

Article

Riemannian Spatio-Temporal Features of Locomotion for Individual Recognition

Jianhai Zhang ^{*}, Zhiyong Feng, Yong Su, Meng Xing and Wanli Xue 

College of Intelligence and Computing, Tianjin University, Tianjin 300350, China; zyfeng@tju.edu.cn (Z.F.); suyong@tju.edu.cn (Y.S.); xingmeng@tju.edu.cn (M.X.); xuewanli@tju.edu.cn (W.X.)

* Correspondence: zhangjianhai@tju.edu.cn; Tel.: +86-137-521-27813

Received: 1 October 2018; Accepted: 20 December 2018; Published: 23 December 2018



Abstract: Individual recognition based on skeletal sequence is a challenging computer vision task with multiple important applications, such as public security, human–computer interaction, and surveillance. However, much of the existing work usually fails to provide any explicit quantitative differences between different individuals. In this paper, we propose a novel 3D spatio-temporal geometric feature representation of locomotion on Riemannian manifold, which explicitly reveals the intrinsic differences between individuals. To this end, we construct mean sequence by aligning related motion sequences on the Riemannian manifold. The differences in respect to this mean sequence are modeled as spatial state descriptors. Subsequently, a temporal hierarchy of covariance are imposed on the state descriptors, making it a higher-order statistical spatio-temporal feature representation, showing unique biometric characteristics for individuals. Finally, we introduce a kernel metric learning method to improve the classification accuracy. We evaluated our method on two public databases: the CMU Mocap database and the UPCV Gait database. Furthermore, we also constructed a new database for evaluating running and analyzing two major influence factors of walking. As a result, the proposed approach achieves promising results in all experiments.

Keywords: individual recognition; Riemannian manifold; spatio-temporal representation; Riemannian mean motion

1. Introduction

Individual identification from locomotion or action is a central problem of computer vision, which has attracted ever-increasing attention for its applications in biometrics and surveillance. This is due to the fact that (a) human locomotions recorded by cameras are non-contact, non-invasive, and non-cooperating, in contrast with the other biometric identification technologies, such as face, fingerprint, DNA, or iris recognition [1]. Moreover, (b) several studies have proved that more than 20 different biological characteristics in human motion are unique for each individual [2]. In the well-known psychology test of point light displays, the human in motion could be rapidly perceived from points of light attached to body joints [3]. The later research has shown that the different types of locomotion could also be discriminated [4,5]. Particularly, humans can recognize their friends or unfamiliar people under brief exposures from the motion of joints [6,7]. Therefore, locomotion-based individual identification has become flourishing in the computer vision community recently.

The observational locomotions are commonly consecutive RGB-based images or depth images. The main challenge arises from the variations in illumination, views, and appearance. Thus, the various approaches have been proposed to alleviate these influences. For example, a cross-view approach has been proposed to reconstruct the 3D structure from a 2D video sequence [8]. However, these 2D image-based approaches need to explore complicated models against many variations. The results, sometimes, are not satisfying. In contrast, with the development of optical motion capture system and

pose estimation techniques, a 3D skeletal sequence is more reliable for two reasons. Firstly, it eliminates many variations mentioned above in a sense that the variations caused by irrelevant factors, such as illumination, appearance and background etc., are never considered any longer. Secondly, each individual has a unique bio-mechanical motion pattern, which heavily depends on body weight, muscular strength, and tissue flexibility. Therefore, we can infer that the 3D joint trajectories have enough discriminative information for individual identification.

In this article, we design a study focused solely on walking and running, which is an integral part of many day-to-day behaviors. They have many advantages, such as periodicity and uniqueness, which make it suitable for recognizing different individuals. It is worth noting that there is an essential difference between individual recognition and human action recognition. In the field of individual identification, the focus is on identifying personal styles of a certain movement. In contrast, the aim of human action recognition is opposite, generalizing the actions with the variations of different people.

Three observations motivate us to consider Riemannian features: (a) The skeletal structures of different people are unified on the Riemannian manifold in order to fairly use collected data. (b) The human skeleton represented on a Riemannian manifold has a specific topological structure that is able to reflect real and rich information compared to Euclidean space. (c) A certain movement from different individuals can be approximately modeled by a unified dynamic system on Riemannian manifold [9], and the real difference between the individuals can be measured by the intrinsic geodesic distance against the mean system. Moreover, the features of the Riemannian manifold are more discriminative than the features on Euclidean space. This is due to the fact that the difference derived by geodesic along the manifold is larger than the difference in the Euclidean metric, which is illustrated in Figure 1b. In particular, the enhancement of subtle differences will essentially facilitate the tasks of individual recognition.

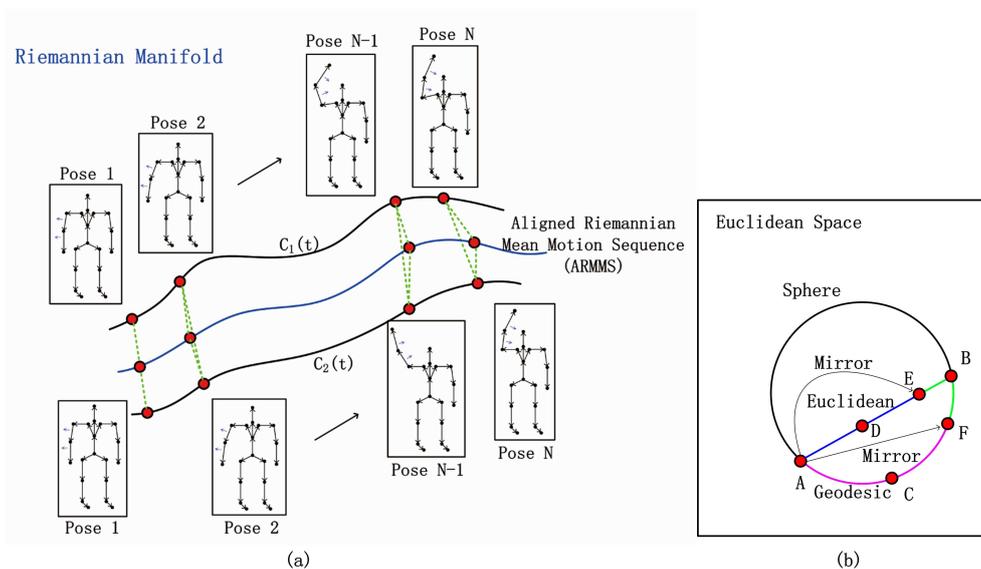


Figure 1. Illustration of the proposed (a) aligned Riemannian mean motion sequence (ARMMS). (b) A and B are the pair of poses in two different sequences $C_1(t)$ and $C_2(t)$. C and D are the mean points on the sphere and plane, respectively. F and E are mirrors of A for manifold and Euclidean cases, respectively. The difference between A and B in the green on the sphere is larger than in the case on the Euclidean plane, i.e., $\widehat{BF} > BE$.

In this paper, we develop Riemannian geometric features on locomotions for individual identification. The proposed approach models the state space of joints on the Riemannian manifold and obtains an aligned Riemannian mean motion sequence (ARMMS), as shown in Figure 1a. The unique spatial feature of each individual is naturally described by the geometrical distance between sequences and ARMMS. We then encode the unique temporal variations of spatial features by the temporal

hierarchy of covariance descriptors. In order to model the nonlinear feature relationship in an efficient manner, the spatial and the temporal features are mapped in high-dimensional feature spaces by the radial basis function (RBF) kernel. Finally, we utilize a kernel metric learning strategy to improve the classification accuracy. The proposed 3D spatio-temporal geometric feature representation based on Riemannian manifold obtains excellent results for individual recognition on both walking and running sequences. The framework is shown in Figure 2.

