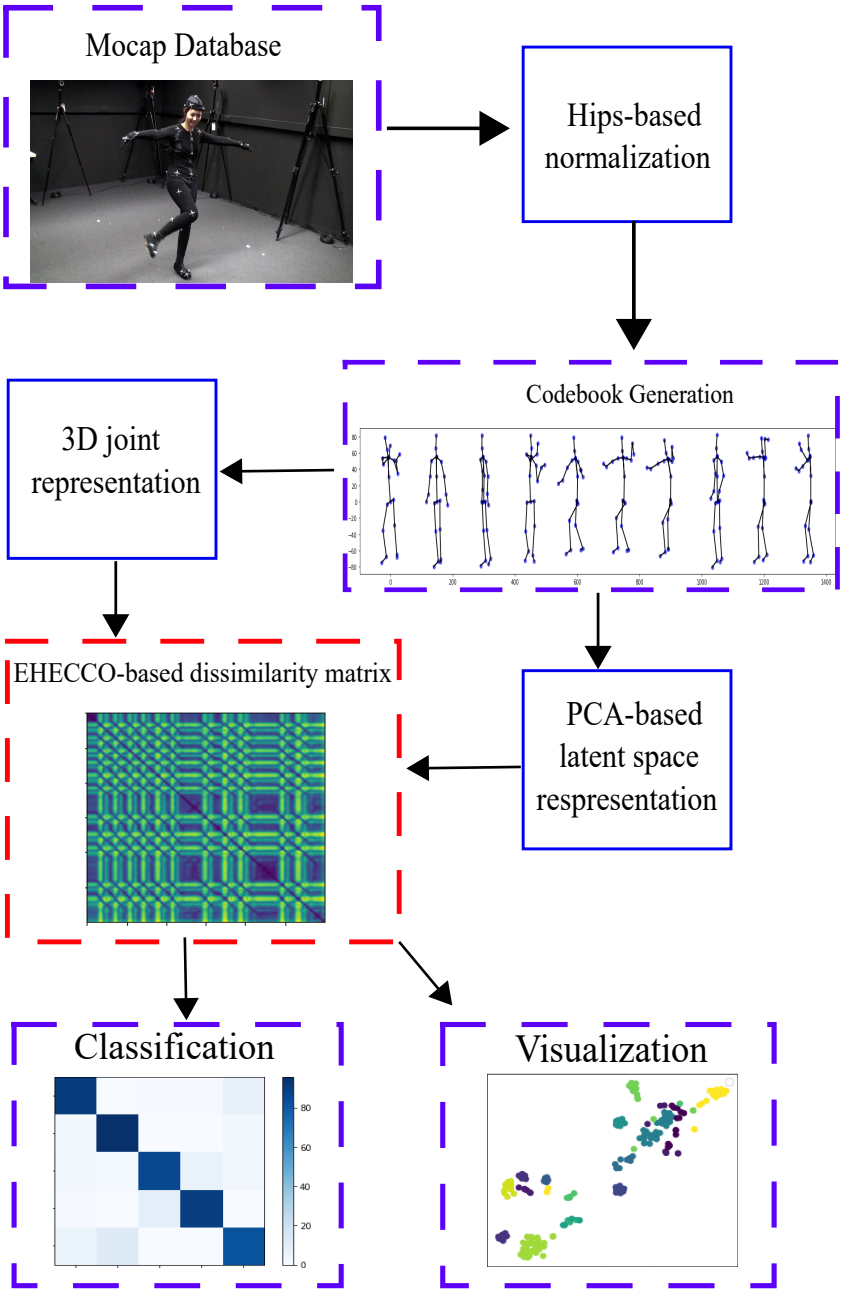

Figure 2. EHECCO-based Mocap data classification framework. Hip joint normalization and spectral clustering-based codebook generation are carried out to extract relevant skeletal poses. Then, 3D joint representation ($\mathcal{X}$) and PCA-based latent projection ($\mathcal{Y}$) are used to support the EHECCO metric from joint probability. Lastly, an SVM classifier is trained from the EHECCO distance that also supports 2D data visualization.

3.1. Mocap Databases

For concrete testing, the following databases are tested for human action classification and analysis from the Mocap data:

- HDM05 for style/subject recognition (http://resources.mpi-inf.mpg.de/HDM05/, accessed on 5 October 2020). This database includes 325 records (from 65 actions) performed by five different subjects. The dataset includes several recorded actions using a Vicon mocap system, where 31 reflective markers are placed on the subject's

bodies [57]. Then, multi-channel time series of BVH files at 120 frames per second is provided. Following the framework proposed by the authors in [27], we built a scheme for style classification (subject recognition). We relate the classes to each of the five subjects who perform the actions as follows: subject 1 (s1) and similarly for the other subjects.

- CMU subset for action recognition (http://mocap.cs.cmu.edu/info.php, accessed on 5 October 2020). Mocap data are obtained from the Carnegie Mellon Graphics Laboratory, holding 12 Vicon infrared MX-40 cameras at 120 Hz with images of four-megapixel resolution. The cameras are placed around a rectangular area, of approximately 3 m $\times$ 8 m, in the center of the room. In particular, multi-channel time series as BVH files with 38 markers are provided. In the same way, as in [26], an action recognition task is carried out from a subset of 150 clips of 15 different motion classes (performed by several subjects): *walking (wal)*, *running (run)*, *sitting (sit)*, *jumping (jum)*, *weight-carrying (wei)*, *climbing (cli)*, *swinging (swn)*, *placing a ball (plb)*, *placing tee (plt)*, *kicking (kic)*, *soccer and basketball playing (soc)*, *boxing (box)*, *swimming (swm)*, *salsa (sal)*, and *Indian Bollywood dancing (InB)*.
- Tennis-Mocap for action recognition and anthropometric analysis (https://drive.google.com/file/d/1-3HAUP4vIBBMz21f7RRgA4b89uNrLxvr/view?usp=sharing, accessed on 5 October 2020). The data are collected from 17 players of the Caldas-Colombia tennis league. The employed motion capture protocol includes the placement of 34 markers for collecting information on body joints. Optitrack Flex V100 (100 Hz) infrared videography is collected from six cameras to acquire sagittal, frontal, and lateral planes. All subjects are encouraged to hit the ball with the same velocity and action as in a tennis match. Moreover, the players are instructed to hit one series continuously by 30 s of each indicated stroke: *serve (Ser)*, *forehand (For)*, *backhand (Bac)*, *volley (Vol)*, *backhand volley (BaV)*, and *smash (Sma)*. In addition, the Tennis database includes the anthropomorphic players' measurements depicted in Table 1.

