

# Gait Recognition from Motion Capture Data

MICHAL BALAZIA and PETR SOJKA, Masaryk University

Gait recognition from motion capture data, as a pattern classification discipline, can be improved by the use of machine learning. This article contributes to the state of the art with a statistical approach for extracting robust gait features directly from raw data by a modification of Linear Discriminant Analysis with Maximum Margin Criterion. Experiments on the CMU MoCap database show that the suggested method outperforms 13 relevant methods based on geometric features and a method to learn the features by a combination of Principal Component Analysis and Linear Discriminant Analysis. The methods are evaluated in terms of the distribution of biometric templates in respective feature spaces expressed in a number of class separability coefficients and classification metrics. Results also indicate a high portability of learned features, what means that we can learn what aspects of walk people generally differ in and extract those as general gait features. Recognizing people without needing group-specific features is convenient, as particular people might not always provide annotated learning data. As a contribution to reproducible research, our evaluation framework and database have been made publicly available. This research makes motion capture technology directly applicable for human recognition.

CCS Concepts: • **Security and privacy** → **Biometrics**; • **Computing methodologies** → *Artificial intelligence*; *Machine learning*; • **Applied computing** → *Surveillance mechanisms*;

Additional Key Words and Phrases: Gait recognition, MoCap, maximal margin criterion

## ACM Reference format:

Michal Balazia and Petr Sojka. 2018. Gait Recognition from Motion Capture Data. *ACM Trans. Multimedia Comput. Commun. Appl.* 14, 1s, Article 22 (February 2018), 18 pages.  
<https://doi.org/10.1145/3152124>

## 1 INTRODUCTION

From the surveillance perspective, gait pattern biometrics is appealing because it can be performed at a distance and without body-invasive equipment or the need for the subject's cooperation. This allows data acquisition without a subject's consent. As the data are collected with a high participation rate and the subjects are not expected to claim their identities, the trait is employed for identification rather than for authentication.

Motion capture technology acquires video clips of individuals and generates structured motion data. The format maintains an overall structure of the human body and holds estimated three-dimensional (3D) positions of the main anatomical landmarks as the person moves. These so-called motion capture data (MoCap) can be collected online by depth sensors (RGB-D) such as Microsoft Kinect, Asus Xtion, or Vicon. To visualize motion capture data (see Figure 1), a simplified stick figure representing the human skeleton (a graph of joints connected by bones) can be

Authors' addresses: M. Balazia and P. Sojka, Masaryk University, Faculty of Informatics, Botanická 68A, 602 00 Brno, Czech Republic.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2018 ACM 1551-6857/2018/02-ART22 \$15.00

<https://doi.org/10.1145/3152124>

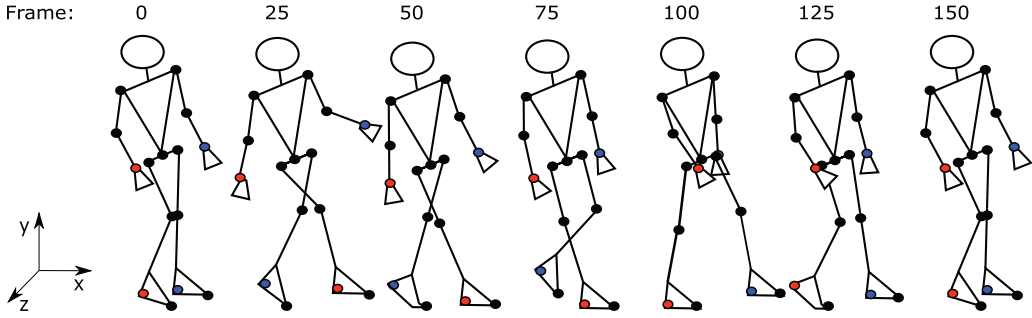


Fig. 1. Motion capture data. Skeleton is represented by a stick figure of 31 joints (only 15 are drawn here). Seven selected video frames of a walk sequence contain 3D spatial coordinates of each joint in time.

automatically recovered from the values of body point spatial coordinates. With recent rapid improvements in MoCap sensor accuracy, we foresee an affordable MoCap technology [21] that can identify people from MoCap data.

The goal of this work is to present a method for learning robust gait features from raw MoCap data. A collection of extracted features builds a gait template that serves as the walker's signature. Templates are stored in a central database. Recognition of a person involves capturing their walk sample, extracting gait features to compose a template, and, finally, querying this database for a set of similar templates to report the most likely identity. The similarity of the two templates is expressed in a single number computed by a similarity/distance function.

Related work is outlined in Section 2, and our methods are described in detail in Section 3. We provide a thorough evaluation of a number of competitive MoCap-based gait recognition methods on a benchmark database and framework described in Section 4.1. Two setups of data separation into learning and evaluation parts are presented in Section 4.2, together with several metrics used for thorough evaluation. Results are presented and discussed in Section 4.3.

## 2 RELATED WORK

Human gait long has been an active subject of study in physical medicine (detection of gait abnormalities [13, 18] and disorders caused by stroke or cerebral palsy [25]), sport (gait regulation [41]), or sociology (age and gender classification [33] or attractiveness evaluation [22]). Inspired by medical studies of Murray [34] and the psychological research of Johansson [24], Cutting and Kozlowski [17] conducted experiments about how participants are able to recognize pedestrians from simply observing the 2D gait pattern generated by light bulbs attached to several joints over their body. These experiments proved that the gait is personally unique and can be potentially used for biometric recognition. The main challenges from the perspective of biometric security [11] are (1) to detect gait cycles in long video footage, (2) to recognize registered participants by their biometric samples, and (3) to retrieve relevant biometric samples from large databases.

Many research groups investigate the discrimination power of gait pattern and develop models of human walk for applications in the automatic recognition from MoCap data of walking people. A number of MoCap-based gait recognition methods have been introduced in the past few years, and new ones continue to emerge. To move forward with the wealth and scope of competitive research, it is necessary to compare the innovative approaches with the state of the art and evaluate them against established evaluation metrics on a benchmark database. New frameworks and databases have been developed recently [14, 26].

Over the past few years, most of the introduced gait features are on a geometric basis. They typically combine static body parameters (bone lengths, person's height) with dynamic gait

features such as step length, walk speed, joint angles and inter-joint distances, along with various statistics (mean, standard deviation, or local/global extremes) of their signals. We are particularly focused on the dynamic parameters. By our definition, gait is a dynamic behavioral trait. Static body parameters are not associated with gait; however, they can be used as another biometric.

What follows is a detailed specification of the 13 gait features extraction methods that we have reviewed in our work to date. Since the idea behind each method has some potential, we have implemented each of them for direct comparison.

