Joint International Workshops on Structural and Syntactic Pattern Recognition
and Statistical Techniques in Pattern Recognition 2016, preprint

Walker-Independent Features for Gait Recognition from Motion Capture Data

Michal Balazia (0000-0001-7153-9984) and Petr Sojka (0000-0002-5768-4007)

Faculty of Informatics, Masaryk University, Botanická 68a, 602 00 Brno, Czech Republic
xbalazia@mail.muni.cz and sojka@fi.muni.cz

Abstract MoCap-based human identification, as a pattern recognition discipline, can be optimized using a machine learning approach. Yet in some applications such as video surveillance new identities can appear on the fly and labeled data for all encountered people may not always be available. This work introduces the concept of learning walker-independent gait features directly from raw joint coordinates by a modification of the Fisher's Linear Discriminant Analysis with Maximum Margin Criterion. Our new approach shows not only that these features can discriminate different people than who they are learned on, but also that the number of learning identities can be much smaller than the number of walkers encountered in the real operation.

1 Introduction

Recent rapid improvement in motion capture (MoCap) sensor accuracy brought affordable technology that can identify walking people. MoCap technology provides video clips of walking individuals containing structural motion data. The format keeps an overall structure of the human body and holds estimated 3D positions of major anatomical landmarks as the person moves. MoCap data can be collected online by a system of multiple cameras (Vicon) or a depth camera (Microsoft Kinect). To visualize motion capture data (see Figure 1), a simplified stick figure representing the human skeleton (graph of joints connected by bones) can be recovered from body point spatial coordinates.

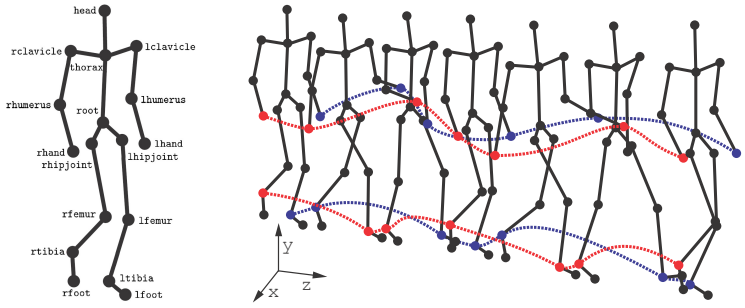

Figure 1. Motion capture data. Skeleton is represented by a stick figure of 31 joints (only 17 are drawn here). Seven selected video frames of a walk sequence contain 3D coordinates of each joint in time. The red and blue lines track trajectories of hands and feet. [18]

Recognizing a person by walk involves capturing and normalizing their walk sample, extracting gait features to compose a template, and finally querying a central database for a set of similar templates to report the most likely identity. This work focuses on extracting robust and discriminative gait features from raw MoCap data.

Many geometric gait features have been introduced over the past few years. They are typically combinations of static body parameters (bone lengths, person's height) [12] with dynamic gait features such as step length, walk speed, joint angles and inter-joint distances [4,1,12,15], along with various statistics (mean, standard deviation or maximum) of their signals [3]. Clearly, these features are schematic and human-interpretable, which is convenient for visualizations and for intuitive understanding, but unnecessary for automatic gait recognition. Instead, this application prefers learning features that maximally separate the identity classes and are not limited by such dispensable factors.

Methods for 2D gait recognition extensively use machine learning models for extracting gait features, such as principal component analysis and multi-scale shape analysis [8], genetic algorithms and kernel principal component analysis [17], radial basis function neural networks [20], or convolutional neural networks [7]. All of those and many other models are reasonable to be utilized also in 3D gait recognition.

In the video surveillance environment data need to be acquired without walker's consent and new identities can appear on the fly. Here and also in other applications where labels for all encountered people may not always be available, we value features that have a high power in distinguishing all people and not exclusively who they were learned on. We call these walker-independent features. The main idea is to statistically learn what aspects of walk people generally differ in and extract those as gait features. The features are learned in a supervised manner, as described in the following section.

2 Learning Gait Features

In statistical pattern recognition, reducing space dimensionality is a common technique to overcome class estimation problems. Classes are discriminated by projecting high-dimensional input data onto low-dimensional sub-spaces by linear transformations with the goal of maximizing the class separability. We are interested in finding an optimal feature space where a gait template is close to those of the same walker and far from those of different walkers.

Let the model of a human body have J joints and all samples be linearly normalized to their average length T. Labeled learning data in the measurement space $\mathcal{G}_L$ are in the form $\{(\mathbf{g}_n, \ell_n)\}_{n=1}^{N_L}$ where

$$\mathbf{g}_n = \left[[\gamma_1(1) \cdots \gamma_J(1)]^\top \cdots [\gamma_1(T) \cdots \gamma_J(T)]^\top \right]^\top \tag{1}$$

is a gait sample (one gait cycle) in which $\gamma_j(t) \in \mathbb{R}^3$ are 3D spatial coordinates of a joint $j \in \{1, \ldots, J\}$ at time $t \in \{1, \ldots, T\}$ normalized with respect to the person's position and walk direction. See that $\mathcal{G}_L$ has dimensionality $D = 3JT$. Each learning sample falls strictly into one of the learning identity classes $\{\mathcal{I}_c\}_{c=1}^{C}$ determined by ℓ_n. A class $\mathcal{I}_c \subseteq \mathcal{G}_L$ has N_c samples. The classes are complete and mutually exclusive. We say that learning samples $(\mathbf{g}_n, \ell_n)$ and $(\mathbf{g}_{n'}, \ell_{n'})$ share a common walker if and only if they belong to the same class, i.e., $(\mathbf{g}_n, \ell_n), (\mathbf{g}_{n'}, \ell_{n'}) \in \mathcal{I}_c \Leftrightarrow \ell_n = \ell_{n'}$.