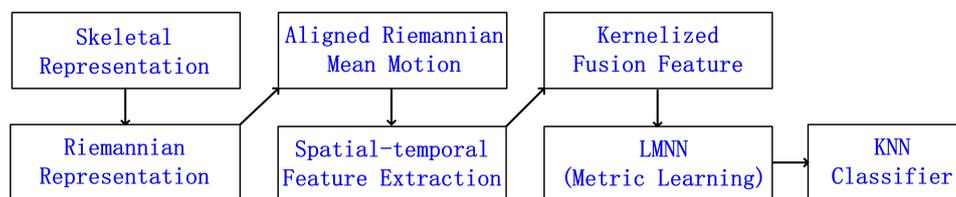


Figure 2. Overview of the proposed framework.

The contribution of the proposed work is threefold: Firstly, we propose efficient spatial and temporal features on the Riemannian manifold, which have rich statistical information and enhances the subtle differences between different individuals. Secondly, we demonstrate the effectiveness and robustness of the proposed approach by conducting extensive experiments on the sequences captured from three different sensors (the Vicon system, Kinect, and a digital camera). Finally, we also built a locomotion database for exploratory factor analysis and recognizing individuals based on walking or running.

The rest of this paper is organized as follows. Section 2 presents a review of related work. The proposed framework is described in Section 3, followed by an experimental evaluation of our method in Section 4. Finally, Section 5 concludes the paper.

2. Related Work

Individual recognition has received ever-increasing attention as a remote biometric identification technology. Most existing techniques have been developed to recognize individuals by their walking or running fashions (gait). The current methods for gait recognition can be classified into two categories: appearance-based approaches and model-based approaches.

One of the earliest appearance-based approach can be seen in [2], they build a descriptor by a binarized silhouette from a side view to recognize each individual. The binary silhouette feature has been also used to train a probabilistic model [10]. In addition, the authors of [11] present a gait recognition method to compute a gait energy image (GEI) descriptor, which is the average of all silhouette images for a single gait cycle. Inspired by the GEI, the authors of [12] introduced the depth energy image (DEI), which is the average of the depth silhouettes taken along a gait cycle. Similar work also include chrono-gait images (CGIs) [13], gait flow images (GFIs) [14], and gait energy volume (GEV) [15]. Although [16] has shown that the GEI required less computational effort and could achieve a stable result for gait recognition, the appearance is easily camouflaged. Moreover, appearance changes caused by the variations in viewing angle would cause difficulties for most of the appearance-based approaches. This problem cannot be easily avoided in practical applications [17–20].

The second category is model-based approaches, which have better performance in the case of view variants [8,21]. Several approaches are proposed to construct 3D gait information through multiple calibrated cameras [22,23]. The authors of [22] apply an image-based rendering on a 3D VH model to reconstruct gait features. The authors of [23] propose an approach to estimate the observation angle at each frame from the walking direction, and images are synthesized from 3D reconstructions. In addition, several approaches extract gait features that are robust to viewing changes [24,25]. The authors of [24] propose a method based on homography to compute view-normalized trajectories of body parts. A self-calibrating view-independent recognition approach is proposed in [25], where lower limbs are estimated based on markerless motion estimation and then reconstructed in the sagittal

plane. The model-based approaches can achieve promising performance. However, the dependence on the multiple calibrated cameras under fully controlled and co-operative environments makes these approaches hard to be implemented in a real-world environment. Moreover, the performance of gait recognition decreases as the viewing angle increases.

In recent years, human representations based on 3D skeleton data have been attracting an increasing amount of attention. Compared with 2D visual data, additional depth information provides more geometric information of motion data. The recent development of deep learning approaches make it easier to obtain 3D human poses from a single view or video sequence [26–28]. These approaches train their models by sequences captured from various motion capture (Mocap) databases (such as Human 3.6 M [29] and KTH [30]). The human skeleton representation can disentangle identity-unrelated factors, such as illumination conditions, dressing, and camera viewpoints, which make it more suitable for representing the discriminative information of individuals. The authors of [31] compute distance–time dependency signals on the the motion sequences selected from the CMU Mocap database, which express the distance variation between two specific joints over time. The authors of [32] extracted a couple of joint angles from two signature poses within a gait cycle to form a gait pattern descriptor, and classified the query subject by the baseline 1-Nearest Neighbor. It can well handle both the variations in joints and the pairwise relationship between joints. These approaches make better use of biometric cues than do the appearance-based approaches. However, these features are not discriminative enough to explain the essential difference between individuals. It should be noted that, as static features, each skeleton is unique. Using this trivial static, skeletal feature itself can still result in perfect recognition. Therefore, untreated position data of joints potentially including the precise static features would make the experiment meaningless. In order to use the Mocap data accurately and fairly, in [32], a prototypical skeleton was constructed and used to represent the bodies of all subjects, and [33] used the angle rotation features to disable static features. In our approach, we model the human poses on the Riemannian manifold, which is a unified representation of the skeleton. This representation has been used in human pose tracking and estimation approaches [9,34].

3. Our Method

In this section, we will introduce the formulations of Riemannian spatio-temporal features. Furthermore, the kernel metric learning method which is used for feature classification will also be depicted in detail.

3.1. Preliminary

Now we present three steps for pre-processing motion sequences: period extraction, alignment, and interpolation. Since human locomotion is a kind of periodic signal, a movement sequence may include several cycles. Capturing a periodic cycle means starting and stopping the sample extraction cycle when certain movements are detected. We extract action cycles by utilizing the speed auto-correlation function of ankle or toe (depending on the structure of the skeleton). This is because the foot is the first contact part with the ground and it moves up the kinematic chain. The speed auto-correlation function possessing the same cycle as the original signal can be utilized to eliminate the influence of the stochastic fluctuations from the signals [35]. The speed of a certain joint is the changing rate of its position in Euclidean space $v = |\vec{v}| = \left| \frac{ds}{dt} \right|$. A speed auto-correlation coefficient R is defined as

$$R = \frac{1}{F - 1 - |\Delta f|} \sum_{i=1}^{F-1-|\Delta f|} v_i v_{i+\Delta f} \quad (1)$$

where the lag parameter Δf is the phase shift in the number of samples, and F is the number of frames. The speed auto-correlation function is represented by a sequence of auto-correlation coefficients over increasing time lags. Since phase shifts can be performed with identical results in both positive and negative directions, the speed auto-correlation curve can be plotted symmetrically with the zeroth

shift located centrally. For a time series of speed signals during walking or running, each cycle can be represented as phase shifts between peak values of auto-correlation coefficients [36].

Similar movements performed by the same or different subjects can have different execution rates in a periodic process. To compensate for these unsynchronized sequences, we use generalized time warping (GTW) proposed by [37], which could find an optimal alignment between many multi-dimensional time series of different modalities. However, the GTW reproduces multiple identical frames for alignment purpose, which is obviously contrary to the law of movement. In order to ensure continuity and smoothness of the motion trajectory, the same frames are replaced by interpolating on Riemannian manifold.