Table 1. Tennis dataset's anthropomorphic measurements. The color represents the measurement group: age (brown), weight (light green), length (red), perimeters (blue), fat fold (pink), and tennis move (black).

Age	Thigh cm (THI)	Height cm (HEI)	Medial calf mm (CAL)
Mass	Calf maximum cm (CALM)	Foot length cm (LFE)	Biceps mm (BIC)
Cephalic cm (CEP)	Relaxed arm cm (ARMR)	Biliocrestal cm (BIL)	Front thigh mm (THIF)
Minimum ankle cm (ANK)	Mesosternal chest cm (MEC)	Humerus cm (HUM)	Forehand (FORE)
Hip max cm (HIP)	Forearm cm (FOR)	Supraspinal mm (SUP)	Smash (SMA)
Contracted arm 90 cm (ARMC)	Bistyloid cm (BIS)	Subscapular mm (SUB)	Backhand (BAC)
Waist cm (WAI)	Biacromial cm (BIA)	Iliac crest mm (ILI)	Serve (SER)
Middle thigh cm (THIM)	Femur knee cm (FEK)	Triceps mm (TRI)	Volley (VOL)
Wrist cm (WRI)	Wingspan cm (WIN)	Abdominal mm (ABD)	Backhand Volley (BAV)

3.2. Method Comparison, Quality Assessment, and Implementation Details

To evaluate the performance of our EHECCO-based framework to classify Mocap data, we compare the results on the public databases (HDM05 and CMU subset) obtained in HAR with relevant state-of-the-art approaches:

Method comparison for HDM05 dataset (style/subject recognition). We compare our own method with the following methods: symmetric positive definite network (SPDNet) [40],

special Euclidean group (SE) [41], special orthogonal group (SO) [42], Lie groups on deep neural networks (LieNet) [31], and works based on 3D sequence to RGB image transformation (Seq2im) [27].

Method comparison for CMU subset (action recognition). We compare our results with the following approaches: motion template combined with a DTW-based classifier (MT+DTW) [58], self-similarity matrix with DTW distance (SSM+DTW) [18], efficient motion retrieval (EMR) [59], and motion words with convolutional neural networks (MW+CNN) [26].

Afterward, regarding the Tennis-Mocap database (own database), we carried out action recognition tasks along with anthropometric analysis using the extracted EHECCO-based patterns together with the measurements presented in Table 1.

As a quality assessment, we use a 10-fold cross-validation strategy based on the well-known average accuracy and confusion matrix performance measures [54]. As an illustrative example, the accuracy for a binary classification case is defined as $A_{cc} = (T_p + T_n)/N$, where T_p and T_n are the true positive and true negative classifier's predictions, respectively, being N the number of studied samples. Similarly, the confusion matrix for a binary classification task includes an array holding the values of T_p and T_n in the main diagonal and the false positive (F_p) and false negative (F_n) predictions on the upper and lower triangular matrix positions.

All our experiments are implemented in Python using the sklearn toolbox for the training and validation of the models and the PyMO library (https://github.com/omimo/PyMO, accessed on 5 October 2020) for the management and representation of Mocap data. The most relevant codes of this paper can be found in a publicly available repository (https://github.com/Ckvalencia/hello-world/blob/master/SHECCO_CMU_sub.ipynb, accessed on 12 April 2021).

4. Results and Discussion

This section describes the classification results obtained by EHECCO-based distance for the Mocap datasets specified in Section 3.1.

4.1. HDM05 and CMU Results: Mocap Classification Benchmark

Figure 3 presents an example of relevant skeletons (codebook generation) for a given Mocap video selected from HDM05 and CMU datasets. For illustration purposes, two classes are investigated: throwing high with the right hand while standing and boxing, for which the 2D PCA projection is dotted with colored points, while the recorded frames are pictured with black points. Note that the frames chosen by the clustering algorithm are distributed so that they cover the entire space. As seen, the algorithm manages to capture the most relevant information about the movement without significant loss of information. Furthermore, the *boxing* record results show how both the codebook generation and the PCA-based projection preserve the cyclic action behavior, e.g., the subject acting several times.

For each database, Figure 4 presents the confusion matrix along with the 2D low-dimensional scatter plot performed by the EHECCO distance matrix $\mathbf{D}$ using the t-distributed stochastic neighbor embedding (t-SNE) algorithm [60]. The scatter plot visually interprets the EHECCO patterns, preserving the spatial relationships in the higher tensor space (nearest-neighbors) [61]. As a result, our EHECCO approach achieves a competitive discrimination performance concerning both subject/style and action recognition tasks, reaching an average accuracy of 88.8 and 90 percentage in HDM05 and CMU subsets, respectively. The scatters also evidence the EHECCO's ability to reveal both local and global data patterns. Of note, some classes hold nonstationary behavior, due to groups overlapping, i.e., see the confusion matrices and the 2D projections for subject two vs. subject five in HDM05: *sal* vs. *cli*, *soc* vs. *sit*, and *plb* vs. *kic* actions for the CMU subset. The behavior of this latter paired comparison is expected because of the Mocap data variations [19]. Overall, the combination of EHECCO with SVM can deal with the intra/interclass variability.