We measure class separability of a given feature space by a representation of the Maximum Margin Criterion (MMC) [11,13] used by the Vapnik's Support Vector Machines (SVM) [19]

$$\mathcal{J} = \frac{1}{2} \sum_{c,c'=1}^{C_L} \left((\mu_c - \mu_{c'})^\top (\mu_c - \mu_{c'}) - \mathrm{tr}\,(\mathbf{\Sigma}_c + \mathbf{\Sigma}_{c'}) \right) \tag{2}$$

which is actually a summation of $\frac{1}{2} C_L (C_L - 1)$ between-class margins. The margin is defined as the Euclidean distance of class means minus both individual variances (traces of scatter matrices $\mathbf{\Sigma}_c = \frac{1}{N_c} \sum_{n=1}^{N_c} \left(\mathbf{g}_n^{(c)} - \mu_c \right) \left(\mathbf{g}_n^{(c)} - \mu_c \right)^\top$ and similarly for $\mathbf{\Sigma}_{c'}$). For the whole labeled data, we denote the between- and within-class and total scatter matrices

$$\begin{aligned}
\mathbf{\Sigma}_\mathrm{B} &= \sum_{c=1}^{C_L} (\mu_c - \mu) (\mu_c - \mu)^\top \\
\mathbf{\Sigma}_\mathrm{W} &= \sum_{c=1}^{C_L} \frac{1}{N_c} \sum_{n=1}^{N_c} \left(\mathbf{g}_n^{(c)} - \mu_c \right) \left(\mathbf{g}_n^{(c)} - \mu_c \right)^\top \\
\mathbf{\Sigma}_\mathrm{T} &= \sum_{c=1}^{C_L} \frac{1}{N_c} \sum_{n=1}^{N_c} \left(\mathbf{g}_n^{(c)} - \mu \right) \left(\mathbf{g}_n^{(c)} - \mu \right)^\top = \mathbf{\Sigma}_\mathrm{B} + \mathbf{\Sigma}_\mathrm{W}
\end{aligned} \tag{3}$$

where $\mathbf{g}_n^{(c)}$ denotes the n-th sample in class $\mathcal{I}_c$ and μ_c and μ are sample means for class $\mathcal{I}_c$ and the whole data set, respectively, that is, $\mu_c = \frac{1}{N_c} \sum_{n=1}^{N_c} \mathbf{g}_n^{(c)}$ and $\mu = \frac{1}{N_L} \sum_{n=1}^{N_L} \mathbf{g}_n$. Now we obtain

$$\begin{aligned}
\mathcal{J} &= \frac{1}{2} \sum_{c,c'=1}^{C_L} (\mu_c - \mu_{c'})^\top (\mu_c - \mu_{c'}) - \frac{1}{2} \sum_{c,c'=1}^{C_L} \mathrm{tr}\,(\mathbf{\Sigma}_c + \mathbf{\Sigma}_{c'}) \\
&= \frac{1}{2} \sum_{c,c'=1}^{C_L} (\mu_c - \mu + \mu - \mu_{c'})^\top (\mu_c - \mu + \mu - \mu_{c'}) - \sum_{c=1}^{C_L} \mathrm{tr}\,(\mathbf{\Sigma}_c) \\
&= \mathrm{tr} \left(\sum_{c=1}^{C_L} (\mu_c - \mu) (\mu_c - \mu)^\top \right) - \mathrm{tr} \left(\sum_{c=1}^{C_L} \mathbf{\Sigma}_c \right) \\
&= \mathrm{tr}\,(\mathbf{\Sigma}_\mathrm{B}) - \mathrm{tr}\,(\mathbf{\Sigma}_\mathrm{W}) = \mathrm{tr}\,(\mathbf{\Sigma}_\mathrm{B} - \mathbf{\Sigma}_\mathrm{W})\,.
\end{aligned} \tag{4}$$

Since $\mathrm{tr}\,(\mathbf{\Sigma}_\mathrm{B})$ measures the overall variance of the class mean vectors, a large one implies that the class mean vectors scatter in a large space. On the other hand, a small $\mathrm{tr}\,(\mathbf{\Sigma}_\mathrm{W})$ implies that classes have a small spread. Thus, a large $\mathcal{J}$ indicates that samples are close to each other if they share a common walker but are far from each other if they are performed by different walkers. Extracting features, that is, transforming the input data in the measurement space into a feature space of higher $\mathcal{J}$, can be used to link new observations of walkers more successfully.

Feature extraction is given by a linear transformation (feature) matrix $\mathbf{\Phi} \in \mathbb{R}^{D \times \widehat{D}}$ from a D-dimensional measurement space $\mathcal{G} = \{\mathbf{g}_n\}_{n=1}^{N}$ of not necessarily labeled gait samples to a $\widehat{D}$-dimensional feature space $\widehat{\mathcal{G}} = \{\widehat{\mathbf{g}}_n\}_{n=1}^{N}$ of gait templates where $\widehat{D} < D$ and each gait sample $\mathbf{g}_n$ is transformed into a gait template $\widehat{\mathbf{g}}_n = \mathbf{\Phi}^\top \mathbf{g}_n$. The objective is to learn a transform $\mathbf{\Phi}$ that maximizes MMC in the feature space

$$\mathcal{J}(\mathbf{\Phi}) = \operatorname{tr}\left(\mathbf{\Phi}^\top (\mathbf{\Sigma}_\mathrm{B} - \mathbf{\Sigma}_\mathrm{W})\, \mathbf{\Phi}\right). \tag{5}$$