An articulated body can be described by a tree-structure open kinematic chain [38], which refers to an assembly of rigid bodies connected by joints to provide constrained motion. In order to obtain richer statistical information about the spatial structure of the poses, we use the orientation of the link between joints i, j . Subsequently, they are represented as a series of vectors on a sphere $x_f^{ij} = \frac{x_i^j - x_f^j}{\|x_i^j - x_f^j\|_2} \in \mathbb{S}^2$. Based on the kinematic chain, the whole pose $x_f \in \mathcal{M}$ can be represented as the Cartesian product of all the relative rotations between consecutive joints [34]:

$$\mathcal{M} = \mathbb{S}^2 \times \mathbb{S}^2 \times \dots \times \mathbb{S}^2. \quad (2)$$

For brevity, in the following, we denote the joint state x_f^{ij} on the Riemannian manifold as x_f^i . The same frames $(x_{f+1}, \dots, x_{w-1})$ between two different frames x_f and x_w , are revised by uniformly interpolating for each joint j ($j = 1, \dots, J$) on a Riemannian manifold, where J is the number of joints. Specifically, for joint j , the uniform interpolation takes place on the geodesic between x_f^j and x_w^j . First, we use the logarithm operation in Equation (3) to linearize the geodesic between x_f^j and x_w^j on the tangent space of x_f :

$$v_f^j = \log_{x_f^j}(x_w^j) \quad (3)$$

where v_f^j is the tangent vector of x_f^j . The tangent vector is then cut into $w - f$ equal parts $(v_{f+1}^j, \dots, v_{w-1}^j)$. Finally, we use the exponential mapping in Equation (4) to obtain the revised results $(\hat{x}_{f+1}^j \dots \hat{x}_{w-1}^j)$ on the manifold.

$$\hat{x}_{f+1}^j = \exp_{x_f^j}(v_f^j). \quad (4)$$

3.2. Riemannian Geometric Features Representation

Individual recognition aims to identify the unique motion pattern of each individual by biometric cues. The active appearance model (AAM) states that the most important aspect of encoding for facial identity is the information about an individual face x relative to the mean face m [39]:

$$x = m + Cb \quad (5)$$

where C is a set of orthogonal modes of shape variation, and b is a set of shape parameters. However, for the articulated objects with properties of high dimensional and nonlinearity, the formulation in Equation (5) cannot model the variations in human pose.

Similar to AAM, we tend to obtain a **mean motion sequence** on the Riemannian manifold, which can be utilized to extract the discriminating features for describing individual motion pattern. To achieve this objective, for each frame of all sequences, we define a **mean pose**, which is composed of the geometric means of all joints. The mean value of a set of points on a manifold lying on a small neighborhood of manifolds can be calculated intrinsically. We use the Karcher mean [40], which is

defined as the point on the manifold that minimizes the sum of squared distances. The Karcher mean of joint j in frame f can be represented as

$$m_f^j = \arg \min_{m_f^j} \sum_{s=1}^S d^2(m_f^j, x_f^j(s)) \quad (6)$$

where S is the number of sequence.

Finally, the mean pose composed by J joints in frame f can be denoted as $\bar{x}_f = \{m_f^j\}_{j=1}^J \in R^J$, and the mean motion sequence containing F mean poses is defined as $\{\bar{x}_f\}_{f=1}^F \in R^{JF}$. Algorithm 1 describes the steps to compute mean sequence. Given a skeleton sequence $\{x_f\}_{f=1}^F \in R^{JF}$ of a certain individual, the most direct way to characterize the differences between individuals is to calculate the geometric distance against the mean sequence. For a vector x_f^j of joint j , the geodesic distance $G_{f,j}$ can be obtained by

$$G_{f,j}(x_f^j, m_f^j) = \left[\left(\log_{m_f^j} x_f^j \right) \left(\log_{m_f^j} x_f^j \right)^T \right]^{\frac{1}{2}}. \quad (7)$$

Finally, the Riemannian geometric features in frame f can be represented as a vector $G_f \in R^J$. As a result, the geometric features of sequence s ($s = 1, \dots, S$) containing F frames can be denoted as a feature vector $G(s) \in R^{FJ}$.

Algorithm 1 Computing mean sequence.

Input: a set of sequences $\{x_f(s)\}, s = (1, \dots, S)$ and $\varepsilon_1, \varepsilon_2 > 0$

Initialization: $m_f^j \leftarrow x_f^j, j = (1, \dots, J)$ and $f = (1, \dots, F)$

for $f = 1$ to F **do**

for $j = 1$ to J **do**

while $|\bar{v}_f^j| < \varepsilon_1$ **do**

 Compute $v_f^j \leftarrow \log_{m_f^j(i)}(x_f^j(s)),$ for $s = 1, \dots, S$

 Compute mean tangent vector $\bar{v}_f^j \leftarrow \frac{1}{S} \sum v(s)$

 Update $m_f^j(i): m_f^j(i+1) \leftarrow \exp_{m_f^j(i)}(\varepsilon_2 \bar{v}_f^j)$

$i = i + 1$

end while

Return: mean joint m_f^j

end for

Return: mean pose $\bar{x}_f = \{m_f^j\}_{j=1}^J$

end for

Return: mean sequence $\{\bar{x}_f\}_{f=1}^F$

3.3. Riemannian Temporal Hierarchy of Covariance Descriptors

The covariance descriptor has also been used for the tasks of human motion analysis. For human action recognition, the pose variations within a sequence is encoded as a sample covariance descriptor [41,42]. Since the symmetric positive definite (SPD) matrices can be treated as points on the Riemannian manifold, they deal with classification tasks by nearest neighbor (NN) using the log-Euclidean metric. In addition, manifolds can be also mapped to a reproducing kernel Hilbert space (RKHS) via the kernel trick [43–45]. However, the elements in the covariance descriptor contain the unique skeleton information of each individual, and this descriptor does not enclose temporal evolution information of motion sequences. Therefore, the traditional method cannot be used directly

in the task of individual identification. In order to encode the temporal dependency of geometric features over time, we use hierarchical covariance matrices, each covering a sub-sequence of the input sequence. Different from *Cov3DJ* proposed by [46], which models variations in the joints in Euclid space, the proposed approach encodes the dynamic variations of Riemannian geometric features into hierarchical covariance matrices. For each sequence, the covariance matrix can be described as a symmetric matrix $C(\cdot, \cdot) \in R^{J \times J}$. The covariance matrix of the range of frames from f_1 to f_2 can be calculated as follows:

$$C(f_1, f_2) = \frac{1}{f_2 - f_1 - 1} \sum_{f=f_1}^{f_2} (G_f - \bar{G}_f)(G_f - \bar{G}_f)^T \tag{8}$$

where \bar{G}_f is the sample mean of G_f , and the T is the transpose operator. The upper triangle of the covariance matrix in this case is $J(J + 1)/2$, which is the length of each sub-descriptor. As a typical experimental setup also used in [46], we use a two-level descriptor and allow overlap in the second levels to construct the temporal covariance descriptors. The hierarchical construction is shown in Figure 3. The top level C_{00} is computed over the entire motion sequence. The lower levels $[C_{10} \dots C_{16}]$ are computed over smaller windows, overlapping the entire sequence.