One more aspect to highlight is comparing the performance EHECCO classification performance with several state-of-the-art results recently reported. Thus, Table 2 shows the accuracy results for the *HDM05* dataset, including the following methods: symmetric positive definite network (SPDNet) [40], special Euclidean group (SE) [41], special orthogonal group (SO) [42], Lie groups on deep neural networks (LieNet) [31], and sequence to RGB image (Seq2im) [27]. The latter employs 3D sequence to RGB image transformation combined with conventional classifiers such as SVM, K-nearest neighbors (KNN), random forest (RF), and convolutional neural networks (CNN). As seen, the EHECCO+SVM combination overcomes the state-of-the-art techniques compared, including those based on deep learning such as Seq2im+CNN. Nevertheless, deep learning approaches often require exhaustive fine-tuning, whereas our EHECCO-based metric provides a data-driven technique as input vector evaluations for nonlinear pattern extraction in RHKS.

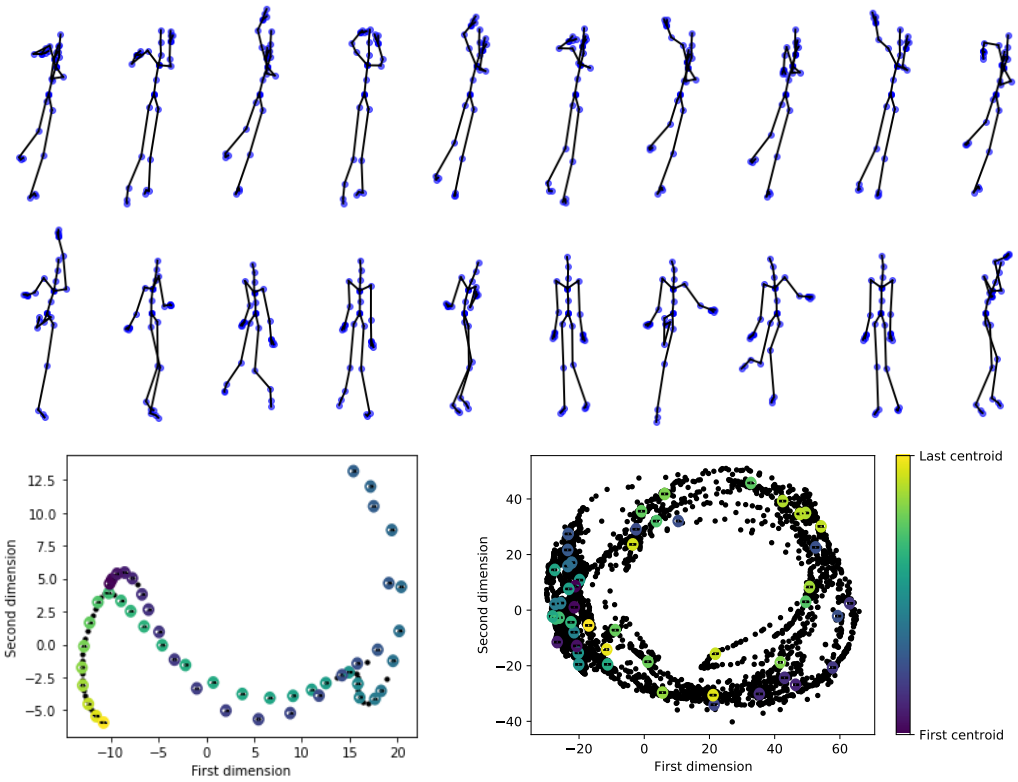

Figure 3. Illustrative results for codebook generation and latent space-based representation (HDM05 and CMU subset datasets). Top: Codebook generation for a Mocap video of the *throwing high with the right hand while standing* class (HDM05). Middle: Codebook generation for a Mocap record of *boxing* class (CMU subset). Bottom left: PCA-based latent space for HDM05 video. Bottom right: PCA-based latent space for CMU subset video. The first two components are shown for visualization purposes. Black markers represent the original input Mocap frames (time series). Color markers represent the chosen frames (codebook).

Furthermore, Table 3 presents the comparison results for the CMU subset, which includes the motion template (MT), self-similarity matrix (SSM), and efficient motion retrieval (EMR) methods [18,58,59], relying on dissimilarity matrices obtained from Mocap data feature extraction techniques and the DTW distance. Although they managed to obtain promising results, their achieved performance is not competitive enough concerning more recent methods. Motion word-(MW)-based methodology [26] yields competitive accuracy. In fact, MW incorporates a deep learning scheme to favor the time series representation. Our EHECOO outperforms most of the compared works, and it is rather similar regarding the achieved accuracy compared to the work proposed in [26]. Hence, EHECOO allows encoding nonlinear Mocap data similarities from both the 3D skeleton and PCA-based latent

space through a joint distribution comparison perspective. Thereby, the EHECCO+SVM pipeline supports both the style and action recognition performance with the benefit of providing the metric interpretability of the extracted representation.

Table 2. Comparing results of Mocap-based style/subject recognition (HDM05 dataset). The average accuracy is reported concerning the cited works vs. our approach—EHECCO+SVM.