Once the transformation is found, all measured samples are transformed into templates (in the feature space) along with the class means and covariances. The templates are compared by the Mahalanobis distance function

$$\widehat{\delta}(\widehat{\mathbf{g}}_n, \widehat{\mathbf{g}}_{n'}) = \sqrt{(\widehat{\mathbf{g}}_n - \widehat{\mathbf{g}}_{n'})^\top \widehat{\mathbf{\Sigma}}_\mathrm{T}^{-1} (\widehat{\mathbf{g}}_n - \widehat{\mathbf{g}}_{n'})}. \tag{6}$$

We show that solution to the optimization problem in Equation (5) can be obtained by eigendecomposition of the matrix $\mathbf{\Sigma}_\mathrm{B} - \mathbf{\Sigma}_\mathrm{W}$. An important property to notice about the objective $\mathcal{J}(\mathbf{\Phi})$ is that it is invariant w.r.t. rescalings $\mathbf{\Phi} \to \alpha\mathbf{\Phi}$. Hence, we can always choose $\mathbf{\Phi} = \mathbf{f}_1 \| \cdots \| \mathbf{f}_{\widehat{D}}$ such that $\mathbf{f}_{\widehat{d}}^\top \mathbf{f}_{\widehat{d}} = 1$, since it is a scalar itself. For this reason we can reduce the problem of maximizing $\mathcal{J}(\mathbf{\Phi})$ into the constrained optimization problem

$$\begin{aligned} \max \quad & \sum_{\widehat{d}=1}^{\widehat{D}} \mathbf{f}_{\widehat{d}}^\top (\mathbf{\Sigma}_\mathrm{B} - \mathbf{\Sigma}_\mathrm{W})\, \mathbf{f}_{\widehat{d}} \\ \text{subject to} \quad & \mathbf{f}_{\widehat{d}}^\top \mathbf{f}_{\widehat{d}} - 1 = 0 \qquad \forall \widehat{d} = 1, \ldots, \widehat{D}. \end{aligned} \tag{7}$$

To solve the above optimization problem, let us consider the Lagrangian

$$\mathcal{L}\left(\mathbf{f}_{\widehat{d}}, \lambda_{\widehat{d}}\right) = \sum_{\widehat{d}=1}^{\widehat{D}} \mathbf{f}_{\widehat{d}}^\top (\mathbf{\Sigma}_\mathrm{B} - \mathbf{\Sigma}_\mathrm{W})\, \mathbf{f}_{\widehat{d}} - \lambda_{\widehat{d}} \left(\mathbf{f}_{\widehat{d}}^\top \mathbf{f}_{\widehat{d}} - 1\right) \tag{8}$$

with multipliers $\lambda_{\widehat{d}}$. To find the maximum, we derive it with respect to $\mathbf{f}_{\widehat{d}}$ and equate to zero

$$\frac{\partial \mathcal{L}\left(\mathbf{f}_{\widehat{d}}, \lambda_{\widehat{d}}\right)}{\partial \mathbf{f}_{\widehat{d}}} = \left((\mathbf{\Sigma}_\mathrm{B} - \mathbf{\Sigma}_\mathrm{W}) - \lambda_{\widehat{d}} \mathbf{I}\right) \mathbf{f}_{\widehat{d}} = 0 \tag{9}$$

which leads to

$$(\mathbf{\Sigma}_\mathrm{B} - \mathbf{\Sigma}_\mathrm{W})\, \mathbf{f}_{\widehat{d}} = \lambda_{\widehat{d}} \mathbf{f}_{\widehat{d}} \tag{10}$$

where $\lambda_{\widehat{d}}$ are the eigenvalues of $\mathbf{\Sigma}_\mathrm{B} - \mathbf{\Sigma}_\mathrm{W}$ and $\mathbf{f}_{\widehat{d}}$ are the corresponding eigenvectors. Putting it all together,

$$(\mathbf{\Sigma}_\mathrm{B} - \mathbf{\Sigma}_\mathrm{W})\, \mathbf{\Phi} = \mathbf{\Lambda}\mathbf{\Phi} \tag{11}$$

where $\mathbf{\Lambda} = \operatorname{diag}\left(\lambda_1, \ldots, \lambda_{\widehat{D}}\right)$ is the eigenvalue matrix. Therefore,

$$\mathcal{J}(\mathbf{\Phi}) = \operatorname{tr}\left(\mathbf{\Phi}^\top (\mathbf{\Sigma}_\mathrm{B} - \mathbf{\Sigma}_\mathrm{W})\, \mathbf{\Phi}\right) = \operatorname{tr}\left(\mathbf{\Phi}^\top \mathbf{\Lambda}\mathbf{\Phi}\right) = \operatorname{tr}(\mathbf{\Lambda}) \tag{12}$$

is maximized when $\mathbf{\Lambda}$ has $\widehat{D}$ largest eigenvalues and $\mathbf{\Phi}$ contains the corresponding leading eigenvectors.