- **AhmedF** by Ahmed et al. [1] chooses 20 joint relative distance signals and 16 joint relative angle signals across the whole body, compared using Dynamic Time Warping (DTW).
- **AhmedM** by Ahmed et al. [2] extracts the mean, standard deviation, and skew during one gait cycle of horizontal distances (projected on the Z axis) among feet, knees, wrists, and shoulders; mean and standard deviation during one gait cycle of vertical distances (Y coordinates) of head, wrists, shoulders, knees, and feet; and, finally, the mean area during one gait cycle of the triangle of root and two feet.
- **AliS** by Ali et al. [3] measures the mean areas during one gait cycle of lower limb triangles.
- **AnderssonVO** by Andersson and Araujo [4] calculates gait attributes as mean and standard deviation during one gait cycle of local extremes of the signals of lower body angles, step length as a maximum of feet distance, stride length as a length of two steps, cycle time and velocity as a ratio of stride length, and cycle time. In addition, they extract the mean and standard deviation during one gait cycle of each bone length, and height as the sum of the bone lengths between head and root plus the averages of the bone lengths between root and both feet.
- **BallA** by Ball et al. [9] measures mean, standard deviation, and maximum during one gait cycle of lower limb angle pairs as follows: upper leg relative to the Y axis, lower leg relative to the upper leg, and the foot relative to the Z axis.
- **DikovskiB** by Dikovski et al. [19] selects the mean during one gait cycle of step length, height, and all bone lengths; the mean, standard deviation, minimum, maximum, and mean difference of subsequent frames during one gait cycle of all major joint angles; and the angle between the lines of the shoulder joints and the hip joints.
- **JiangS** by Jiang et al. [23] measures angle signals between the Y axis and four major lower body (thigh and calf) bones. The signals are compared using DTW.
- **KrzeszowskiT** by Krzeszowski et al. [28] observes the signals of rotations of eight major bones (humerus, ulna, thigh, and calf) around all three axes, the person's height, and step length. These signals are compared using the DTW distance function.
- **KwolekB** by Kwolek et al. [30] processes signals of bone angles around all axes, the person's height, and step length. The gait cycles are normalized to 30 frames.
- **NareshKumarMS** by Naresh Kumar and Venkatesh Babu [29] is an interesting approach that extracts all joint trajectories around all three axes and compares gait templates by a distance function of their covariance matrices.
- **PreisJ** by Preis et al. [35] takes height, length of legs, torso, both lower legs, both thighs, both upper arms, both forearms, step length, and speed.
- **SedmidubskyJ** by Sedmidubsky et al. [36] concludes that only the two shoulder-hand signals are discriminatory enough to be used for recognition. These temporal data are compared using the DTW distance function.
- **SinhaA** by Sinha et al. [37] combines all features of BallA and PreisJ with mean areas during one gait cycle of upper body and lower body and then mean, standard deviation, and

maximum distances during one gait cycle between the centroid of the upper body polygon and the centroids of four limb polygons.

The aforementioned features are schematic and human-interpretable, which is convenient for visualizations and for intuitive understanding of the models but unnecessary for automatic gait recognition. Instead, to refrain from ad-hoc schemes and to explore beyond the limits of human interpretability, we prefer learning gait features that maximally separate the identity classes in the feature space. The features calculated by statistical observation of large amounts of data are expected to have a much higher discriminative potential, which will be the subject of experimental evaluation in Section 4.

Methods for 2D gait recognition extensively use machine-learning models for extracting gait features, such as principal component analysis and multi-scale shape analysis [15], genetic algorithms and kernel principal component analysis [39], radial basis function neural networks [42], or convolutional neural networks [12]. All of these and many other models can be reasonably utilized in 3D gait recognition as well. The following section provides a scheme for learning the features directly from raw data by (i) a modification of Fisher's Linear Discriminant Analysis with Maximum Margin Criterion and (ii) a combination of Principal Component Analysis and Linear Discriminant Analysis.

### 3 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 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  $N_L$  learning samples of  $C_L$  walkers be linearly normalized to their average length  $T$ . Labeled learning data in a sample (measurement) space have the form  $\mathcal{G}_L = \{(\mathbf{g}_n, \ell_n)\}_{n=1}^{N_L}$ , where

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

is a gait sample (one gait cycle) in which  $\gamma_j(t) \in \mathbb{R}^3$  are 3D spatial coordinates of joint  $j \in \{1, \dots, J\}$  at time  $t \in \{1, \dots, T\}$  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  $\{\mathcal{I}_c\}_{c=1}^{C_L}$  labeled by  $\ell_n$ . A class  $\mathcal{I}_c \subseteq \mathcal{G}_L$  has  $N_c$  samples. The classes are complete and mutually exclusive. We say that 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.,  $(\mathbf{g}_n, \ell_n), (\mathbf{g}_{n'}, \ell_{n'}) \in \mathcal{I}_c \Leftrightarrow \ell_n = \ell_{n'}$ . For the whole labeled data, we denote the between- and within-class and total scatter matrices

$$\begin{aligned} \Sigma_B &= \sum_{c=1}^{C_L} (\mu_c - \mu) (\mu_c - \mu)^\top \\ \Sigma_W &= \sum_{c=1}^{C_L} \frac{1}{N_c} \sum_{n=1}^{N_c} (\mathbf{g}_n^{(c)} - \mu_c) (\mathbf{g}_n^{(c)} - \mu_c)^\top \\ \Sigma_T &= \sum_{c=1}^{C_L} \frac{1}{N_c} \sum_{n=1}^{N_c} (\mathbf{g}_n^{(c)} - \mu) (\mathbf{g}_n^{(c)} - \mu)^\top = \Sigma_B + \Sigma_W, \end{aligned} \quad (2)$$

where  $\mathbf{g}_n^{(c)}$  denotes the  $n$ th sample in class  $\mathcal{I}_c$  and  $\mu_c$  and  $\mu$  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  $\mu = \frac{1}{N_L} \sum_{n=1}^{N_L} \mathbf{g}_n$ . Margin is defined as

the Euclidean distance of class means minus both individual variances (traces of scatter matrices)

$$\Sigma_c = \frac{1}{N_c} \sum_{n=1}^{N_c} (\mathbf{g}_n^{(c)} - \mu_c) (\mathbf{g}_n^{(c)} - \mu_c)^\top. \quad (3)$$

We measure class separability of a given feature space by a representation of the Maximum Margin Criterion (MMC) [27, 31] used by the Vapnik's Support Vector Machines (SVM) [40] which maximizes the sum of  $\frac{1}{2}C_L(C_L - 1)$  between-class margins

$$\begin{aligned} \mathcal{J} &= \frac{1}{2} \sum_{c, c'=1}^{C_L} ((\mu_c - \mu_{c'})^\top (\mu_c - \mu_{c'}) - \text{tr}(\Sigma_c + \Sigma_{c'})) \\ &= \frac{1}{2} \sum_{c, c'=1}^{C_L} (\mu_c - \mu_{c'})^\top (\mu_c - \mu_{c'}) - \frac{1}{2} \sum_{c, c'=1}^{C_L} \text{tr}(\Sigma_c + \Sigma_{c'}) \\ &= \frac{1}{2} \sum_{c, c'=1}^{C_L} (\mu_c - \mu + \mu - \mu_{c'})^\top (\mu_c - \mu + \mu - \mu_{c'}) - \sum_{c=1}^{C_L} \text{tr}(\Sigma_c) \\ &= \text{tr}\left(\sum_{c=1}^{C_L} (\mu_c - \mu) (\mu_c - \mu)^\top\right) - \text{tr}\left(\sum_{c=1}^{C_L} \Sigma_c\right) \\ &= \text{tr}(\Sigma_B) - \text{tr}(\Sigma_W) = \text{tr}(\Sigma_B - \Sigma_W). \end{aligned} \quad (4)$$

Since  $\text{tr}(\Sigma_B)$  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  $\text{tr}(\Sigma_W)$  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 sample 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  $\Phi \in \mathbb{R}^{D \times \widehat{D}}$  from a  $D$ -dimensional sample 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  $\widehat{\mathbf{g}}_n = \Phi^\top \mathbf{g}_n$ . The objective is to learn a transform  $\Phi$  that maximizes MMC in the feature space

$$\mathcal{J}(\Phi) = \text{tr}(\Phi^\top (\Sigma_B - \Sigma_W) \Phi). \quad (5)$$