Method	Accuracy (%)
SPDNet [40]	61.45
SE [41]	70.26
SO [42]	71.31
LieNet [31]	75.78
Seq2Im+SVM [27]	70.70
Seq2Im+KNN [27]	66.82
Seq2IM+RF [27]	80.62
Seq2Im+CNN (fine-tuning)[27]	83.33
EHECCO+SVM	88.80

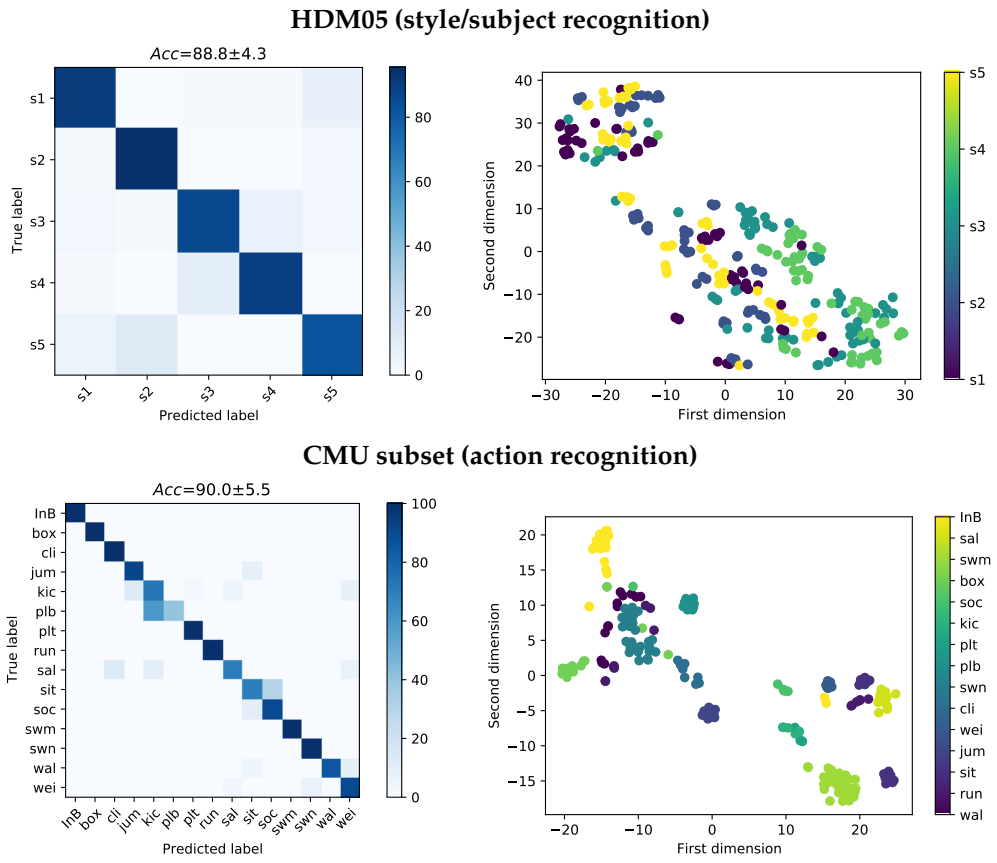

Figure 4. EHECCO-based classification results for HDM05 and CMU subset databases. **Top left**: HDM05's confusion matrix (style/subject recognition). **Top right**: HDM05 t-SNE-based 2D projection from EHECCO distance. **Bottom left**: CMU subset's confusion matrix (action recognition). **Bottom right**: CMU subset t-SNE-based 2D projection from EHECCO distance.

Table 3. Comparing results of Mocap-based action recognition (CMU subset database). The average accuracy is reported concerning the cited works vs. our approach—EHECCO+SVM.

Method	Accuracy (%)
MT+DTW [58]	82.9
SSM+DTW [18]	85.3
EMR [59]	86.7
MW+CNN [26]	90.7
EHECCO+SVM	90.0

4.2. Tennis-Mocap Results: Classification and Anthropomorphic Analysis

Figure 5 depicts the codebook generation (relevant poses) for some videos of the Tennis-Mocap database. Usually, the alienation angles, linear velocities, and angular velocities are factors to be evaluated in the training of a professional tennis player [46,62]. Nevertheless, the analysis of the action execution is costly and involves kinetic analysis with additional instrumentation [63]. Our method shows a valuable tool based only on kinematic information provided by optical sensors. Indeed, our EHECCO-based approach allows encoding the relevant poses characterizing from the time series (tennis action) without any manual frame segmentation or preprocessing. As seen, the provided codebook encodes the most relevant information in the first execution of each record and some significant variations in the posterior executions of the action.

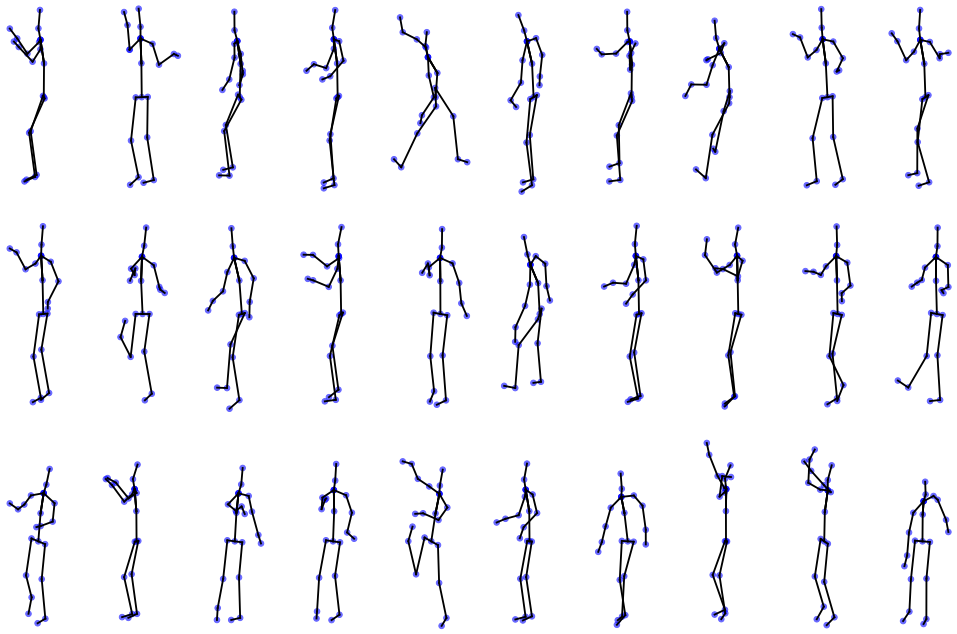

Figure 5. Illustrative results for codebook generation (Tennis-Mocap dataset). Top: *forehand*; Middle: *volley*; Bottom: *Smash*.