In the following we discuss how to calculate the eigenvectors of $\mathbf{\Sigma}_{\mathrm{B}} - \mathbf{\Sigma}_{\mathrm{W}}$ and to determine an optimal dimensionality $\widehat{D}$ of the feature space. Rewrite $\mathbf{\Sigma}_{\mathrm{B}} - \mathbf{\Sigma}_{\mathrm{W}} = 2\mathbf{\Sigma}_{\mathrm{B}} - \mathbf{\Sigma}_{\mathrm{T}}$. Note that the null space of $\mathbf{\Sigma}_{\mathrm{T}}$ is a subspace of that of $\mathbf{\Sigma}_{\mathrm{B}}$ since the null space of $\mathbf{\Sigma}_{\mathrm{T}}$ is the common null space of $\mathbf{\Sigma}_{\mathrm{B}}$ and $\mathbf{\Sigma}_{\mathrm{W}}$. Thus, we can simultaneously diagonalize $\mathbf{\Sigma}_{\mathrm{B}}$ and $\mathbf{\Sigma}_{\mathrm{T}}$ to some $\mathbf{\Lambda}$ and $\mathbf{I}$

$$\begin{aligned} \mathbf{\Psi}^{\top}\mathbf{\Sigma}_{\mathrm{B}}\mathbf{\Psi} &= \mathbf{\Lambda} \\ \mathbf{\Psi}^{\top}\mathbf{\Sigma}_{\mathrm{T}}\mathbf{\Psi} &= \mathbf{I} \end{aligned} \tag{13}$$

with the $D \times \operatorname{rank}(\mathbf{\Sigma}_{\mathrm{T}})$ eigenvector matrix

$$\mathbf{\Psi} = \mathbf{\Omega}\mathbf{\Theta}^{-\frac{1}{2}}\mathbf{\Xi} \tag{14}$$

where $\mathbf{\Omega}$ and $\mathbf{\Theta}$ are the eigenvector and eigenvalue matrices of $\mathbf{\Sigma}_{\mathrm{T}}$, respectively, and $\mathbf{\Xi}$ is the eigenvector matrix of $\mathbf{\Theta}^{-1/2}\mathbf{\Omega}^{\top}\mathbf{\Sigma}_{\mathrm{B}}\mathbf{\Omega}\mathbf{\Theta}^{-1/2}$. To calculate $\mathbf{\Psi}$, we use a fast two-step algorithm in virtue of Singular Value Decomposition (SVD). SVD expresses a real $r \times s$ matrix $\mathbf{A}$ as a product $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{\top}$ where $\mathbf{D}$ is a diagonal matrix with decreasing non-negative entries, and $\mathbf{U}$ and $\mathbf{V}$ are $r \times \min\{r, s\}$ and $s \times \min\{r, s\}$ eigenvector matrices of $\mathbf{A}\mathbf{A}^{\top}$ and $\mathbf{A}^{\top}\mathbf{A}$, respectively, and the non-vanishing entries of $\mathbf{D}$ are square roots of the non-zero corresponding eigenvalues of both $\mathbf{A}\mathbf{A}^{\top}$ and $\mathbf{A}^{\top}\mathbf{A}$. See that $\mathbf{\Sigma}_{\mathrm{T}}$ and $\mathbf{\Sigma}_{\mathrm{B}}$ can be expressed in the forms

$$\begin{aligned} \mathbf{\Sigma}_{\mathrm{T}} &= \mathbf{X}\mathbf{X}^{\top} \quad \text{where} \quad \mathbf{X} = \frac{1}{\sqrt{N_L}}\left[(\mathbf{g}_1 - \mu)\cdots(\mathbf{g}_{N_L} - \mu)\right] \quad \text{and} \\ \mathbf{\Sigma}_{\mathrm{B}} &= \mathbf{\Upsilon}\mathbf{\Upsilon}^{\top} \quad \text{where} \quad \mathbf{\Upsilon} = \left[(\mu_1 - \mu)\cdots(\mu_{C_L} - \mu)\right], \end{aligned} \tag{15}$$

respectively. Hence, we can obtain the eigenvectors $\mathbf{\Omega}$ and the corresponding eigenvalues $\mathbf{\Theta}$ of $\mathbf{\Sigma}_{\mathrm{T}}$ through the SVD of $\mathbf{X}$ and analogically $\mathbf{\Xi}$ of $\mathbf{\Theta}^{-1/2}\mathbf{\Omega}^{\top}\mathbf{\Sigma}_{\mathrm{B}}\mathbf{\Omega}\mathbf{\Theta}^{-1/2}$ through the SVD of $\mathbf{\Theta}^{-1/2}\mathbf{\Omega}^{\top}\mathbf{\Upsilon}$. The columns of $\mathbf{\Psi}$ are clearly the eigenvectors of $2\mathbf{\Sigma}_{\mathrm{B}} - \mathbf{\Sigma}_{\mathrm{T}}$ with the corresponding eigenvalues $2\mathbf{\Lambda} - \mathbf{I}$. Therefore, to constitute the transform $\mathbf{\Phi}$ by maximizing the MMC, we should choose the eigenvectors in $\mathbf{\Psi}$ that correspond to the eigenvalues of at least $\frac{1}{2}$ in $\mathbf{\Lambda}$. Note that $\mathbf{\Lambda}$ contains at most $\operatorname{rank}(\mathbf{\Sigma}_{\mathrm{B}}) = C_L - 1$ positive eigenvalues, which gives an upper bound on the feature space dimensionality $\widehat{D}$. Algorithm 1 [5] provided below is an efficient way of learning the transform $\mathbf{\Phi}$ for MMC on given labeled learning data $\mathcal{G}_L$.