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

$$\widehat{\delta}(\widehat{\mathbf{g}}_n, \widehat{\mathbf{g}}_{n'}) = \sqrt{(\widehat{\mathbf{g}}_n - \widehat{\mathbf{g}}_{n'})^\top \widehat{\Sigma}_T^{-1} (\widehat{\mathbf{g}}_n - \widehat{\mathbf{g}}_{n'})}. \quad (6)$$

We show that a solution to the optimization problem in Equation (5) can be obtained by eigen-decomposition of the matrix  $\Sigma_B - \Sigma_W$ . An important property to notice about the objective  $\mathcal{J}(\Phi)$  is that it is invariant w.r.t. rescalings  $\Phi \rightarrow \alpha \Phi$ . Hence, we can always choose  $\Phi = \mathbf{f}_1 \parallel \dots \parallel \mathbf{f}_{\widehat{D}}$  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

$\mathcal{J}(\Phi)$  into the constrained optimization problem

$$\begin{aligned} \max \quad & \sum_{\hat{d}=1}^{\hat{D}} \mathbf{f}_{\hat{d}}^{\top} (\Sigma_B - \Sigma_W) \mathbf{f}_{\hat{d}} \\ \text{subject to} \quad & \mathbf{f}_{\hat{d}}^{\top} \mathbf{f}_{\hat{d}} - 1 = 0 \quad \forall \hat{d} \in \{1, \dots, \hat{D}\}. \end{aligned} \quad (7)$$

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

$$\mathcal{L}(\mathbf{f}_{\hat{d}}, \lambda_{\hat{d}}) = \sum_{\hat{d}=1}^{\hat{D}} \mathbf{f}_{\hat{d}}^{\top} (\Sigma_B - \Sigma_W) \mathbf{f}_{\hat{d}} - \lambda_{\hat{d}} (\mathbf{f}_{\hat{d}}^{\top} \mathbf{f}_{\hat{d}} - 1) \quad (8)$$

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

$$\frac{\partial \mathcal{L}(\mathbf{f}_{\hat{d}}, \lambda_{\hat{d}})}{\partial \mathbf{f}_{\hat{d}}} = ((\Sigma_B - \Sigma_W) - \lambda_{\hat{d}} \mathbf{I}) \mathbf{f}_{\hat{d}} = 0, \quad (9)$$

which leads to

$$(\Sigma_B - \Sigma_W) \mathbf{f}_{\hat{d}} = \lambda_{\hat{d}} \mathbf{f}_{\hat{d}}, \quad (10)$$

where  $\lambda_{\hat{d}}$  are the eigenvalues of  $\Sigma_B - \Sigma_W$  and  $\mathbf{f}_{\hat{d}}$  are the corresponding eigenvectors. Putting it all together,

$$(\Sigma_B - \Sigma_W) \Phi = \Lambda \Phi, \quad (11)$$

where  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_{\hat{D}})$  is the eigenvalue matrix. Therefore,

$$\mathcal{J}(\Phi) = \text{tr}(\Phi^{\top} (\Sigma_B - \Sigma_W) \Phi) = \text{tr}(\Phi^{\top} \Lambda \Phi) = \sum_{\hat{d}=1}^{\hat{D}} \mathbf{f}_{\hat{d}}^{\top} \lambda_{\hat{d}} \mathbf{f}_{\hat{d}} = \sum_{\hat{d}=1}^{\hat{D}} \lambda_{\hat{d}} \mathbf{f}_{\hat{d}}^{\top} \mathbf{f}_{\hat{d}} = \sum_{\hat{d}=1}^{\hat{D}} \lambda_{\hat{d}} = \text{tr}(\Lambda) \quad (12)$$

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

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

$$\begin{aligned} \Psi^{\top} \Sigma_B \Psi &= \Lambda \\ \Psi^{\top} \Sigma_T \Psi &= \mathbf{I} \end{aligned} \quad (13)$$

with the  $D \times \text{rank}(\Sigma_T)$  eigenvector matrix

$$\Psi = \Omega \Theta^{-\frac{1}{2}} \Xi, \quad (14)$$

where  $\Omega$  and  $\Theta$  are the eigenvector and corresponding eigenvalue matrices of  $\Sigma_T$ , respectively, and  $\Xi$  is the eigenvector matrix of  $\Theta^{-1/2} \Omega^{\top} \Sigma_B \Omega \Theta^{-1/2}$ . To calculate  $\Psi$ , 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;  $\mathbf{U}$  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  $\Sigma_T$  and  $\Sigma_B$  can have the forms

$$\begin{aligned}\Sigma_T &= \mathbf{X}\mathbf{X}^\top \text{ where } \mathbf{X} = \frac{1}{\sqrt{N_L}} [(\mathbf{g}_1 - \mu) \cdots (\mathbf{g}_{N_L} - \mu)] \text{ and} \\ \Sigma_B &= \mathbf{Y}\mathbf{Y}^\top \text{ where } \mathbf{Y} = [(\mu_1 - \mu) \cdots (\mu_{C_L} - \mu)],\end{aligned}\quad (15)$$

respectively. Hence, we can obtain the eigenvectors  $\Omega$  and the corresponding eigenvalues  $\Theta$  of  $\Sigma_T$  through the SVD of  $\mathbf{X}$  and analogically  $\Xi$  of  $\Theta^{-1/2} \Omega^\top \Sigma_B \Omega \Theta^{-1/2}$  through the SVD of  $\Theta^{-1/2} \Omega^\top \mathbf{Y}$ . The columns of  $\Psi$  are clearly the eigenvectors of  $2\Sigma_B - \Sigma_T$  with the corresponding eigenvalues  $2\Delta - \mathbf{I}$ . Therefore, to constitute the transform  $\Phi$  by maximizing the MMC, we should choose the eigenvectors in  $\Psi$  that correspond to the eigenvalues of at least  $1/2$  in  $\Delta$ . Note that  $\Delta$  contains at most  $\text{rank}(\Sigma_B) = C_L - 1$  positive eigenvalues, which gives an upper bound on the feature space dimensionality  $\hat{D}$ .

We found inspiration in the Fisher Linear Discriminant Analysis (LDA) [20] that uses Fisher's criterion

$$\mathcal{J}(\Phi_{\text{LDA}}) = \text{tr} \left( \frac{\Phi_{\text{LDA}}^\top \Sigma_B \Phi_{\text{LDA}}}{\Phi_{\text{LDA}}^\top \Sigma_W \Phi_{\text{LDA}}} \right). \quad (16)$$

However, since the rank of  $\Sigma_W$  is at most  $N_L - C_L$ , it is a singular (non-invertible) matrix if  $N_L$  is less than  $D + C_L$  or, analogously might be unstable if  $N_L \ll D$ . Small sample size is a substantial difficulty, as it is necessary to calculate  $\Sigma_W^{-1}$ . To alleviate this, the measured data can be first projected to a lower-dimensional space using Principal Component Analysis (PCA), resulting in a two-stage PCA+LDA feature extraction technique [10] originally introduced for face recognition,

$$\begin{aligned}\mathcal{J}(\Phi_{\text{PCA}}) &= \text{tr}(\Phi_{\text{PCA}}^\top \Sigma_T \Phi_{\text{PCA}}) \\ \mathcal{J}(\Phi_{\text{LDA}}) &= \text{tr} \left( \frac{\Phi_{\text{LDA}}^\top \Phi_{\text{PCA}}^\top \Sigma_B \Phi_{\text{PCA}} \Phi_{\text{LDA}}}{\Phi_{\text{LDA}}^\top \Phi_{\text{PCA}}^\top \Sigma_W \Phi_{\text{PCA}} \Phi_{\text{LDA}}} \right),\end{aligned}\quad (17)$$

and the final transform is  $\Phi = \Phi_{\text{PCA}} \Phi_{\text{LDA}}$ . Given that there are  $\bar{D}$  principal components, then regardless of the dimensionality  $D$  there are at least  $\bar{D} + 1$  independent data points. Thus, if the  $\bar{D} \times \bar{D}$  matrix  $\Phi_{\text{PCA}}^\top \Sigma_W \Phi_{\text{PCA}}$  is estimated from  $N_L - C_L$  independent observations and providing the  $C_L \leq \bar{D} \leq N_L - C_L$ , then we can always invert  $\Phi_{\text{PCA}}^\top \Sigma_W \Phi_{\text{PCA}}$  and in this way obtain the LDA estimate. Note that this method is sub-optimal for multi-class problems [32] as PCA keeps at most  $N_L - C_L$  principal components, whereas at least  $N_L - 1$  of them are necessary to not lose information. PCA+LDA in this form has been used for silhouette-based (2D) gait recognition by Su et al. [38] and is included in our experiments with MoCap (3D).