Figure 3 shows a covariance descriptor with J joints, resulting in a descriptor for each sequence $C(s) \in R^{\frac{7J(J+1)}{2}}$ ($s = 1, \dots, S$). For a more detailed and rigorous treatment, we refer the reader to the standard texts [46].

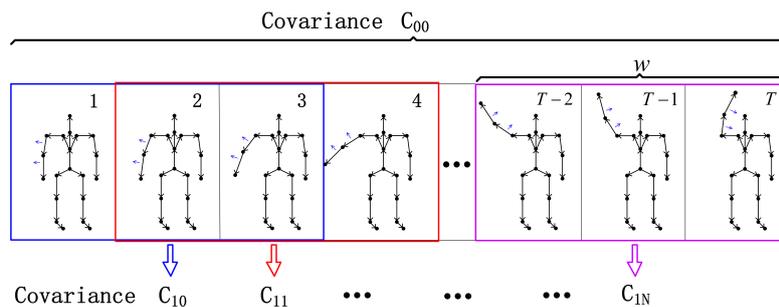


Figure 3. Temporal construction of the covariance descriptor.

3.4. Kernel Metric Learning

As a result from the previous two stages, for each training sample n ($n = 1 \dots N$), we can obtain the spatial geometric feature $G(n) \in R^{FJ}$ and the temporal hierarchy feature $C(n) \in R^{\frac{7J(J+1)}{2}}$. The objective of the feature fusion stage is to combine spatial and temporal information that improves individual recognition accuracy. The spatio-temporal features for all sequences contains the unique motion cue of each individual. However, the variation in human motion space is highly nonlinear. The performance of the large margin nearest neighbor (LMNN) [47] might degrade when the given data has a nonlinear structure. Therefore, we utilize a kernel metric learning approach to learn a low-rank distance metric in a high dimensional feature space. The metric learning approach is inspired by the kernel extension of the LMNN (KLMNN) algorithm proposed by [48,49], which is known to deliver a significant performance improvement in the recognition system. Firstly, we introduce the kernel trick to map the input spatial and temporal features to high-dimensional feature spaces. Usually, the kernel function can be written as

$$k(x_i, x_j) = \Phi(x_i)^T \Phi(x_j) \tag{9}$$

where $\Phi(\cdot)$ is an implicit nonlinear mapping, and $k(\cdot, \cdot)$ is the induced kernel function.

In this context, for the spatial geometric and the temporal hierarchy features, we use the radial basis function (RBF) kernel:

$$k_s(G(i), G(j)) = \exp\left(-\frac{\|G(i) - G(j)\|^2}{2\delta^2}\right) \quad (10)$$

$$k_t(C(i), C(j)) = \exp\left(-\frac{\|C(i) - C(j)\|^2}{2\beta^2}\right) \quad (11)$$

where δ and β are estimated as the mean of the training data pairwise distances.

After obtaining the kernelized spatial geometric and temporal hierarchy features, here we introduce a fusion scheme based on the Hadamard product of kernel functions. In particular, the pairwise similarity matrix for the all sequences is defined as the element-wise multiplication between two kernels $k_{st} = k_s \otimes k_t \in R^{N \times N}$. The n -th column of the kernel matrix k_{st} is a feature vector of the n -th sequence, which, for brevity, is denoted by $k_{st}(n)$. Instead of computing simple Euclidean distance, we use the Mahalanobis distance metric to measure the distance between these kernelized features. Considering the k -NN classifier, the loss function is a measurement of violations made by impostor samples and distancing among target neighbors. The distance between two target neighbors $k_{st}(i), k_{st}(j) \in R^N$ that we want to be closest to each other, is defined as

$$D(k_{st}(i), k_{st}(j)) = \|L(k_{st}(i) - k_{st}(j))\|_2^2. \quad (12)$$

On the other hand, we want the distance between the $k_{st}(i)$ and the impostor $k_{st}(l)$, which is closer to $k_{st}(i)$ without being targets, to keep away from each other in the function of linear transformation L . Therefore, in our model, the loss function is represented by both push and pull components, trying to pull the targets and push the impostors simultaneously. Meanwhile, a penalty term $\|L^T L - I\|_F$ is imposed to maintain orthogonal independence amongst the different coordinates of the metric space. It is defined as

$$\begin{aligned} \varepsilon(L) = & \mu \sum_{i,j \rightarrow i} \sum_{l \rightarrow i} [D(k_{st}(i), k_{st}(j)) - D(k_{st}(l), k_{st}(i)) + \zeta] \\ & + (1 - \mu) \sum_{i,j \rightarrow i} [D(k_{st}(i), k_{st}(j))] + \lambda \|L^T L - I\|_F \end{aligned} \quad (13)$$

where the notations $j \rightarrow i$ and $l \rightarrow i$ denote that j is the index of targets of sample i , while l is the index of impostors of sample i . The key idea is to learn the transformation L to be a non-square matrix of size $p \times N$, with $p \leq N$. Thus, L defines a mapping from the high-dimensional space to a low-dimensional embedding. In addition, the final regularization term in Equation (13) enforces that the equivalent metric $L^T L$ should remain close to the identity matrix. The λ and μ are the weighting parameters.

The optimal transformation L that minimizes Equation (13) can be found by solving the global optimization problem:

$$L^* = \arg \min_L \varepsilon(L). \quad (14)$$

In order to solve Equation (14) by the gradient descent approach, we compute the derivative term of $\varepsilon(L)$, as follows:

$$\begin{aligned} \frac{1}{2} \frac{\partial \varepsilon}{\partial L} = & L \sum_{i,j \rightarrow i} \sum_{l \rightarrow i} [M_{i,j} - M_{i,l}] \\ & + 2\lambda L (L^T L - I) \end{aligned} \quad (15)$$

where $M_{i,j} = (k_{st}(i) - k_{st}(j))(k_{st}(i) - k_{st}(j))^T$ and $M_{i,l} = (k_{st}(i) - k_{st}(l))(k_{st}(i) - k_{st}(l))^T$.

For giving a reasonable initialization before starting the iterations, in the feature learning stage, we use linear discriminant analysis (LDA) [50] to initialize L on the feature matrix. The optimization is performed until the maximal number of iterations is reached or the gradient below a set threshold value. In our experiment, we learn the transformation $L \in R^{p \times N}$ with maximal $epoch = 40$ and the threshold value $\delta = 0.01$.

4. Experiments

In this section, we present the results of individual identification tasks on three databases. Here, two cases are considered: (a) the 3D locations of the joints are provided (Section 4.2), and (b) the 3D joint locations are unknown (Section 4.3). In every experiment, the procedure is repeated 10 times and the accuracy is averaged. The parameters are set as $\lambda = 0.1$ and $\mu = 0.2$. On all databases, i.e., CMU, UPCVgait, and our database, the classification performances are obtained by means of stratified five-fold cross-validation. Four pieces are used for training, and one piece is used for testing. For the compared methods in CMU and UPCV, we stimulate the feature extraction and use a unified classifier for experiments.

4.1. Database

In our experiments, two publicly available databases are exploited to evaluate our approach: the CMU Mocap database [51] and the UPCV Gait database [52].

For the CMU database, the experimenter wore a black jump suit with 41 markers, and motions were recorded with an optical marker-based Vicon system. The video surveillance environment of 30 m² was surrounded by 12 cameras with a sampling rate of 120 Hz at heights ranging from 2 to 4 m. Highly accurate 3D poses were obtained in the form of relative body point coordinates in each video frame of the motion sequences. Each pose was stored in a bone-rotational form, which is a generic representation method for describing the deformation of marked joints over time.