Algorithm 1 LearnTransformationMatrixMMC($\mathcal{G}_L$)

split $\mathcal{G}_L = \{(\mathbf{g}_n, \ell_n)\}_{n=1}^{N_L}$ into classes $\{\mathcal{I}_c\}_{c=1}^{C_L}$ of $N_c = |\mathcal{I}_c|$ samples
compute overall mean $\mu = \frac{1}{N_L}\sum_{n=1}^{N_L}\mathbf{g}_n$ and individual class means $\mu_c = \frac{1}{N_c}\sum_{n=1}^{N_c}\mathbf{g}_n^{(c)}$
compute $\mathbf{\Sigma}_{\mathrm{B}} = \sum_{c=1}^{C_L}(\mu_c - \mu)(\mu_c - \mu)^{\top}$
compute $\mathbf{X} = \frac{1}{\sqrt{N_L}}\left[(\mathbf{g}_1 - \mu)\cdots(\mathbf{g}_{N_L} - \mu)\right]$
compute $\mathbf{\Upsilon} = \left[(\mu_1 - \mu)\cdots(\mu_{C_L} - \mu)\right]$
compute eigenvectors $\mathbf{\Omega}$ and corresponding eigenvalues $\mathbf{\Theta}$ of $\mathbf{\Sigma}_{\mathrm{T}}$ through SVD of $\mathbf{X}$
compute eigenvectors $\mathbf{\Xi}$ of $\mathbf{\Theta}^{-1/2}\mathbf{\Omega}^{\top}\mathbf{\Sigma}_{\mathrm{B}}\mathbf{\Omega}\mathbf{\Theta}^{-1/2}$ through SVD of $\mathbf{\Theta}^{-1/2}\mathbf{\Omega}^{\top}\mathbf{\Upsilon}$
compute eigenvectors $\mathbf{\Psi} = \mathbf{\Omega}\mathbf{\Theta}^{-1/2}\mathbf{\Xi}$
compute eigenvalues $\mathbf{\Lambda} = \mathbf{\Psi}^{\top}\mathbf{\Sigma}_{\mathrm{B}}\mathbf{\Psi}$
return transform $\mathbf{\Phi}$ as eigenvectors in $\mathbf{\Psi}$ that correspond to the eigenvalues of at least $1/2$ in $\mathbf{\Lambda}$

3 Experiments and Results

3.1 Database

For the evaluation purposes we have extracted a large number of samples from the general MoCap database from CMU [9] as a well-known and recognized database of structural human motion data. It contains numerous motion sequences, including a considerable number of gait sequences. Motions are recorded with an optical marker-based Vicon system. People wear a black jumpsuit and have 41 markers taped on. The tracking space of $30\,\mathrm{m}^2$, surrounded by 12 cameras of sampling rate of 120 Hz in the height from 2 to 4 meters above ground, creates a video surveillance environment. Motion videos are triangulated to get highly accurate 3D data in the form of relative body point coordinates (with respect to the root joint) in each video frame and stored in the standard ASF/AMC data format. Each registered participant is assigned with their respective skeleton described in an ASF file. Motions in the AMC files store bone rotational data, which is interpreted as instructions about how the associated skeleton deforms over time.

These MoCap data, however, contain skeleton parameters pre-calibrated by the CMU staff. Skeletons are unique for each walker and even a trivial skeleton check could result in 100% recognition. In order to use the collected data in a fairly manner, a prototypical skeleton is constructed and used to represent bodies of all subjects, shrouding the unique skeleton parameters of individual walkers. Assuming that all walking subjects are physically identical disables the skeleton check as a potentially unfair classifier. Moreover, this is a skeleton-robust solution as all bone rotational data are linked with a fixed skeleton. To obtain realistic parameters, it is calculated as the mean of all skeletons in the provided ASF files.

We calculate 3D joint coordinates using bone rotational data and the prototypical skeleton. One cannot directly use raw values of joint coordinates, as they refer to absolute positions in the tracking space, and not all potential methods are invariant to person's position or walk direction. To ensure such invariance, the center of the coordinate system is moved to the position of root joint $\gamma_{\mathrm{root}}(t) = [0, 0, 0]^\top$ for each time t and axes are adjusted to the walker's perspective: the X axis is from right (negative) to left (positive), the Y axis is from down (negative) to up (positive), and the Z axis is from back (negative) to front (positive). In the AMC file structure notation it is achieved by zeroing the root translation and rotation (`root 0 0 0 0 0 0`) in all frames of all motion sequences.

Since the general motion database contains all motion types, we extracted a number of sub-motions that represent gait cycles. First, an exemplary gait cycle was identified, and clean gait cycles were then filtered out using the DTW distance over bone rotations. The similarity threshold was set high enough so that even the least similar sub-motion still semantically represents a gait cycle. Finally, subjects that contributed with less than 10 samples were excluded. The final database has 54 walking subjects that performed 3,843 samples in total, which makes an average of about 71 samples per subject.

3.2 Evaluation Setups and Metrics

Learning data $\mathcal{G}_L = \{(\mathbf{g}_n, \ell_n)\}_{n=1}^{N_L}$ of C_L identities and evaluation data $\mathcal{G}_E = \{(\mathbf{g}_n, \ell_n)\}_{n=1}^{N_E}$ of C_E identity classes have to be disjunct at all times. Evaluation is performed exclusively

on the evaluation part, taking no observations of the learning part into account. In the following we introduce two setups of data separation: homogeneous and heterogeneous. The homogeneous setup learns the transformation matrix on $1/3$ samples of C_L identities and is evaluated on templates derived from other $2/3$ samples of the same $C_E = C_L$ identities. The heterogeneous setup learns the transform on all samples in C_L identities and is evaluated on all templates derived from other C_E identities. For better clarification we refer to Figure 2. Note that unlike in homogeneous setup, in heterogeneous setup there is no walker identity ever used for both learning and evaluation at the same time.