On given labeled learning data  $\mathcal{G}_L$ , Algorithm 1 and Algorithm 2 [5, 6] provided below are efficient ways of learning the transforms  $\Phi$  for MMC and PCA+LDA, respectively.

## 4 EVALUATION

This article provides an extended evaluation. In the following, we describe the evaluation database and framework, setups for data separation into learning and evaluation parts, evaluation metrics, and results with discussion.

### 4.1 Database and Framework

For evaluation purposes we have extracted a large number of gait samples from the MoCap database obtained from the CMU Graphics Lab [16], which is available under the Creative Commons license. It is a well-known and recognized database of structural human motion data and contains a considerable number of gait sequences. Motions are recorded with an optical marker-based Vicon system. People wear a black jumpsuit with 41 markers taped onto it. The tracking space of  $30\text{m}^2$  is surrounded by 12 cameras with a sampling rate of 120Hz at heights ranging from 2 to 4m



**ALGORITHM 1:** LearnTransformationMatrixMMC( $\mathcal{G}_L$ )

- 
- 1: split  $\mathcal{G}_L = \{(\mathbf{g}_n, \ell_n)\}_{n=1}^{N_L}$  into classes  $\{I_c\}_{c=1}^{C_L}$  of  $N_c = |I_c|$  samples
  - 2: compute overall mean  $\mu = \frac{1}{N_L} \sum_{n=1}^{N_L} \mathbf{g}_n$  and individual class means  $\mu_c = \frac{1}{N_c} \sum_{n=1}^{N_c} \mathbf{g}_n^{(c)}$
  - 3: compute  $\Sigma_B = \sum_{c=1}^{C_L} (\mu_c - \mu)(\mu_c - \mu)^\top$
  - 4: compute  $\mathbf{X} = \frac{1}{\sqrt{N_L}} \left[ (\mathbf{g}_1 - \mu) \cdots (\mathbf{g}_{N_L} - \mu) \right]$
  - 5: compute  $\mathbf{Y} = \left[ (\mu_1 - \mu) \cdots (\mu_{C_L} - \mu) \right]$
  - 6: compute eigenvectors  $\Omega$  and corresponding eigenvalues  $\Theta$  of  $\Sigma_T$  through SVD of  $\mathbf{X}$
  - 7: compute eigenvectors  $\Xi$  of  $\Theta^{-1/2} \Omega^\top \Sigma_B \Omega \Theta^{-1/2}$  through SVD of  $\Theta^{-1/2} \Omega^\top \mathbf{Y}$
  - 8: compute eigenvectors  $\Psi = \Omega \Theta^{-1/2} \Xi$
  - 9: compute eigenvalues  $\Delta = \Psi^\top \Sigma_B \Psi$
  - 10: return transform  $\Phi$  as eigenvectors in  $\Psi$  that correspond to the eigenvalues of at least  $1/2$  in  $\Delta$
- 

**ALGORITHM 2:** LearnTransformationMatrixPCALDA( $\mathcal{G}_L$ )

- 
- 1: split  $\mathcal{G}_L = \{(\mathbf{g}_n, \ell_n)\}_{n=1}^{N_L}$  into classes  $\{I_c\}_{c=1}^{C_L}$  of  $N_c = |I_c|$  samples
  - 2: compute  $\Sigma_B = \sum_{c=1}^{C_L} (\mu_c - \mu)(\mu_c - \mu)^\top$
  - 3: compute  $\Sigma_W = \sum_{c=1}^{C_L} \frac{1}{N_c} \sum_{n=1}^{N_c} \left( \mathbf{g}_n^{(c)} - \mu_c \right) \left( \mathbf{g}_n^{(c)} - \mu_c \right)^\top$
  - 4: compute eigenvectors  $\Phi_{PCA}$  of  $\Sigma_T = \Sigma_B + \Sigma_W$  that correspond to  $\bar{D}$  largest eigenvalues (we set  $\bar{D} = C_L$ )
  - 5: compute eigenvectors  $\Phi_{LDA}$  of  $(\Phi_{PCA}^\top \Sigma_W \Phi_{PCA})^{-1} (\Phi_{PCA}^\top \Sigma_B \Phi_{PCA})$
  - 6: return transform  $\Phi = \Phi_{PCA} \Phi_{LDA}$
- 

above ground, thereby creating 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 are 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 to each walker, and even a trivial skeleton check could result in 100% recognition. To fairly use the collected data, a prototypical skeleton is constructed and used to represent bodies of all subjects, shrouding the skeleton parameters. Assuming that all walking individuals are physically identical disables the skeleton check from being a potentially unfair classifier. Moreover, this is a skeleton-robust solution, as all bone rotational data are linked to one specific skeleton. To obtain realistic parameters, it is calculated as the mean of all skeletons in the provided ASF files.

The raw data are in the form of bone rotations or, if combined with the prototypical skeleton, 3D joint coordinates. The bone rotational data are taken from the AMC files without any pre-processing. We calculate the joint coordinates using the 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 a person's position or walk direction. To ensure such invariance, the center of the coordinate system is moved to the position of root joint  $\gamma_{\text{root}}(t) = [0, 0, 0]^\top$  for each time  $t$  and the 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 setting the root translation and rotation to zero (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 a threshold for their DTW distance on bone rotations in time. The distance threshold was explicitly set low enough so that even the least similar sub-motions still semantically represent gait cycles. Setting this threshold higher might also qualify sub-motions that do not resemble gait cycles anymore. Finally, subjects that contributed with fewer than 10 samples were excluded. The final database [8] has 54 walking subjects that performed 3,843 samples in total, resulting in an average of about 71 samples per subject.

As a contribution to reproducible research, we release our database and framework to improve the development, evaluation, and comparison of methods for gait recognition from MoCap data. They are intended also for our fellow researchers and reviewers to reproduce the results of our experiments. Our recent article [7] provides a manual and comments on reproducing the experiments. With this manual, a reader should be able to reproduce the evaluation and to use the implementation in their own application. The source codes and data are available online at our research group web page [8] with link to our departmental Git repository.

The evaluation framework comprises (i) source codes of the state-of-the-art human-interpretable geometric features that we have reviewed in our work to date as well as own our two approaches where gait features are learned by MMC (see Algorithm 1) and by PCA+LDA (see Algorithm 2), the Random method (no features and random classifier), and the Raw method (raw data and DTW distance function). Depending on whether the raw data are in the form of bone rotations or joint coordinates, the methods are referred to with BR or JC subscripts, respectively. The framework includes (ii) a mechanism for evaluating four class separability coefficients of feature space and four classifier performance metrics. It also contains (iii) a tool for learning a custom classifier and for classifying a custom probe on a custom gallery. We provide (iv) an experimental database along with source codes for its extraction from the CMU MoCap database.