Regarding the classification results, as can be seen at the top of Figure 6, accuracies over 80% are attained. The lowest performance must be analyzed in conjunction with the action, where the upper limb's position in the most relevant poses makes these classes closer. Nevertheless, each record classified contains 12 to 16 continuous stroke executions without segmentation, so the confused actions depend on the execution speed after 30 s. The latter can also be corroborated by the 2D t-SNE data projection, where both the action and the players' expertise are presented. As seen, intra and interclass variability are revealed, corroborating the EHECCO's ability to highlight nonlinear patterns related to the player's performance (style/expertise) and the action behavior. However, movements

such as smash, serve, and forehand involve a significant arm span in execution, being difficult to separate. Moreover, they involve major upper-body power/strength as referred in [64]. Though the "arm span" measure is used in anthropometric tennis studies, it has no statistical significance in the early stages when classifying competitive and non-competitive players [65].

Lastly, the bottom of Figure 6 displays the Pearson's correlation-based analysis (absolute value) to compute the linear dependencies between the mean 1D t-SNE-based projection of the players' samples (from EHECOO metric) and the Tennis-Mocap dataset anthropometric measurements (see Table 1). In particular, the correlation analysis is carried out concerning the six movements performed by the players to find the incidence of each physical variable in the execution of the studied actions.

As seen, fat fold variables are highly correlated with each other, similarly to the perimeter variables. Moreover, the tennis actions share substantial correlations with the players' perimeter measurements (blue), specifically with the forehand, backhand, and volley classes. Notably, EHECOO-based interpretability follows the fact that anthropometric characteristics related to the size of the limbs and other parts of the body have a more significant influence on players' performance than features related to age, weight, height, and strength [66].

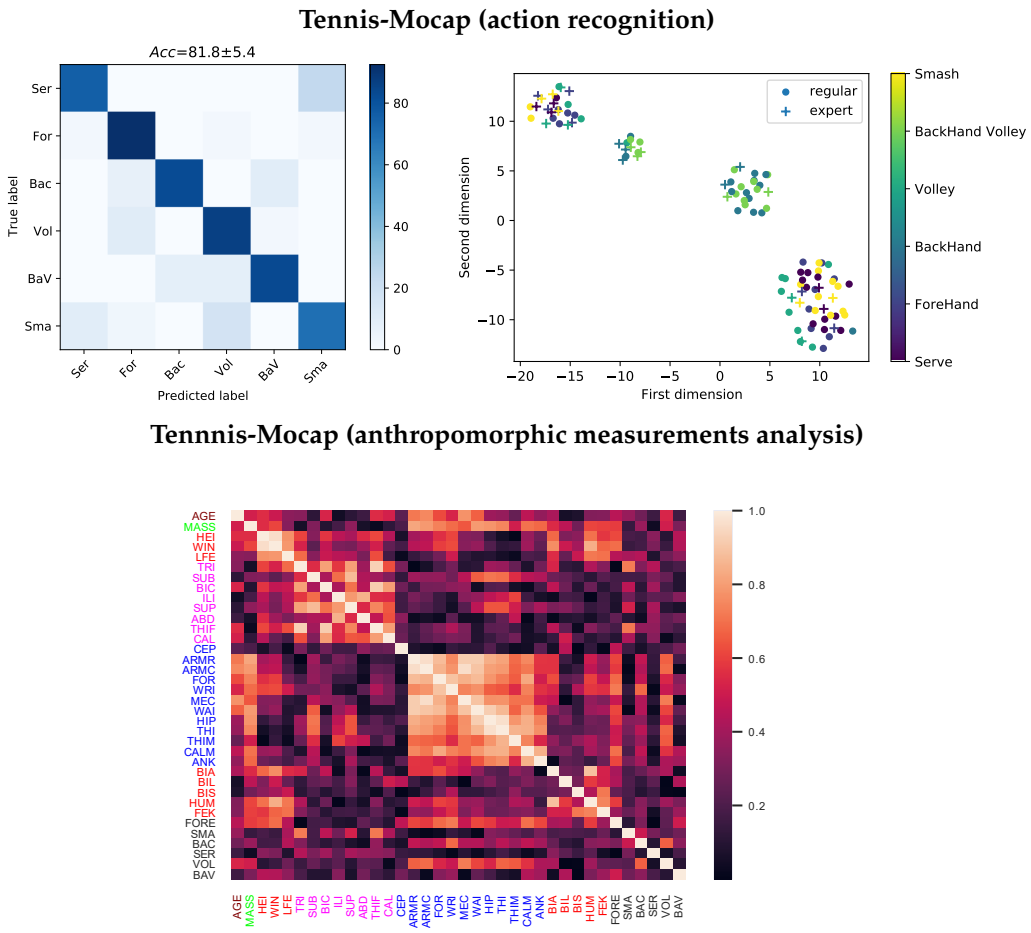

Figure 6. EHECCO-based classification and anthropomorphic measurement results for Tennis-Mocap database. **Top left**: confusion matrix (action recognition). **Top right**: t-SNE-based 2D projection from EHECCO distance. **Bottom left**: Absolute value of the Pearson's correlation coefficient between the EHECCO first t-SNE-based mean projection of each player's videos and his/her anthropomorphic measurements. The most relevant correlations are shown.

5. Concluding Remarks

We introduced a new enhanced Hilbert embedding-based framework from a cross-covariance operator, termed EHECCO, to represent and discriminate joint probability distributions in RKHS. Our approach favors the extraction of relevant nonlinear dependencies from input vectors to support the time series classification. In this sense, an EHECCO-based framework is tested to support Mocap data classification concerning style/subject and action recognition as well as anthropometric analysis. The introduced framework includes a codebook generation and a PCA-based latent space extraction for coding the most relevant frames and patterns from the Mocap series. Then, our EHECCO-based metric is computed to feed an SVM classifier. Provided experiments include the well-known public databases *HDM05* and *CMU subset* and our own dataset, *Tennis-Mocap* (also publicly available). As shown, EHECCO obtains competitive classification performances for both style and action recognition, outperforming state-of-the-art approaches. Moreover, EHECOO codes the intra and interclass variability and favors the interpretation of relevant anthropometric variables correlated with subject expertise and performed actions.