The UPCVgait database consists of data with respect to walking sequences from 30 individuals (15 males and 15 females), which targets pose-based gender and identity recognition tasks. For each individual, there were five sequences captured during a small time cycle. Each person was asked to walk in a straight line through a corridor at a normal speed, and the Kinect sensor was placed 1.70 m above the ground, at the left of the walking path, with the sensor's principal direction at an angle of 30° relative to the walking line. The length of sequence ranges from 55 to 120 frames, depending on the walking speed of each individual. Poses with 20 key joints were estimated using Microsoft Kinect SDK at approximately 30 fps.

4.2. Evaluation with Known 3D Poses

The CMU Mocap database contains an extensive collection of data with respect to walking sequences of different subjects. However, the length of sequences varies largely, which may lead to sample imbalance, and there was no unified rule for walking process. In this experiment, subjects with more than 20 cycles and who walked straight were chosen. We select seven walking subjects (#7, #8, #12, #15, #16, #35, #39) that performed 140 walking cycles in total, with 20 cycles per subject. As shown in Figure 4(left), we used a skeleton with 19 joints. To quantitatively evaluate the performance, we compared our method with seven baselines [31,53–58].

As shown in Table 1, the proposed approach achieves a 98.70% classification rate, exceeding the second-best approach by more than 2%. Our results in this table correspond to a feature dimension of $p = 112$ and 150 frames per cycle in the alignment stage. This result proves that the proposed feature representation can reveal the unique spatio-temporal characteristics of each individual. Moreover, the proposed approach models individual motion pattern by measuring the distance from the mean motion sequence, which is more distinguishable and reliable than other compared approaches.

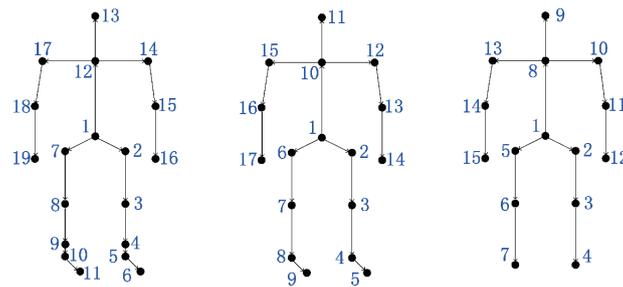


Figure 4. Skeletal structures: **left:** the CMU Mocap database, **middle:** the UPCVgait database, and **right:** our database.

Table 1. Recognition rate of all individual recognition methods tested with the CMU Mocap database.

Methods	Ball et al. [53]	Ahmed et al. [54]	Andersson et al. [55]	Jiang et al. [56]
Accuracy	46.10%	91.20%	92.60%	88.80%
Methods	Preis et al. [57]	Sedmidubsky et al. [31]	Sinha et al. [58]	Ours
Accuracy	56.00%	81.60%	96.20%	98.70%

In the UPCVgait database, we used the skeleton with 17 joints as shown in Figure 4(middle). The distorted poses in each sequence were removed in the cycle extraction stage. This is because, as shown in Figure 5, the distorted poses at the end of a sequence would show weak correlation to other poses. Since the frame rate of this database is lower than the CMU Mocap database, we set 50 frames per cycle in the GTW alignment stage. To quantitatively evaluate the performance, we compared our method with five baselines [53,57,59–61]. Among these methods, the authors in [59] applied the covariance and its dissimilarity measure concept on the skeletal trajectories. Similarly, the authors in [61] fused information from feature representations from both Euclidean and Riemannian spaces. One of the feature they used was the kernelized covariance matrix. Compared with the vector of locally aggregated descriptors (VLAD) encoding method used in [61], the proposed features not only preserve the temporal information but also depict intrinsic distance on the Riemannian manifold. As shown in Table 2, our approach achieves a 97.90% classification rate, exceeding the second-best approach [61] by more than 1%. Our results in this table correspond to a feature dimension of $p = 110$.

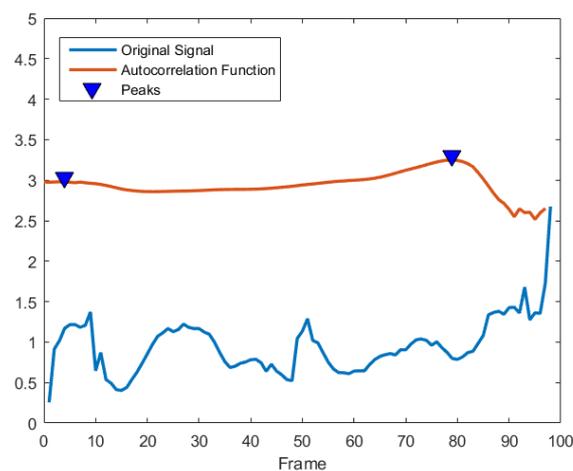


Figure 5. Motion cycle extraction by autocorrelation function.

Table 2. Recognition rate of all individual recognition methods tested with the UPCVgait database.

Methods	Ball et al. [53]	Preis et al. [57]	Kumar and Babu [59]
Accuracy	14.10%	43.00%	89.20%
Methods	Theodorakopoulos et al. [60]	Kastaniotis et al. [61]	Ours
Accuracy	94.80%	96.20%	97.90%

4.3. Evaluation with Unknown Poses

In practical applications, such as video surveillance and forensics investigation, it is very difficult to obtain real 3D human motion data. Moreover, the current available walking Mocap databases are rather small and consider little about the factors that contribute to motion performance. Contrary to the above databases captured by special sensors, we built a database that records the walking and running sequences by digital camera.

In our database, we recorded walking and running sequences from 20 individuals (10 males and 10 females). In addition, shoes and carrying condition were used for an exploratory factor analysis of walking recognition. For the first factor, we recorded the three types of shoes for each individual including sport, slipper, and leather shoes. For the second factor, we recorded the sequences where the person carries a backpack of eight kilograms on his/her back. For each impact factor, we recorded 10 cycles of motion for each individual. The camera was placed at 1.20 m above the ground and at 15 degrees relative to the starting point of the walking line. The 3D poses were captured by a unified deep neural network that presents a two-stage cascaded structure [62]. An overview of our capture system is illustrated in Figure 6. In our database, each individual was asked to walk and run at his or her preferred speed through a straight course. The length of the course was approximately 6 m. We used a 7D camera manufactured by Canon Inc. The original image size and frame rate were 1280×720 pixels and 50 fps, respectively. In order to reconstruct the 3D pose from each frame, each frame was cut to a man-centered size of 256×256 . Figure 7 shows example frames in walking sequences of Subject 6. The source codes and data (RGB and 3D skeleton sequences) will be available at our research group website.