## 4.2 Evaluation Setup and Metrics

In the following, we introduce two setups of data separation: homogeneous and heterogeneous. The homogeneous setup learns the transformation matrix on a fraction of samples of  $C_L$  identities and is evaluated on templates derived from the remaining samples of the same  $C_E = C_L$  identities. The heterogeneous setup learns the transform on all samples of  $C_L$  identities and is evaluated on all templates derived from all samples of other  $C_E$  identities. An abstraction of this concept is depicted in Figure 2. Note that in heterogeneous setup no walker identity is ever used for both learning and evaluation at the same time.

A use of the homogeneous setup can be any system to recognize people that cooperated at learning the features. For example, a small company entrance authentication where all employees cooperated to register. MoCap data of each employee's gait pattern can be registered along with a picture of their face. On the other hand, the heterogeneous setup can be used at person re-identification. During video surveillance, new identities can appear on the fly, and labeled data for all the people encountered may not always be available. In this scenario, forensic investigators may ask for tracking the suspects.

The 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  $(C_L, C_E)$ . This parametrization specifies how many learning and how many evaluation identity classes are randomly selected from the database. The evaluation of each setup is repeated 3 times, selecting new random identity classes each time and reporting the average result.

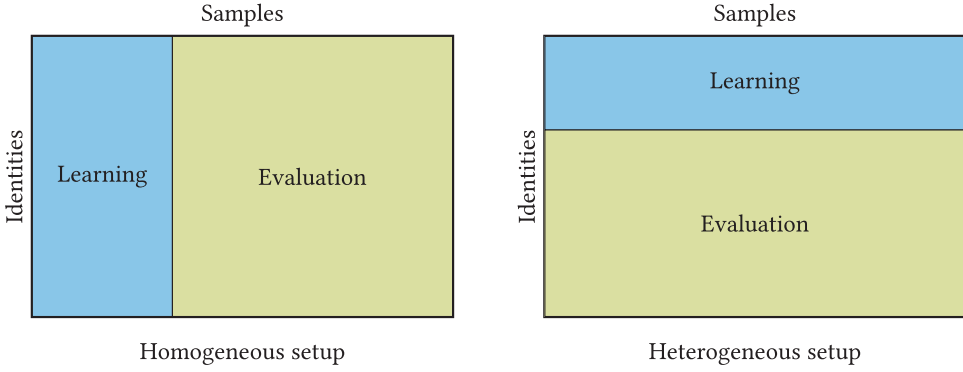


Fig. 2. Abstraction of the concept of homogeneous and heterogeneous setups. Consider a matrix (visualized as a rectangle) of all database samples where each row contains samples of one identity class. The homogeneous setup separates all identities into  $1/3$  learning samples and  $2/3$  evaluation samples (left), whereas the heterogeneous setup takes  $1/3$  learning identities and  $2/3$  evaluation identities with all their samples (right). Setting the sizes of learning and evaluation parts is a subject of setup configuration.

In both setups, the class separability coefficients are calculated directly on the full evaluation part, whereas the classification metrics are estimated with the 10-fold cross-validation taking one dis-labeled fold as a testing set and the other nine labeled folds 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  $\ell_{\arg\min_i \widehat{\delta}(\widehat{\mathbf{g}}^{\text{test}}, \widehat{\mathbf{g}}_i^{\text{gallery}})}$  of the gallery's closest identity class.

What follows is a list of evaluation metrics. Recognition rate is often perceived as the ultimate qualitative measure; however, it is more explanatory to include an evaluation in terms of class separability of the feature space. The class separability measures give an estimate of the recognition potential of the extracted features and do not reflect eventual combination with an unsuitable classifier:

- *Davies-Bouldin Index*: DBI

$$\text{DBI} = \frac{1}{C_E} \sum_{c=1}^{C_E} \max_{1 \leq c' \leq C_E, c' \neq c} \frac{\sigma_c + \sigma_{c'}}{\widehat{\delta}(\widehat{\mu}_c, \widehat{\mu}_{c'})}, \quad (18)$$

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 elements in identity class  $\mathcal{I}_c$  to its centroid and analogically for  $\sigma_{c'}$ . Templates of low intra-class distances and of high inter-class distances have a low DBI.

- *Dunn Index*: DI

$$\text{DI} = \frac{\min_{1 \leq c < c' \leq C_E} \widehat{\delta}(\widehat{\mu}_c, \widehat{\mu}_{c'})}{\max_{1 \leq c \leq C_E} \sigma_c}, \quad (19)$$

with  $\sigma_c$  from the above DBI. Since this criterion seeks classes with high intra-class similarity and low inter-class similarity, a high DI is more desirable.

- *Silhouette Coefficient*: SC

$$\text{SC} = \frac{1}{N_E} \sum_{n=1}^{N_E} \frac{b(\widehat{\mathbf{g}}_n) - a(\widehat{\mathbf{g}}_n)}{\max \{a(\widehat{\mathbf{g}}_n), b(\widehat{\mathbf{g}}_n)\}}, \quad (20)$$

where  $a(\widehat{\mathbf{g}}_n) = \frac{1}{N_c} \sum_{n'=1}^{N_c} \widehat{\delta}(\widehat{\mathbf{g}}_n, \widehat{\mathbf{g}}_{n'})$  is the average distance from  $\widehat{\mathbf{g}}_n$  to other samples within the same identity class and  $b(\widehat{\mathbf{g}}_n) = \min_{1 \leq c' \leq C_E, c' \neq c} \frac{1}{N_{c'}} \sum_{n'=1}^{N_{c'}} \widehat{\delta}(\widehat{\mathbf{g}}_n, \widehat{\mathbf{g}}_{n'})$  is the average distance of  $\widehat{\mathbf{g}}_n$  to the samples in the closest class. It is clear that  $-1 \leq \text{SC} \leq 1$  and a SC close to one mean that classes are appropriately separated.

- *Fisher's Discriminant Ratio*: FDR

$$\text{FDR} = \frac{\frac{1}{C_E} \sum_{c=1}^{C_E} \widehat{\delta}(\widehat{\mu}_c, \widehat{\mu})}{\frac{1}{N_E} \sum_{c=1}^{C_E} \sum_{n=1}^{N_c} \widehat{\delta}(\widehat{\mathbf{g}}_n, \widehat{\mu}_c)}. \quad (21)$$

A high FDR is preferred for seeking classes with low intra-class sparsity and high inter-class sparsity.

Apart from analyzing the distribution of templates in the feature space, it is schematic to combine the features with a rank-based classifier and to evaluate the system based on distance distribution with respect to a query. For obtaining a more applied performance evaluation, we evaluate the following:

- *Cumulative Match Characteristic*: CMC  
Sequence of Rank- $k$  (for  $k$  on X axis from 1 up to  $C_E$ ) recognition rates (Y axis) to measure ranking capabilities of a recognition method. Its headline Rank-1 is the Correct Classification Rate (CCR).
- *False Accept Rate vs. False Reject Rate*: FAR/FRR  
Two sequences of the error rates (Y axis) as functions of the discrimination threshold (X axis). Each method has a value  $e$  of this threshold giving Equal Error Rate (EER=FAR=FRR).
- *Receiver Operating Characteristic*: ROC  
Sequence of True Accept Rate (TAR) and False Accept Rate with a varied discrimination threshold. For a given threshold, the system signalizes both TAR (Y axis) and FAR (X axis). The value of Area Under Curve (AUC) is computed as the integral of the ROC curve.
- *Recall vs. Precision*: RCL/PCN  
Sequence of the rates with a varied discrimination threshold. For a given threshold the system signals both RCL (X axis) and PCN (Y axis). The value of Mean Average Precision (MAP) is computed as the area under the RCL/PCN curve.