As future work, the authors plan to include other anthropometric and sports measurements to enhance the proposed framework, i.e., the arm span will be more sensitive in elite players' classification [65]. Moreover, EHECCO-based HAR applications from conventional video camaras [8], WiFi human sensing [10], and RFID [11] data will be carried out. Further, we plan to test the EHECCO metric on other types of time series, i.e., brain activity data [67]. Additionally, more elaborate classifiers and deep learning schemes can benefit from our EHECCO metric [68]. Finally, an extension of the EHECCO distance for the joint distribution of multiple spaces, not only two, is a research line of interest.

Author Contributions: Conceptualization, A.M.A.-M., J.D.P.-G. and G.C.-D.; methodology, C.K.V.-M., J.D.P.-G. and A.M.A.-M.; software, C.K.V.-M. and J.D.P.-G.; validation, L.F.V.-M., J.D.P.-G. and A.M.A.-M.; formal analysis, A.M.A.-M., and G.C.-D.; investigation, C.K.V.-M.; data curation, C.K.V.-M. and J.D.P.-G.; writing—original draft preparation, C.K.V.-M., A.M.A.-M. and G.C.-D.; writing—review and editing, A.M.A.-M. and G.C.-D.; visualization, C.K.V.-M. and L.F.V.-M.; supervision, A.M.A.-M. and G.C.-D. All authors have read and agreed to the published version of the manuscript.

Funding: Under grants provived by: "Convocatoria Doctorados Nacionales COLCIENCIAS 727 de 2015"; "Convocatoria Doctorados Nacionales COLCIENCIAS 647 de 2014" (Minciencias), and "Gestión de la Innovación y Desarrollo Tecnológico", Universidad Autónoma de Occidente, Cali-Colombia.

Institutional Review Board Statement: Ethical review and approval was waived for this study due to all public data studied here were previously submitted to ethical reviews.

Informed Consent Statement: Description and informed consents of the databases can be found at the following links: HDM05: http://resources.mpi-inf.mpg.de/HDM05/ (accessed on 5 October 2020), CMU : http://mocap.cs.cmu.edu/info.php (accessed on 5 October 2020) and Tennis-Mocap: https://github.com/jdpulgarin/Tennis-MoCap/blob/main/Copyright.md (accessed on 5 October 2020).

Data Availability Statement: The databases used in this study are public and can be found at the following links: HDM05: http://resources.mpi-inf.mpg.de/HDM05/ (accessed on 5 October 2020), CMU subset: http://mocap.cs.cmu.edu/info.php (accessed on 5 October 2020), and Tennis-Mocap: https://github.com/jdpulgarin/Tennis-MoCap (accessed on 5 October 2020).

Conflicts of Interest: The authors declare that this research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Kadu, H.; Kuo, C. Automatic human mocap data classification. *IEEE Trans. Multimed.* **2014**, *16*, 2191–2202. [CrossRef]
2. Kotsifakos, A. Case study: Model-based vs. distance-based search in time series databases. In Proceedings of the Exploratory Data Analysis (EDA) Workshop in SIAM International Conference on Data Mining (SDM), Philadelphia, PA, USA, 23–26 April 2014.
3. Anantasech, P.; Ratanamahatana, C. Enhanced Weighted Dynamic Time Warping for Time Series Classification. In Proceedings of the Third International Congress on Information and Communication Technology, London, UK, 27–28 February 2019; pp. 655–664.