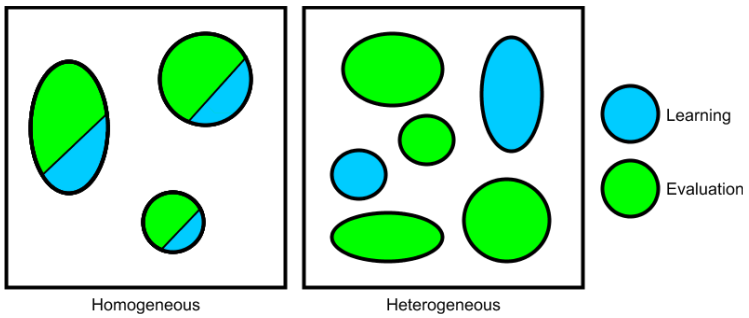

Figure 2. Abstraction of data separation for homogeneous setup of $C_L = C_E = 3$ learning-and-evaluation classes (left) and for heterogeneous setup of $C_L = 2$ learning classes and $C_E = 4$ evaluation classes (right). Black square represents a database and ellipses are identity classes.

Homogeneous setup is parametrized by a single number $C_L = C_E$ of learning-and-evaluation identity classes, whereas the heterogeneous setup has the form (C_L, C_E) specifying how many learning and how many evaluation identity classes are randomly selected from the database. Evaluation of each setup is repeated 3 times, selecting new random C_L and C_E identity classes each time and reporting the average result. Please note that in the heterogeneous setup the learning identities are disjunct from the evaluation identities, that is, there is no single identity used for both learning and evaluation.

Correct Classification Rate (CCR) is a standard qualitative measure; however, if a method has a low CCR, we cannot directly say if the system is failing because of bad features or a bad classifier. Providing an evaluation in terms of class separability of the feature space gives an estimate on the recognition potential of the extracted features and do not reflect eventual combination with an unsuitable classifier. Quality of features extraction algorithms is reflected in the Davies-Bouldin Index (DBI)

$$\mathrm{DBI} = \frac{1}{C_E} \sum_{c=1}^{C_E} \max_{1 \le c' \le C_E, c' \neq c} \frac{\sigma_c + \sigma_{c'}}{\widehat{\delta}(\widehat{\mu}_c, \widehat{\mu}_{c'})} \tag{16}$$

where $\sigma_c = \frac{1}{N_c} \sum_{n=1}^{N_c} \widehat{\delta}(\widehat{\mathbf{g}}_n, \widehat{\mu}_c)$ is the average distance of all templates in identity class $\mathcal{I}_c$ to its centroid, and similarly for $\sigma_{c'}$. Templates of low intra-class distances and of high inter-class distances have a low DBI. DBI is measured on the full evaluation part, whereas CCR is estimated with 10-fold cross-validation taking one dis-labeled fold as a testing set and other nine as gallery. Test templates are classified by the winner-takes-all

strategy, in which a test template $\widehat{\mathbf{g}}^{\text{test}}$ gets assigned with the label $\ell_{\text{argmin}_i \widehat{\delta}\left(\widehat{\mathbf{g}}^{\text{test}}, \widehat{\mathbf{g}}_i^{\text{gallery}}\right)}$ of the gallery's closest identity class.

Based on Section 3.1, our database has 54 identity classes in total. We performed the series of experiments **A**, **B**, **C**, **D** below. The experiments **A** and **B** are to compare the homogeneous and heterogeneous setup, whereas **C** and **D** examine how performance of the system in the heterogeneous setup improves with increasing number of learning identities. The results are illustrated in Figure 3 and in Figure 4 in the next section.

A homogeneous setup with $C_L = C_E \in \{2, \ldots, 27\}$;
B heterogeneous setup with $C_L = C_E \in \{2, \ldots, 27\}$;
C heterogeneous setup with $C_L \in \{2, \ldots, 27\}$ and $C_E = 27$;
D heterogeneous setup with $C_L \in \{2, \ldots, 52\}$ and $C_E = 54 - C_L$.

3.3 Results

Experiments **A** and **B** compare homogeneous and heterogeneous setups by measuring the drop in performance on an identical number of learning and evaluation identities ($C_L = C_E$). Top plot in Figure 3 shows the measured values of DBI and CCR metrics in both alternatives, which not only appear comparable but also in some configurations the heterogeneous setup has an even higher CCR. Bottom plot expresses heterogeneous setup as a percentage of the homogeneous setup in each of the particular metrics. Here we see that with raising number of identities the heterogeneous setup approaches 100% of the fully homogeneous alternative.

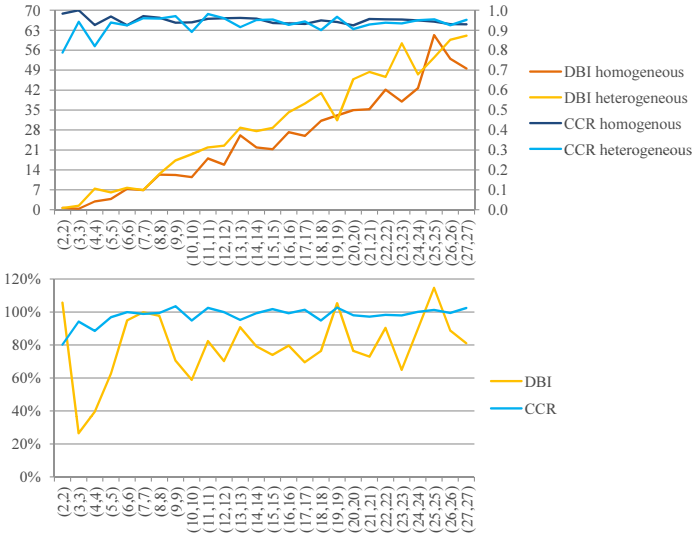

Figure 3. DBI (left vertical axis) and CCR (right vertical axis) for experiments **A** of homogeneous setup and **B** of heterogeneous setup (top) with (C_L, C_E) configurations (horizontal axes) and their percentages (bottom).