These measures reflect how class-separated the feature space is, how often the walk pattern of a person is classified correctly and how difficult it is to confuse different people. They do not, in fact, provide complementary information, although a quality evaluation framework should be able to evaluate the most popular measures. Different applications use different evaluation measures. For example, a hotel lobby authentication system could use a high Rank-3 at the CMC, while a city-level person tracking system is likely to need the ROC curve leaning towards the upper left corner.

Finally, the evaluation incorporates two scalability measures: average distance computation time (DCT) in milliseconds and average template dimensionality (TD).

### 4.3 Results

In this section, we provide comparative evaluation results of the feature extraction methods in terms of evaluation metrics defined in Section 4.2. To ensure a fair comparison, we evaluate all methods on the same experimental database and framework described in Section 4.1. Table 1

Table 1. Class Separability Coefficients, Classification Metrics, and Scalability Metrics for All Implemented Methods

Method	Implementation		Class separability coefficients				Classification metrics				Scalability		Evaluation total time
	gait features	dist. function	DBI	DI	SC	FDR	CCR	EER	AUC	MAP	DCT	TD	
AhmedF	JRD + JRA	DTW+ $L_1$	185.8	1.708	-0.164	0.77	0.891	0.374	0.677	0.254	45	5,254	10.3d
AhmedM	HDF + VDF	$L_1$	216.2	0.842	-0.246	0.954	0.657	0.38	0.659	0.165	<1	24	48.6m
	triangle areas	$L_1$	501.5	0.26	-0.463	1.175	0.225	0.384	0.679	0.111	<1	2	40.9m
AnderssonVO	all described	$L_1$	142.3	1.297	-0.102	1.127	0.84	0.343	0.715	0.251	<1	68	45.7m
BallA	all described	$L_1$	161	1.458	-0.163	1.117	0.75	0.346	0.711	0.231	<1	18	48.5m
Dikovskib	Dataset 3	$L_1$	144.5	1.817	-0.135	<b>1.227</b>	0.881	0.363	0.695	0.254	<1	71	50.7m
JiangS	angles	DTW+ $L_1$	206.6	1.802	-0.249	0.85	0.811	0.395	0.657	0.242	8	584	1.9d
KrzeszowskiT	all described	DTW+ $L_1$	154.1	1.982	-0.147	0.874	0.915	0.392	0.662	0.275	35	3,795	8.1d
KwolekB	g_all	$L_1$	150.9	1.348	-0.084	1.175	0.896	0.358	0.723	0.323	<1	660	1.1h
NareshKumarMS	normalized JC	cov. matrices	<b>118.6</b>	1.618	-0.086	1.09	0.801	0.459	0.631	0.217	8	13,950	1.8d
PreisJ	static + dynamic	$L_1$	1,980.6	0.055	-0.512	1.067	0.143	0.401	0.626	0.067	<1	13	48.7m
SedmidubskyJ	$\mathcal{S}_{C_L H_L} + \mathcal{S}_{C_R H_R}$	DTW+ $L_1$	398.1	1.35	-0.425	0.811	0.543	0.388	0.657	0.149	<1	292	1.4d
SinhaA	all described	$L_1$	214.8	1.112	-0.215	1.101	0.674	0.356	0.697	0.191	<1	45	49.6m
MMC <sub>BR</sub>	MMC on BR	Mahalanobis	154.2	1.638	<b>0.062</b>	1.173	0.925	<b>0.297</b>	<b>0.748</b>	0.353	<1	53	2.6h
MMC <sub>JC</sub>	MMC on JC	Mahalanobis	130.3	1.891	0.051	1.106	0.918	0.378	0.721	0.315	<1	51	3.0h
PCA+LDA <sub>BR</sub>	PCA+LDA on BR	Mahalanobis	182	1.596	-0.015	0.984	0.918	0.361	0.695	0.276	<1	54	4.7h
PCA+LDA <sub>JC</sub>	PCA+LDA on JC	Mahalanobis	174.4	1.309	-0.091	0.827	0.863	0.44	0.643	0.201	<1	54	10.9h
Random	none						0.042					<b>0</b>	27.9m
Raw <sub>BR</sub>	BR	DTW+ $L_2$	163.7	<b>2.092</b>	0.011	0.948	<b>0.966</b>	0.315	0.743	<b>0.358</b>	70	8,229	16.1d
Raw <sub>JC</sub>	JC	DTW+ $L_2$	155.1	1.954	-0.12	0.897	0.926	0.377	0.679	0.283	161	13,574	36.7d

Homogeneous setup, evaluated on the full database with nested cross-validation. We advise readers to read the original articles for better understanding of the implementation column. The last column contains the total evaluation times (seconds s, minutes m, hours h, days d).

presents the implementation details of all the methods and the results of class separability coefficients, classification metrics, and scalability.

The goal of the MMC-based learning is to find a linear discriminant that maximizes the misclassification margin. This optimization technique appears to be more effective than designing geometric gait features. A variety of class-separability coefficients and classification metrics allows insights from different statistical perspectives. The results in Table 1 indicate that the proposed MMCB method (on bone rotational data) is a leading concept for rank-based classifier systems: highest SC and AUC, lowest EER, and competitive DBI, DI, FDR, CCR, and MAP. In terms of recognition rate, the MMC method was only outperformed by the Raw method, which is implemented here as a form of baseline. We interpret the high scores as a sign of robustness.

Apart from the performance merits, the MMC method is also efficient: relatively low-dimensional templates and Mahalanobis distance ensure fast distance computations and thus contribute to high scalability. Note that even if the Raw method has some of the best results, it can hardly be used in practice due to its extreme consumption of time and space resources. On the other hand, Random has no features but cannot be considered a serious recognition method. To illustrate the evaluation time, calculating the distance matrix (a matrix of distances between all evaluation templates) took a couple minutes for the MMC method, almost nothing for the Random method, and more than two weeks for the Raw method. The learning time of the MMC and PCA+LDA methods increases with the number of learning samples  $N_L$  (counting  $\Sigma_W$  and  $\Sigma_T$  matrices); however, this is not an issue in the field of gait recognition as training the models suffers from the opposite problem, undersampling.

To reproduce the experiments in Table 1, instructions to follow are to be found at Reference [8]. Please note that some methods are rather slow—total evaluation times (learning included) in the last column were measured on a computer with Intel Xeon CPU E5-2650 v2 @ 2.60GHz and 256GB RAM.

An additional experiment was carried out with the traditional 10-fold cross-validation in which gait cycles of a common walk sequence are always kept together in the same fold. This is to prevent a situation in which two consecutive gait cycles are split between testing and training folds which would cause a potential overtraining. The classification metrics of methods include CCR, EER, AUC, MAP, and the CMC curve up to Rank-10. Displayed in Table 2, the results suggest that the methods based on machine learning outperform the methods of hand-designed features in terms of all measured metrics. Note that the results are slightly worse than those in Table 1 due avoidance of the overtraining cases and that the order of the tested methods is preserved. We interpret this order as a demonstrative result of comparison of all tested methods based on their classification potential.

Finally, the experiments **A**, **B**, **C**, and **D** described below compare the homogeneous and heterogeneous setup and examine how quality of the system in the heterogeneous setup improves with an increased number of learning identities. The main idea is to show that we can learn what aspects of walk people generally differ in and extract those as general gait features. Recognizing people without needing group-specific features is convenient as particular people might not always provide annotated learning data. Results are illustrated in Figure 3 and in Figure 4,

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

Experiments **A** and **B** compare homogeneous and heterogeneous setups by measuring the drop in the quality measures on an identical number of learning and evaluation identities ( $C_L = C_E$ ).