4. Fawaz, H.; Forestier, G.; Weber, J.; Idoumghar, L.; Muller, P. Deep learning for time series classification: A review. *Data Min. Knowl. Discov.* **2019**, *33*, 917–963. [CrossRef]
5. Bicego, M.; Murino, V.; Figueiredo, M. Similarity-based classification of sequences using hidden Markov models. *Pattern Recognit.* **2004**, *37*, 2281–2291. [CrossRef]
6. Bicego, M.; Murino, V. Investigating hidden Markov models' capabilities in 2D shape classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 281–286. [CrossRef]
7. Tanisaro, P.; Heidemann, G. Time series classification using time warping invariant echo state networks. In Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, USA, 18–20 December 2016; pp. 831–836.
8. Nurai, T.; Naqvi, W. A research protocol of an observational study on efficacy of microsoft kinect azure in evaluation of static posture in normal healthy population. *Research Square.* **2021**, *1*, 1–9.
9. Yu, T.; Jin, H.; Tan, W.T.; Nahrstedt, K. SKEPRID: Pose and illumination change-resistant skeleton-based person re-identification. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2018**, *14*, 1–24. [CrossRef]
10. Jiang, W.; Xue, H.; Miao, C.; Wang, S.; Lin, S.; Tian, C.; Murali, S.; Hu, H.; Sun, Z.; Su, L. Towards 3D human pose construction using wifi. In Proceedings of the 26th Annual International Conference on Mobile Computing and Networking, London, UK, 21–25 September 2020; pp. 1–14.
11. Yang, C.; Wang, X.; Mao, S. RFID-Pose: Vision-Aided Three-Dimensional Human Pose Estimation With Radio-Frequency Identification. *IEEE Trans. Reliab.* **2020**. [CrossRef]
12. Božek, P.; Pivarčiová, E. Registration of holographic images based on the integral transformation. *Comput. Informatics* **2012**, *31*, 1369–1383.
13. Jozef, C.; Bozek, P.; Pivarciová, E. A new system for measuring the deflection of the beam with the support of digital holographic interferometry. *J. Electr. Eng.* **2015**, *66*, 53–56. [CrossRef]
14. de Souza, C.; Gaidon, A.; Cabon, Y.; Murray, N.; López, A. Generating human action videos by coupling 3D game engines and probabilistic graphical models. *Int. J. Comput. Vis.* **2019**, *128*, *1–32*. [CrossRef]
15. Alarcón-Aldana, A.; Callejas-Cuervo, M.; Bo, A. Upper Limb Physical Rehabilitation Using Serious Videogames and Motion Capture Systems: A Systematic Review. *Sensors* **2020**, *20*, 5989. [CrossRef]
16. Jedlička, P.; Krňoul, Z.; Kanis, J.; Železnỳ, M. Sign Language Motion Capture Dataset for Data-driven Synthesis. In Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives, Marseille, France, 11–16 May 2020; pp. 101–106.
17. Protopapadakis, E.; Voulodimos, A.; Doulamis, A.; Camarinopoulos, S.; Doulamis, N.; Miaoulis, G. Dance pose identification from motion capture data: A comparison of classifiers. *Technologies* **2018**, *6*, 31. [CrossRef]
18. Sun, C.; Junejo, I.; Foroosh, H. Motion retrieval using low-rank subspace decomposition of motion volume. In *Computer Graphics Forum*; Blackwell Publishing Ltd.: Oxford, UK, 2011; Volume 30, pp. 1953–1962.
19. Sebernegg, A.; Kán, P.; Kaufmann, H. Motion Similarity Modeling–A State of the Art Report. *arXiv* **2020**, arXiv:2008.05872.
20. Vrigkas, M.; Nikou, C.; Kakadiaris, I. A review of human activity recognition methods. *Front. Robot. AI* **2015**, *2*, 28. [CrossRef]
21. Gedat, E.; Fechner, P.; Fiebelkorn, R.; Vandenhouten, R. Human action recognition with hidden Markov models and neural network derived poses. In Proceedings of the 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, Serbia, 14–16 September 2017; pp. 000157–000162.
22. Principe, J. *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2010.
23. Pulgarin-Giraldo, J.; Alvarez-Meza, A.; Van Vaerenbergh, S.; Santamaría, I.; Castellanos-Dominguez, G. Analysis and classification of MoCap data by hilbert space embedding-based distance and multikernel learning. In Proceedings of the 23rd Iberoamerican Congress on Pattern Recognition, Madrid, Spain, 19–22 November 2018; pp. 186–193.
24. Williams, C.; Rasmussen, C. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006; Volume 2.
25. Milios, D.; Camoriano, R.; Michiardi, P.; Rosasco, L.; Filippone, M. Dirichlet-based gaussian processes for large-scale calibrated classification. *arXiv* **2018**, arXiv:1805.10915.
26. Aristidou, A.; Cohen-Or, D.; Hodgins, J.; Chrysanthou, Y.; Shamir, A. Deep motifs and motion signatures. *ACM Trans. Graph. (TOG)* **2018**, *37*, 1–13. [CrossRef]
27. Laraba, S.; Brahimi, M.; Tilmanne, J.; Dutoit, T. 3D skeleton-based action recognition by representing motion capture sequences as 2D-RGB images. *Comput. Animat. Virtual Worlds* **2017**, *28*, e1782. [CrossRef]
28. Dridi, N.; Hadzagic, M. Akaike and bayesian information criteria for hidden markov models. *IEEE Signal Process. Lett.* **2018**, *26*, 302–306. [CrossRef]
29. Singh, A.; Principe, J. Information theoretic learning with adaptive kernels. *Signal Process.* **2011**, *91*, 203–213. [CrossRef]
30. Blandon, J.; Valencia, C.; Alvarez, A.; Echeverry, J.; Alvarez, M.; Orozco, A. Shape classification using hilbert space embeddings and kernel adaptive filtering. In *International Conference Image Analysis and Recognition*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 245–251.
31. Huang, Z.; Wan, C.; Probst, T.; Van Gool, L. Deep learning on lie groups for skeleton-based action recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6099–6108.