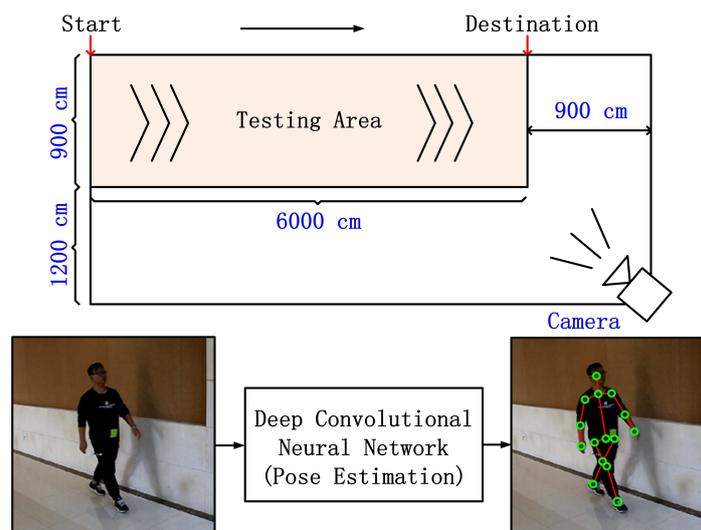


Figure 6. Overview of the capture system and the captured images. The top is the testing field of our database, and the bottom is one frame in a single gait sequence of our database, which is captured from the prescribed walking direction.

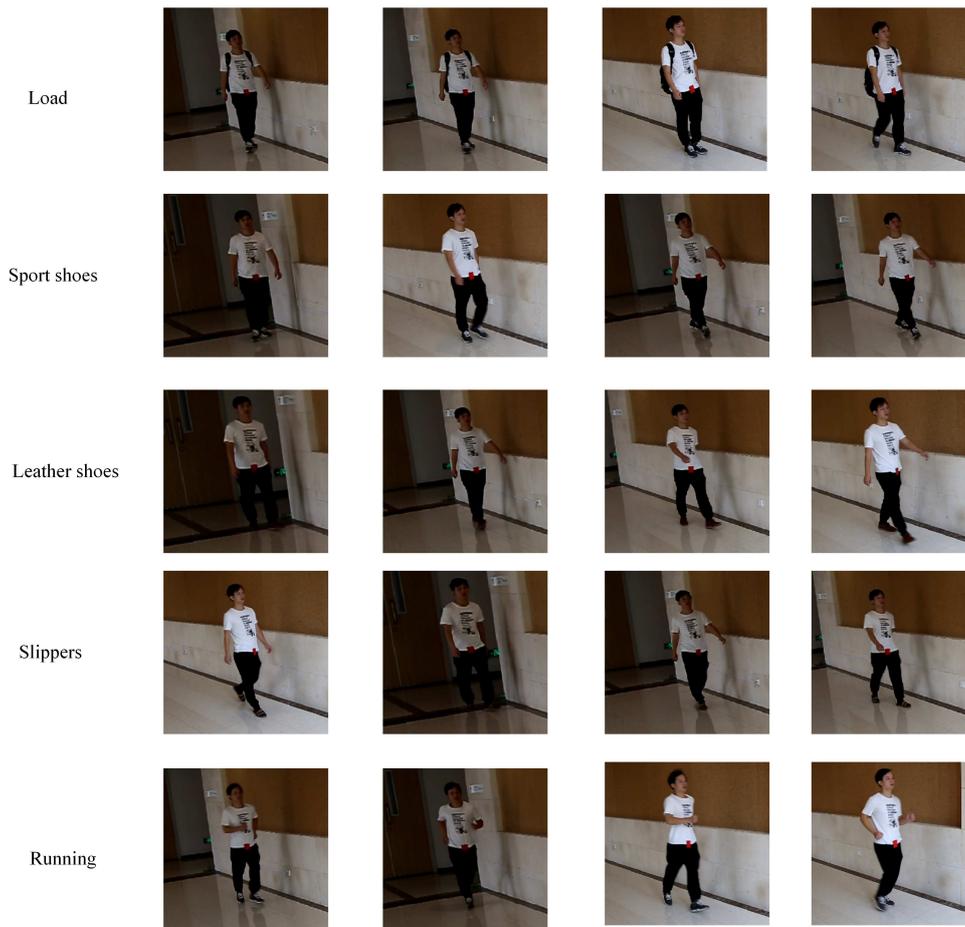


Figure 7. Example frames of Subject 6 in our database.

Since the proposed features were extracted from the Riemannian manifold, which essentially describes the system of human movement, once the skeleto-muscular structure was influenced by other external factors (such as shoes or baggage), the motor system changed. Therefore, we modeled different factors separately. In the second row of Table 3, we report the recognition accuracies of walking sequences in our database. From the third row to the fifth row are the recognition accuracies with different influence factors. In addition, in the last row of Table 3, the result suggests that each individual also has a unique motion pattern for running. We can see that our features are robust under conditions of many factors.

Table 3. Recognition rate of the proposed method on different sequences in our database.

Feature Dimension	$p = 100$	$p = 100$	$p = 120$	$p = 130$	$p = 140$	$p = 150$	$p = 160$
Walking	87.1%	90.2%	92.0%	93.4%	96.5%	98.3%	99.7%
Load	82.1%	84.3%	88.8%	91.3%	93.1%	96.1%	96.2%
Leather	85.2%	87.5%	90.7%	93.2%	95.0%	97.6%	98.2%
Slippers	84.5%	86.8%	87.3%	91.5%	93.2%	96.9%	97.3%
Running	82.5%	85.6%	88.9%	92.2%	95.5%	96.4%	96.8%

Figure 8 shows the recognition accuracy of the proposed approach in walking sequences versus feature dimensions. We can see that the fusion feature (KRGF+KRTHC) outperforms both the kernelized Riemannian geometric feature (KRGF) and the kernelized Riemannian temporal hierarchy of covariance (KRTHC) feature by a considerable margin. Specifically, the KRTHC feature can model pose variations within a sequence, but it may ignore the important local differences between individuals. Combining the local features encoded by the KRGF and the large-scale temporal features produced by

the KRTHC feature, the proposed approach can simultaneously capture the varying spatial distribution of motion and its temporal variations.

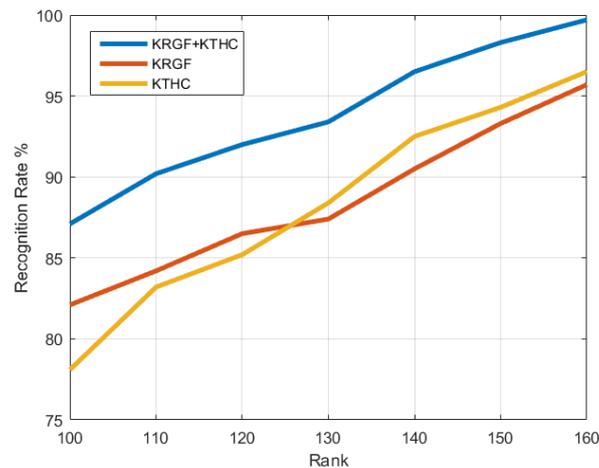


Figure 8. The recognition rate for three proposed features with different feature dimensions: the kernelized Riemannian geometric feature (KRGF), the kernelized Riemannian temporal hierarchy of covariance (KRTHC) feature, and their fusion (KRGF + KRTHC) on the walking sequence in our database.

5. Conclusions

In this work, we focus on the problem of individual identification from the locomotion sequence. We model the human motion system on the Riemannian manifold, which is an essential characterization of human movement. The aligned Riemannian mean motion sequence provides the common information among different individuals. The spatial state descriptors of individuals can be intuitively expressed by the geometrical distance between a testing sequence and the aligned Riemannian mean motion sequence. In order to take the temporal structure of the human movement into consideration, we modeled the dynamic variations of static geometric features by the temporal hierarchy covariance descriptors. The spatial and temporal features were then efficiently combined by a kernel metric learning method, which was used to obtain the most discriminative information from the feature space. Classification results using the proposed method achieve state-of-the-art results in both walking and running sequences. Combined with pose estimation techniques performed via deep learning methods, our approach can be used to recognize individuals in real scenarios.