Experiments **C** and **D** investigate on the impact of the number of learning identities in the heterogeneous setup. Observing from the Figure 4, the performance grows quickly on the first configurations with very few learning identities, which we can interpret as an analogy to the Pareto (80–20) principle. Specifically, the results of experiment **C** say that 8 learning identities achieve almost the same performance (66.78 DBI and 0.902 CCR) to as if learned on 27 identities (68.32 DBI and 0.947 CCR). The outcome of experiment **D** indicates a similar growth of performance and we see that yet 14 identities can be enough to learn the transformation matrix to distinguish 40 completely different people (0.904 CCR).

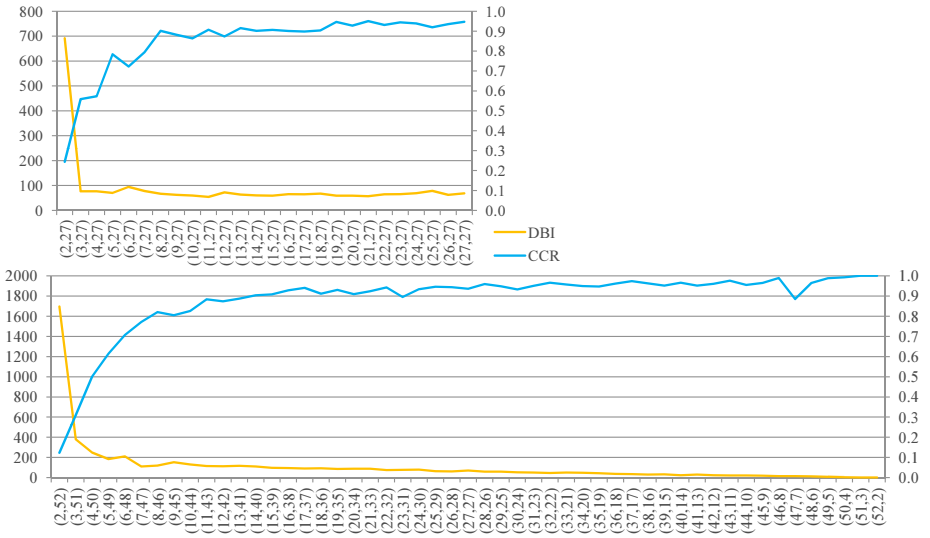

Figure 4. DBI (left vertical axes) and CCR (right vertical axes) for experiments **C** (top) and **D** (bottom) on heterogeneous setup with (C_L, C_E) configurations (horizontal axes).

The proposed method and seven other state-of-the-art methods [2,4,6,10,12,14,16] have been subjected to extensive simulations on homogeneous setup in our recent research paper [5]. A variety of class-separability coefficients and classification metrics allows insights from different statistical perspectives. Results indicate that the proposed method is a leading concept for rank-based classifier systems: lowest Davies-Bouldin Index, highest Dunn Index, highest (and exclusively positive) Silhouette Coefficient, second highest Fisher's Discriminant Ratio and, combined with rank-based classifier, the best Cumulative Match Characteristic, False Accept Rate and False Reject Rate trade-off, Receiver Operating Characteristic (ROC) and recall-precision trade-off scores along with Correct Classification Rate, Equal Error Rate, Area Under ROC Curve and Mean Average Precision. We interpret the high scores as a sign of robustness. Apart from performance merits, the MMC method is also efficient: low-dimensional templates ($\widehat{D} \leq C_L - 1 = C_E - 1 = 53$) and Mahalanobis distance ensure fast distance computations and thus contribute to high scalability.

4 Conclusions

Despite many advanced optimization techniques used in statistical pattern recognition, a common practice of state-of-the-art MoCap-based human identification is still to design geometric gait features by hand. As the first contribution of this paper, the proposed method does not involve any ad-hoc features; on the contrary, they are computed from a much larger space beyond the limits of human interpretability. The features are learned directly from raw joint coordinates by a modification of the Fisher's LDA with MMC so that the identities are maximally separated. We believe that MMC is a suitable criterion for optimizing gait features; however, our future work will continue with research on further potential optimality criterions and machine learning approaches. Furthermore, we are in the process of developing an evaluation framework with implementation details and source codes of all related methods, data extraction drive from the general CMU MoCap database and the evaluation mechanism to support reproducible research.

Second contribution lies in showing the possibility of building a representation on a problem and using it on another (related) problem. Simulations on the CMU MoCap database show that our approach is able to build robust feature spaces without pre-registering and labeling all potential walkers. In fact, we can take different people (experiments **A** and **B**) and just a fraction of them (experiments **C** and **D**). We have observed that with an increasing volume of identities the heterogeneous evaluation setup is on par with the homogeneous setup, that is, it does not matter what identities we learn the features on. One does not have to rely on the availability of all walkers for learning. This is particularly important for a system to aid video surveillance applications where encountered walkers never supply labeled data. Multiple occurrences of individual walkers can now be linked together even without knowing their actual identities.