Table 2. Classification Metrics for All Implemented Methods

method	CCR	EER	AUC	MAP	CMC
Raw <sub>JC</sub>	0.872	0.321	0.731	0.317	1.00
MMC <sub>BR</sub>	0.868	0.305	0.739	0.332	0.95
Raw <sub>BR</sub>	0.867	0.333	0.701	0.259	0.90
MMC <sub>JC</sub>	0.861	0.325	0.72	0.309	0.85
PCA+LDA <sub>BR</sub>	0.845	0.335	0.682	0.247	0.80
KwolekB	0.823	0.367	0.711	0.296	0.75
KrzyszowskiT	0.802	0.348	0.717	0.273	0.70
PCA+LDA <sub>JC</sub>	0.79	0.417	0.634	0.189	0.65
DikovskiB	0.787	0.376	0.679	0.227	0.60
AhmedF	0.771	0.371	0.664	0.22	0.55
AnderssonVO	0.76	0.352	0.703	0.228	0.50
NareshKumarMS	0.717	0.459	0.613	0.19	0.45
JiangS	0.692	0.407	0.637	0.204	0.40
Balla	0.667	0.356	0.698	0.207	0.35
SinhaA	0.598	0.362	0.69	0.176	0.30
AhmedM	0.58	0.392	0.646	0.145	0.25
SedmidubskyJ	0.464	0.394	0.65	0.138	0.20
AliS	0.186	0.394	0.662	0.096	0.15
PreisJ	0.131	0.407	0.618	0.066	0.10
Random	0.039				0.05

Homogeneous setup, evaluated on the full database with 10-fold cross-validation. Methods are ordered by their CCR score.

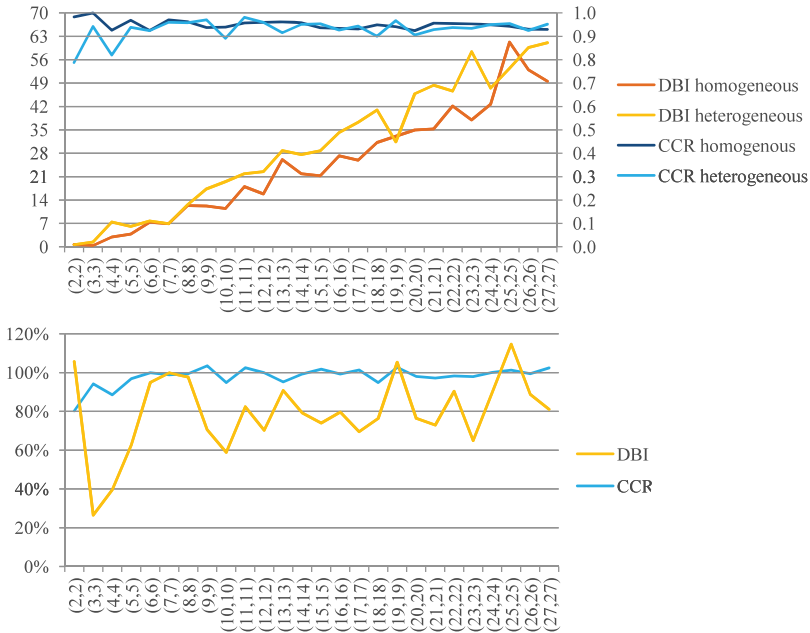


Fig. 3. DBI (left vertical axis) and CCR (right vertical axis) for experiments **A** of homogeneous setup and **B** of heterogeneous setup (top) with  $(C_L, C_E)$  configurations (horizontal axes) and their percentages (bottom).

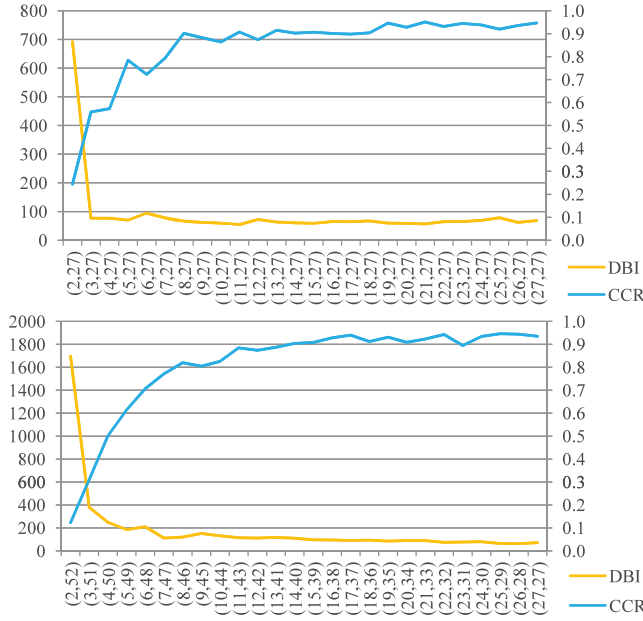


Fig. 4. DBI (left vertical axes) and CCR (right vertical axes) for experiments **C** (top) and **D** (bottom) on heterogeneous setup with  $(C_L, C_E)$  configurations (horizontal axes).

Please note that based on Section 4.1 our database has 54 identity classes in total, and they can be split into the learning and evaluation parts in the numbers of at most (27, 27). The top plot in Figure 3 shows the 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. The bottom plot expresses heterogeneous setup as a percentage of the homogeneous setup in each of the particular metrics. Here we see that with raising the number of identities the heterogeneous setup approaches 100% of the homogeneous setup. Therefore, the gait features learned by the introduced MMC method are walker independent, which means that they can be learned on anybody.

Experiments **C** and **D** investigate the impact of the number of learning identities in the heterogeneous setup. Experiment **D** has only been evaluated up until (27, 27), as results with a learning part larger than an evaluation part would be insignificant. As can be seen in 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) as if they had been 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 14 identities can be enough to learn the transformation matrix to distinguish 40 completely different people (0.904 CCR).

## 5 CONCLUSION

A common practice in state-of-the-art MoCap-based human identification is to design geometric gait features by hand. However, the field of pattern recognition has not recently advanced to a point where best results are frequently obtained using a machine-learning approach. Finding optimal features for MoCap-based gait recognition is no exception. This work introduces the concept of



learning robust and discriminative features directly from raw MoCap data by a modification of the Fisher Linear Discriminant Analysis with Maximum Margin Criterion and by combining Principal Component Analysis and Linear Discriminant Analysis, both with the goal of maximal separation of identity classes, avoiding the potential singularity problem caused by undersampling. Instead of instinctively drawing ad-hoc features, these methods are computed from a much larger space beyond the limits of human interpretability. The collection of features learned by MMC achieves leading scores in four class separability coefficients and therefore has a great potential for gait recognition applications. This is demonstrated on our extracted and normalized MoCap database of 54 walkers with 3,843 gait cycles by outperforming other 13 methods in numerous evaluation metrics. MMC is suitable for optimizing gait features; however, our future work will continue with research on further potential optimality criteria and multi-linear machine-learning approaches.

The 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 use just a fraction of them (experiments **C** and **D**). We have observed that already on 5 identities the heterogeneous evaluation setup exceeds 95% of the homogeneous setup and improves with an increasing volume of identities. This means that one does not have to rely on the availability of all walkers for learning; instead, the features can be learned on anybody. The CCR of over 90% when learning on 14 identities and evaluating on other 40 identities shows that an MMC-based system once learned on a fixed set of walker identities can be used to distinguish many more additional walkers. This is particularly important for a system that supports video surveillance applications where the encountered walkers never supply labeled data. Multiple occurrences of individual walkers can now be linked together even without knowing their actual identities.