32. Kamilaris, A.; Prenafeta-Boldú, F. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
33. Duin, R.; Pekalska, E. *Dissimilarity Representation For Pattern Recognition, The: Foundations And Applications*; World Scientific: Hackensack, NJ, USA, 2005; Volume 64.
34. García-Vega, S.; Álvarez-Meza, A.; Castellanos-Domínguez, G. MoCap Data Segmentation and Classification Using Kernel Based Multi-channel Analysis. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 495–502.
35. Müller, M. Dynamic time warping. *Information Retrieval for Music and Motion*; Springer: Cham, Switzerland, 2007; pp. 69–84.
36. Jeong, Y.; Jeong, M.; Omitaomu, O. Weighted dynamic time warping for time series classification. *Pattern Recognit.* **2011**, *44*, 2231–2240. [CrossRef]
37. Liu, X.; Sarker, M.; Milanova, M.; OGorman, L. Video-Based Monitoring and Analytics of Human Gait for Companion Robot. In Proceedings of the New Approaches for Multidimensional Signal Processing: Proceedings of International Workshop, NAMSP 2020, Sofia, Bulgaria, 9–11 July 2021; Volume 216, p. 15.
38. Liu, L.; Li, P.; Chu, M.; Cai, H. Stochastic gradient support vector machine with local structural information for pattern recognition. *Int. J. Mach. Learn. Cybern.* **2021**, *1*,1–18.
39. Smola, A.; Gretton, A.; Song, L.; Schölkopf, B. Algorithmic Learning Theory. In Proceedings of the 18th International Conference, ALT 2007, Sendai, Japan, 1–4 October 2007; Chapter A Hilbert Space Embedding for Distributions; Springer: Berlin/Heidelberg, Germany, 2007; pp. 13–31.
40. Huang, Z.; Van Gool, L. A riemannian network for spd matrix learning. *arXiv* **2016**, arXiv:1608.04233.
41. Vemulapalli, R.; Arrate, F.; Chellappa, R. Human action recognition by representing 3d skeletons as points in a lie group. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 588–595.
42. Vemulapalli, R.; Chellapa, R. Rolling rotations for recognizing human actions from 3d skeletal data. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4471–4479.
43. Gretton, A.; Bousquet, O.; Smola, A.; Schölkopf, B. Measuring statistical dependence with Hilbert-Schmidt norms. In *International Conference on Algorithmic Learning Theory*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 63–77.
44. Song, L.; Fukumizu, K.; Gretton, A. Kernel embeddings of conditional distributions: A unified kernel framework for nonparametric inference in graphical models. *IEEE Signal Process. Mag.* **2013**, *30*, 98–111. [CrossRef]
45. Zhao, J.; Xie, X.; Xu, X.; Sun, S. Multi-view learning overview: Recent progress and new challenges. *Inf. Fusion* **2017**, *38*, 43–54. [CrossRef]
46. Shimizu, T.; Hachiuma, R.; Saito, H.; Yoshikawa, T.; Lee, C. Prediction of future shot direction using pose and position of tennis player. In Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports, Nice, France, 21–25 October 2019; pp. 59–66.
47. Muandet, K.; Fukumizu, K.; Sriperumbudur, B.; Schölkopf, B. Kernel mean embedding of distributions: A review and beyond. *arXiv* **2016**, arXiv:1605.09522.
48. Sriperumbudur, B.; Gretton, A.; Fukumizu, K.; Schölkopf, B.; Lanckriet, G. Hilbert space embeddings and metrics on probability measures. *J. Mach. Learn. Res.* **2010**, *11*, 1517–1561.
49. Berlinet, A.; Thomas-Agnan, C. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2011.
50. Carter, T. An introduction to information theory and entropy. *Complex Syst. Summer Sch. Santa Fe* **2007**, *1*, 1–139.
51. Smola, A.; Gretton, A.; Song, L.; Schölkopf, B. A Hilbert space embedding for distributions. *International Conference on Algorithmic Learning Theory*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 13–31.
52. Gretton, A.; Borgwardt, K.; Rasch, M.; Schölkopf, B.; Smola, A. A Kernel Two-sample Test. *J. Mach. Learn. Res.* **2012**, *13*, 723–773.
53. Schölkopf, B.; Smola, A. *Learning with Kernels*; The MIT Press: Cambridge, MA, USA, 2002.
54. Géron, A. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*; O'Reilly Media: Sebastopol, CA, USA, 2019.
55. Jolliffe, I.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [CrossRef]
56. Álvarez-Meza, A.; Cárdenas-Peña, D.; Castellanos-Dominguez, G. Unsupervised kernel function building using maximization of information potential variability. In *Iberoamerican Congress on Pattern Recognition*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 335–342.
57. Müller, M.; Röder, T.; Clausen, M.; Eberhardt, B.; Krüger, B.; Weber, A. *Documentation Mocap Database hdm05*; University of Bonn: Bonn, Germany, 2007.
58. Müller, M.; Röder, T. Motion templates for automatic classification and retrieval of motion capture data. In Proceedings of the 2006 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, Vienna, Austria, 2–4 September 2006; pp. 137–146.
59. Kapadia, M.; Chiang, I.; Thomas, T.; Badler, N.; Kider, J. Efficient motion retrieval in large motion databases. In Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, Orlando, FL, USA, 21–23 March 2013; pp. 19–28.
60. Arora, S.; Hu, W.; Kothari, P.K. An analysis of the t-sne algorithm for data visualization. In Proceedings of the 31st Conference On Learning Theory, Stockholm, Sweden, 5–9 July 2018; pp. 1455–1462.

61. Lee, J.A.; Renard, E.; Bernard, G.; Dupont, P.; Verleysen, M. Type 1 and 2 mixtures of Kullback–Leibler divergences as cost functions in dimensionality reduction based on similarity preservation. *Neurocomputing* **2013**, *112*, 92–108. [CrossRef]
62. Landlinger, J.; Lindinger, S.; Stöggl, T.; Wagner, H.; Müller, E. Key factors and timing patterns in the tennis forehand of different skill levels. *J. Sports Sci. Med.* **2010**, *9*, 643.
63. Delgado-Garcia, G.; Vanrenterghem, J.; Munoz-Garcia, A.; Molina-Molina, A.; Soto-Hermoso, V.M. Does stroke performance in amateur tennis players depend on functional power generating capacity? *J. Sport. Med. Phys. Fit.* **2019**, *59*, 760–766. [CrossRef] [PubMed]
64. Fett, J.; Ulbricht, A.; Ferrauti, A. Impact of Physical Performance and Anthropometric Characteristics on Serve Velocity in Elite Junior Tennis Players. *J. Strength Cond. Res.* **2020**, *34*, 192–202. [CrossRef] [PubMed]
65. Tsoulfa, K.; Dalamitros, A.; Manou, V.; Stavropoulos, N.; Kellis, S. Can a one-day field testing discriminate between competitive and noncompetitive preteen tennis players? *J. Phys. Educ. Sport* **2016**, *16*, 1075–1077. [CrossRef]
66. Coulibaly, S.; Kouassi, F.; Beugré, J.B.; Kouadio, J.; Assi, A.; Sonan, N.; Kouamé, N.; Pineau, J.C. Left and right-hand correspondence of the anthropometrical parameters of the upper and manual lateral limb within professional tennis players. *Gazz. Med. Ital. Arch. Per. Sci. Med.* **2017**, *176*, 338–344.
67. García-Murillo, D.G.; Alvarez-Meza, A.; Castellanos-Dominguez, G. Single-Trial Kernel-Based Functional Connectivity for Enhanced Feature Extraction in Motor-Related Tasks. *Sensors* **2021**, *21*, 2750. [CrossRef]
68. Pomponi, J.; Scardapane, S.; Uncini, A. Bayesian neural networks with maximum mean discrepancy regularization. *Neurocomputing* **2021**. [CrossRef]