Author Contributions: Conceptualization: J.Z. and Y.S.; data curation: M.X.; formal analysis: Z.F.; funding acquisition: Z.F.; methodology, J.Z.; supervision: W.X.; writing—original draft: J.Z.; writing—review & editing: J.Z., Z.F., and Y.S.

Funding: This work was funded by the National Key Technology R&D Program of China grant number 2017YFB1401200.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mason, J.E.; Traoré, I.; Woungang, I. *Machine Learning Techniques for Gait Biometric Recognition: Using the Ground Reaction Force*; Springer: Berlin, Germany, 2016.
2. Kale, A.; Sundaresan, A.; Rajagopalan, A.N.; Cuntoor, N.P.; Roy-Chowdhury, A.K.; Kruger, V.; Chellappa, R. Identification of humans using gait. *IEEE Trans. Image Process.* **2004**, *13*, 1163–1173. [[CrossRef](#)]
3. Johansson, G. Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* **1973**, *14*, 201–211. [[CrossRef](#)]

4. Dittrich, W.H. Action categories and the perception of biological motion. *Perception* **1993**, *22*, 15–22. [[CrossRef](#)] [[PubMed](#)]
5. Bingham, G.P.; Schmidt, R.C.; Rosenblum, L.D. Dynamics and the orientation of kinematic forms in visual event recognition. *J. Exp. Psychol. Hum. Percept. Perform.* **1995**, *21*, 1473. [[CrossRef](#)] [[PubMed](#)]
6. Cutting, J.E.; Kozlowski, L.T. Recognizing friends by their walk: Gait perception without familiarity cues. *Bull. Psychon. Soc.* **1977**, *9*, 353–356. [[CrossRef](#)]
7. Stevenage, S.V.; Nixon, M.S.; Vince, K. Visual analysis of gait as a cue to identity. *Appl. Cognit. Psychol.* **1999**, *13*, 513–526. [[CrossRef](#)]
8. Wu, Z.; Huang, Y.; Wang, L.; Wang, X.; Tan, T. A comprehensive study on cross-view gait based human identification with deep cnns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 209–226. [[CrossRef](#)] [[PubMed](#)]
9. Leonardos, S.; Zhou, X.; Daniilidis, K. Articulated motion estimation from a monocular image sequence using spherical tangent bundles. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 587–593.
10. Cheng, M.H.; Ho, M.F.; Huang, C.L. Gait analysis for human identification through manifold learning and HMM. *Pattern Recognit.* **2008**, *41*, 2541–2553. [[CrossRef](#)]
11. Han, J.; Bhanu, B. Individual recognition using gait energy image. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 316. [[CrossRef](#)]
12. Rougier, C.; Auvinet, E.; Meunier, J.; Mignotte, M.; De Guise, J.A. Depth energy image for gait symmetry quantification. In Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, Boston, MA, USA, 30 August–3 September 2011; pp. 5136–5139.
13. Wang, C.; Zhang, J.; Pu, J.; Yuan, X.; Wang, L. Chrono-Gait Image: A Novel Temporal Template for Gait Recognition. In Proceedings of the European Conference on Computer Vision, Heraklion, Crete, Greece, 5–11 September 2010; pp. 257–270.
14. Lam, T.H.W.; Cheung, K.H.; Liu, J.N.K. Gait flow image: A silhouette-based gait representation for human identification. *Pattern Recognit.* **2011**, *44*, 973–987. [[CrossRef](#)]
15. Sivapalan, S.; Chen, D.; Denman, S.; Sridharan, S.; Fookes, C. Gait energy volumes and frontal gait recognition using depth images. In Proceedings of the 2011 International Joint Conference on Biometrics (IJCB), Washington, DC, USA, 11–13 October 2011; pp. 1–6.
16. Iwama, H.; Okumura, M.; Makihara, Y.; Yagi, Y. The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 1511–1521. [[CrossRef](#)]
17. Kale, A.; Chowdhury, A.K.R.; Chellappa, R. Towards a View Invariant Gait Recognition Algorithm. In Proceedings of the Advanced Video and Signal Based Surveillance, Miami, FL, USA, 22 July 2003; pp. 143–150.
18. Kusakunniran, W.; Wu, Q.; Zhang, J.; Li, H. Gait Recognition Under Various Viewing Angles Based on Correlated Motion Regression. *IEEE Trans. Circuits Syst. Video Technol.* **2012**, *22*, 966–980. [[CrossRef](#)]
19. Kusakunniran, W.; Wu, Q.; Zhang, J.; Li, H. Support vector regression for multi-view gait recognition based on local motion feature selection. In Proceedings of the Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 974–981.
20. Jean, F.; Bergevin, R.; Albu, A.B. Computing and evaluating view-normalized body part trajectories. *Image Vis. Comput.* **2009**, *27*, 1272–1284. [[CrossRef](#)]
21. Ariyanto, G.; Nixon, M.S. Model-based 3D gait biometrics. In Proceedings of the 2011 International Joint Conference on Biometrics (IJCB), Washington, DC, USA, 11–13 October 2011; pp. 1–7.
22. Bodor, R.; Drenner, A.; Fehr, D.; Masoud, O.; Papanikolopoulos, N. View-independent human motion classification using image-based reconstruction. *Image Vis. Comput.* **2009**, *27*, 1194–1206. [[CrossRef](#)]
23. Iwashita, Y.; Ogawara, K.; Kurazume, R. Identification of people walking along curved trajectories. *Pattern Recognit. Lett.* **2014**, *48*, 60–69. [[CrossRef](#)]
24. Jean, F.; Albu, A.B.; Bergevin, R. Towards view-invariant gait modeling: Computing view-normalized body part trajectories. *Pattern Recognit.* **2009**, *42*, 2936–2949. [[CrossRef](#)]
25. Goffredo, M.; Bouchrika, I.; Carter, J.N.; Nixon, M.S. Self-calibrating view-invariant gait biometrics. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2010**, *40*, 997–1008. [[CrossRef](#)] [[PubMed](#)]
26. Tome, D.; Russell, C.; Agapito, L. Lifting from the deep: Convolutional 3d pose estimation from a single image. *arXiv* **2017**, arXiv:1701.00295.