Acknowledgments Authors thank to the anonymous reviewers for their detailed commentary and suggestions. Data used in this project was created with funding from NSF EIA-0196217 and was obtained from http://mocap.cs.cmu.edu. Our extracted database is available at https://gait.fi.muni.cz/ to support results reproducibility.

References

1. Ahmed, F., Paul, P.P., Gavrilova, M.L.: DTW-Based Kernel and Rank-Level Fusion for 3D Gait Recognition Using Kinect. The Visual Computer 31(6), 915–924 (2015), https://doi.org/10.1007/s00371-015-1092-0
2. Ahmed, M., Al-Jawad, N., Sabir, A.: Gait Recognition Based on Kinect Sensor. Proc. SPIE 9139, 91390B–91390B–10 (2014), http://dx.doi.org/10.1117/12.2052588B
3. Ali, S., Wu, Z., Li, X., Saeed, N., Wang, D., Zhou, M.: Transactions on Computational Science XXVI: Special Issue on Cyberworlds and Cybersecurity, chap. Applying Geometric Function on Sensors 3D Gait Data for Human Identification, pp. 125–141. Springer, Berlin, Heidelberg (2016), https://doi.org/10.1007/978-3-662-49247-5_8
4. Andersson, V.O., Araujo, R.M.: Person Identification Using Anthropometric and Gait Data from Kinect Sensor. In: Proc. of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI-15). pp. 425–431. AAAI Press (2015), http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9680
5. Balazia, M., Sojka, P.: Learning Robust Features for Gait Recognition by Maximum Margin Criterion. In: Proc. of 23rd International Conference on Pattern Recognition, ICPR 2016. pp. 901–906. IEEE (Dec 2016)

6. 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. pp. 225–226. HRI '12, ACM, New York, NY, USA (2012), http://doi.acm.org/10.1145/2157689.2157767
7. Castro, F.M., Marín-Jiménez, M.J., Guil, N., de la Blanca, N.P.: Automatic Learning of Gait Signatures for People Identification. CoRR abs/1603.01006 (2016), https://arxiv.org/abs/1603.01006
8. Choudhury, S.D., Tjahjadi, T.: Robust View-Invariant Multiscale Gait Recognition. Pattern Recognition 48(3), 798–811 (2015), http://www.sciencedirect.com/science/article/pii/S0031320314003835
9. CMU Graphics Lab: Carnegie-Mellon Motion Capture (MoCap) Database (2003), http://mocap.cs.cmu.edu
10. Dikovski, B., Madjarov, G., Gjorgjevikj, D.: Evaluation of Different Feature Sets for Gait Recognition Using Skeletal Data from Kinect. In: 37th Intl. Convention on Information and Communication Technology, Electronics and Microelectronics. pp. 1304–1308 (May 2014), http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6859769
11. Kocsor, A., Kovács, K., Szepesvári, C.: Margin Maximizing Discriminant Analysis, pp. 227–238. Springer Berlin Heidelberg, Berlin, Heidelberg (2004), https://doi.org/10.1007/978-3-540-30115-8_23
12. Kwolek, B., Krzeszowski, T., Michalczuk, A., Josinski, H.: 3D Gait Recognition Using Spatio-Temporal Motion Descriptors. In: Proc. of Intelligent Information and Database Systems: 6th Asian Conference, ACIIDS 2014, Bangkok, Thailand, Part II. LNCS, vol. 8398, pp. 595–604. Springer (Apr 2014), https://doi.org/10.1007/978-3-319-05458-2_61
13. Li, H., Jiang, T., Zhang, K.: Efficient and Robust Feature Extraction by Maximum Margin Criterion. IEEE Transactions on Neural Networks 17(1), 157–165 (Jan 2006), https://doi.org/10.1109/TNN.2005.860852
14. Preis, J., Kessel, M., Werner, M., Linnhoff-Popien, C.: Gait Recognition with Kinect. In: 1st International Workshop on Kinect in Pervasive Computing, New Castle, UK, June 18–22. pp. 1–4 (2012), https://www.researchgate.net/publication/239862819_Gait_Recognition_with_Kinect
15. Ramu Reddy, V., Chakravarty, K., Aniruddha, S.: Person Identification in Natural Static Postures Using Kinect. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) Computer Vision – ECCV 2014 Workshops: Zurich, Switzerland, September 6–7 and 12, 2014, Proceedings, Part II. LNCS, vol. 8926, pp. 793–808. Springer (2015), https://doi.org/10.1007/978-3-319-16181-5_60
16. Sinha, A., Chakravarty, K., Bhowmick, B.: Person Identification Using Skeleton Information from Kinect. In: ACHI 2013: Proc. of the Sixth Intl. Conf. on Advances in CHI. pp. 101–108 (2013), https://www.thinkmind.org/index.php?view=article&articleid=achi_2013_4_50_20187
17. Tafazzoli, F., Bebis, G., Louis, S.J., Hussain, M.: Genetic Feature Selection for Gait Recognition. J. of Electron. Imaging 24(1), 013036 (Feb 2015), https://doi.org/10.1117/1.JEI.24.1.013036
18. Valcik, J., Sedmidubsky, J., Zezula, P.: Assessing Similarity Models for Human-Motion Retrieval Applications. Computer Animation and Virtual Worlds 27(5), 484–500 (2016), http://dx.doi.org/10.1002/cav.1674
19. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer-Verlag, New York, NY, USA (1995)
20. Zeng, W., Wang, C.: View-Invariant Gait Recognition via Deterministic Learning. In: International Joint Conference on Neural Networks (IJCNN). pp. 3465–3472 (Jul 2014), https://doi.org/10.1109/IJCNN.2014.6889507