27. Mehta, D.; Sridhar, S.; Sotnychenko, O.; Rhodin, H.; Shafiei, M.; Seidel, H.P.; Xu, W.; Casas, D.; Theobalt, C. VNect: Real-time 3D Human Pose Estimation with a Single RGB Camera. *arXiv* **2017**, arXiv:1705.01583.
28. Tekin, B.; Marquez Neila, P.; Salzmann, M.; Fua, P. Learning to Fuse 2D and 3D Image Cues for Monocular Body Pose Estimation. In Proceedings of the International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; number EPFL-CONF-230311.
29. Ionescu, C.; Papava, D.; Olaru, V.; Sminchisescu, C. Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1325–1339. [[CrossRef](#)]
30. Schuldt, C.; Laptev, I.; Caputo, B. Recognizing Human Actions: A Local SVM Approach. In Proceedings of the International Conference on Pattern Recognition, Washington, DC, USA, 27 June 2004; Volume 3, pp. 32–36.
31. Sedmidubsky, J.; Valcik, J.; Balazia, M.; Zezula, P. Gait recognition based on normalized walk cycles. In Proceedings of the 8th International Symposium on Visual Computing (ISVC 2012), Crete, Greece, 16–18 July 2012; pp. 11–20.
32. Balazia, M.; Plataniotis, K. Human Gait Recognition from Motion Capture Data in Signature Poses. *IET Biometr.* **2016**, *6*, 129–137. [[CrossRef](#)]
33. Yam, C.Y.; Nixon, M.; Carter, J. Extended model-based automatic gait recognition of walking and running. In *Audio-and Video-Based Biometric Person Authentication*; Springer: Berlin, Germany, 2001; pp. 278–283.
34. Simo-Serra, E.; Torras, C.; Moreno-Noguer, F. 3D human pose tracking priors using geodesic mixture models. *Int. J. Comput. Vision* **2017**, *122*, 388–408. [[CrossRef](#)]
35. Ribeiro, P.C.; Santos-Victor, J. Human Activity Recognition from Video: modeling, feature selection and classification architecture. In Proceedings of the International Workshop on Human Activity Recognition and Modeling (HAREM), Oxford, UK, 9 September 2005.
36. Moenilssen, R.; Helbostad, J.L. Estimation of gait cycle characteristics by trunk accelerometry. *J. Biomech.* **2004**, *37*, 121–126. [[CrossRef](#)]
37. Zhou, F.; De la Torre, F. Generalized time warping for multi-modal alignment of human motion. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 1282–1289.
38. Yamane, K. *Simulating and Generating Motions of Human Figures*; Springer Publishing Company, Incorporated: Berlin, Germany, 2010; pp. 1–11.
39. Rhodes, G.; Carey, S.; Byatt, G.; Proffitt, F. Coding spatial variations in faces and simple shapes: A test of two models. *Vis. Res.* **1998**, *38*, 2307–2321. [[CrossRef](#)]
40. Karcher, H. Riemannian center of mass and mollifier smoothing. *Commun. Pure Appl. Math.* **1977**, *30*, 509–541. [[CrossRef](#)]
41. Turaga, P.; Chellappa, R. Nearest-neighbor search algorithms on non-Euclidean manifolds for computer vision applications. In Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing, Hennai, India, 12–15 December 2010; pp. 282–289.
42. Guo, K.; Ishwar, P.; Konrad, J. Action Recognition From Video Using Feature Covariance Matrices. *IEEE Trans. Image Process.* **2013**, *22*, 2479–2494. [[PubMed](#)]
43. Harandi, M.T.; Sanderson, C.; Wiliem, A.; Lovell, B.C. Kernel analysis over Riemannian manifolds for visual recognition of actions, pedestrians and textures. In Proceedings of the Applications of Computer Vision, Breckenridge, CO, USA, 9–11 January 2012; pp. 433–439.
44. Jayasumana, S.; Hartley, R.; Salzmann, M.; Li, H.; Harandi, M. Optimizing over Radial Kernels on Compact Manifolds. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 7 June 2014; pp. 3802–3809.
45. Vemulapalli, R.; Pillai, J.K.; Chellappa, R. Kernel learning for extrinsic classification of manifold features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1782–1789.
46. Hussein, M.E.; Torki, M.; Gowayyed, M.A.; El-Saban, M. Human Action Recognition Using a Temporal Hierarchy of Covariance Descriptors on 3D Joint Locations. *IJCAI* **2013**, *13*, 2466–2472.
47. Weinberger, K.Q.; Saul, L.K. Distance metric learning for large margin nearest neighbor classification. *J. Mach. Learn. Res.* **2009**, *10*, 207–244.

48. Torresani, L.; Lee, K.C. Large margin component analysis. In Proceedings of the Neural Information Processing Systems, Vancouver, BC, Canada, 3–6 December 2007; pp. 1385–1392.
49. Yang, L.; Cheng, J.; Liu, H. Person Re-Identification Based on Kernel Large Margin Nearest Neighbor Classification. In *International Conference in Communications, Signal Processing, and Systems*; Springer: Berlin, Germany, 2016; pp. 783–791.
50. Balakrishnama, S.; Ganapathiraju, A. Linear Discriminant Analysis A Brief Tutorial. *Proc. Int. Jt. Conf. Neural Netw.* **1998**, *3*, 387–391.
51. CMU. Carnegie Mellon University Motion Capture Database. 2014. Available online: <http://mocap.cs.cmu.edu> (accessed on 1 September 2018).
52. Kastaniotis, D.; Theodorakopoulos, I.; Theoharatos, C.; Economou, G.; Fotopoulos, S. A framework for gait-based recognition using Kinect. *Pattern Recognit. Lett.* **2015**, *68*, 327–335. [[CrossRef](#)]
53. Ball, A.; Rye, D.; Ramos, F.; Velonaki, M. Unsupervised clustering of people from ‘skeleton’ data. In Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction, Boston, MA, USA, 5–8 March 2012; pp. 225–226.
54. Ahmed, M.; Al-Jawad, N.; Sabir, A.T. Gait recognition based on Kinect sensor. *Real-Time Image Video Process.* **2014**, *9139*, 91390B.
55. Andersson, V.; Dutra, R.; Araújo, R. Anthropometric and human gait identification using skeleton data from Kinect sensor. In Proceedings of the 29th Annual ACM Symposium on Applied Computing, Gyeongju, Korea, 24–28 March 2014; pp. 60–61.
56. Jiang, S.; Wang, Y.; Zhang, Y.; Sun, J. Real time gait recognition system based on kinect skeleton feature. In *Asian Conference on Computer Vision*; Springer: Berlin, Germany, 2014; pp. 46–57.
57. Preis, J.; Kessel, M.; Werner, M.; Linnhoff-Popien, C. Gait recognition with kinect. In Proceedings of the 1st International Workshop on Kinect in Pervasive Computing, New Castle, UK, 8–22 June 2012; pp. P1–P4.
58. Sinha, A.; Chakravarty, K.; Bhowmick, B. Person identification using skeleton information from kinect. In Proceedings of the Computer-Human Interactions, Cape Town, South Africa, 2–6 September 2013; pp. 101–108.
59. Kumar, M.; Babu, R.V. Human gait recognition using depth camera: a covariance based approach. In Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing, Mumbai (Bombay), India, 16–19 December 2012; p. 20.
60. Theodorakopoulos, I.; Kastaniotis, D.; Economou, G.; Fotopoulos, S. Pose-based human action recognition via sparse representation in dissimilarity space. *J. Vis. Commun. Image Represent.* **2014**, *25*, 12–23. [[CrossRef](#)]
61. Kastaniotis, D.; Theodorakopoulos, I.; Economou, G.; Fotopoulos, S. Gait based recognition via fusing information from Euclidean and Riemannian manifolds. *Pattern Recognit. Lett.* **2016**, *84*, 245–251. [[CrossRef](#)]
62. Zhou, X.; Huang, Q.; Sun, X.; Xue, X.; Wei, Y. Weakly-supervised Transfer for 3D Human Pose Estimation in the Wild. *arXiv* **2016**, arXiv:1704.02447.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).