In the name of reproducible research, our third contribution is the provision of the evaluation framework and database. All our data and source codes have been made available [8] under the Creative Commons Attribution license (CC-BY) for database and the Apache 2.0 license for software, which grant free use and allow for experimental evaluation. We hope that they will contribute to a smooth development and evaluation of further novel MoCap-based gait recognition methods. We encourage all readers and developers of MoCap-based gait recognition methods to contribute to the framework with new algorithms, data, and improvements.

## ACKNOWLEDGMENTS

The authors thank the reviewers for their detailed commentary and suggestions. The data used in this project were created with funding from NSF EIA-0196217 and were obtained from <http://mocap.cs.cmu.edu> [16]. Our extracted database and evaluation framework are available online at <https://gait.fi.muni.cz> to support reproducibility of results.

## REFERENCES

- [1] Faisal Ahmed, Padma Polash Paul, and Marina L. Gavrilova. 2015. DTW-based kernel and rank-level fusion for 3D gait recognition using kinect. *Vis. Comput.* 31, 6 (2015), 915–924. DOI: <http://dx.doi.org/10.1007/s00371-015-1092-0>
- [2] Mohammed Ahmed, Naseer Al-Jawad, and Azhin Sabir. 2014. Gait recognition based on kinect sensor. *Proc. SPIE 9139, Real-Time Image and Video Processing 2014*, 91390B. DOI: <http://dx.doi.org/10.1117/12.2052588>
- [3] Sajid Ali, Zhongke Wu, Xulong Li, Nighat Saeed, Dong Wang, and Mingquan Zhou. 2016. Applying geometric function on sensors 3D gait data for human identification. In *Transactions on Computational Science XXVI: Special Issue on Cyberworlds and Cybersecurity*. Springer, Berlin, 125–141. DOI: [http://dx.doi.org/10.1007/978-3-662-49247-5\\_8](http://dx.doi.org/10.1007/978-3-662-49247-5_8)
- [4] Virginia O. Andersson and Ricardo M. Araujo. 2015. Person identification using anthropometric and gait data from kinect sensor. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI-15)*. AAAI Press, 425–431.

- [5] Michal Balazia and Petr Sojka. 2016. Learning robust features for gait recognition by maximum margin criterion. In *Proceedings of 23rd International Conference on Pattern Recognition (ICPR'16)*. IEEE, 901–906. DOI : <http://dx.doi.org/10.1109/ICPR.2016.7899750>
- [6] Michal Balazia and Petr Sojka. 2016. Walker-independent features for gait recognition from motion capture data. In *Proceedings of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition*, Antonio Robles-Kelly, Marco Loog, Battista Biggio, Francisco Escolano, and Richard Wilson (Eds.). Springer International Publishing, Cham, 310–321. DOI : [http://dx.doi.org/10.1007/978-3-319-49055-7\\_28](http://dx.doi.org/10.1007/978-3-319-49055-7_28)
- [7] Michal Balazia and Petr Sojka. 2017. An evaluation framework and database for mocap-based gait recognition methods. In *Proceedings of the 1st International Workshop on Reproducible Research in Pattern Recognition (RRPR'16)*, Bertrand Kerautret, Miguel Colom, and Pascal Monasse (Eds.). Springer International Publishing, Cham, 33–47. DOI : [http://dx.doi.org/10.1007/978-3-319-56414-2\\_3](http://dx.doi.org/10.1007/978-3-319-56414-2_3)
- [8] Michal Balazia and Petr Sojka. 2017. Gait Recognition from Motion Capture Data. Retrieved May 30, 2017 from <https://gait.fi.muni.cz/>.
- [9] Adrian Ball, David Rye, Fabio Ramos, and Mari Velonaki. 2012. Unsupervised clustering of people from ‘skeleton’ data. In *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI'12)*. ACM, New York, NY, 225–226. DOI : <http://dx.doi.org/10.1145/2157689.2157767>
- [10] Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman. 1997. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* 19, 7 (July 1997), 711–720. DOI : <http://dx.doi.org/10.1109/34.598228>
- [11] N. V. Boulgouris, Konstantinos N. Plataniotis, and Evangelia Micheli-Tzanakou. 2009. *Biometrics: Theory, Methods, and Applications*. Wiley-IEEE Press.
- [12] Francisco M. Castro, Manuel J. Marín-Jiménez, Nicolás Guil, and Nicolas Pérez de la Blanca. 2016. Automatic learning of gait signatures for people identification. In *Advances in Computational Intelligence (IWANN'17)*, I. Rojas, G. Joya, and A. Catala (Eds.). Lecture Notes in Computer Science, vol 10306. Springer, Cham. DOI : [https://doi.org/10.1007/978-3-319-59147-6\\_23](https://doi.org/10.1007/978-3-319-59147-6_23)
- [13] Alexandros Andre Chaaraoui, José Ramón Padilla-López, and Francisco Flórez-Revuelta. 2015. Abnormal gait detection with RGB-D devices using joint motion history features. In *Proceedings of the 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG'15)*, Vol. 7. IEEE, 1–6.
- [14] Xi Chen and Markus Koskela. 2013. Classification of RGB-D and motion capture sequences using extreme learning machine. In *Proceedings of 18th Scandinavian Conference on Image Analysis (SCIA'13)*, Joni-Kristian Kämäräinen and Markus Koskela (Eds.). Springer, 640–651. DOI : [http://dx.doi.org/10.1007/978-3-642-38886-6\\_60](http://dx.doi.org/10.1007/978-3-642-38886-6_60)
- [15] Sruti Das Choudhury and Tardi Tjahjadi. 2015. Robust view-invariant multiscale gait recognition. *Pattern Recogn.* 48, 3 (2015), 798–811. DOI : <http://dx.doi.org/10.1016/j.patcog.2014.09.022>
- [16] CMU Graphics Lab. 2003. Carnegie-Mellon Motion Capture (MoCap) Database. Retrieved from <http://mocap.cs.cmu.edu>.
- [17] James E. Cutting and Lynn T. Kozlowski. 1977. Recognizing friends by their walk: Gait perception without familiarity cues. *Bull. Psychon. Soc.* 9, 5 (1977), 353–356. DOI : <http://dx.doi.org/10.3758/BF03337021>
- [18] Maxime Devanne, Hazem Wannous, Mohamed Daoudi, Stefano Berretti, Alberto Del Bimbo, and Pietro Pala. 2016. Learning shape variations of motion trajectories for gait analysis. In *Proceedings of 23rd International Conference on Pattern Recognition (ICPR'16)*. 895–900. DOI : <http://dx.doi.org/10.1109/ICPR.2016.7899749>
- [19] Bojan Dikovski, Gjorgji Madjarov, and Dejan Gjorgjevikj. 2014. Evaluation of different feature sets for gait recognition using skeletal data from kinect. In *Proceedings of the 37th International Convention on Information and Communication Technology, Electronics and Microelectronics*. 1304–1308. DOI : <http://dx.doi.org/10.1109/MIPRO.2014.6859769>
- [20] Ronald A. Fisher. 1936. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* 7, 2 (1936), 179–188. DOI : <http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x>
- [21] SangUk Han, Madhav Achar, SangHyun Lee, and Feniosky Peña-Mora. 2013. Empirical assessment of a RGB-D sensor on motion capture and action recognition for construction worker monitoring. *Vis. Eng.* 1, 1 (2013), 6. DOI : <http://dx.doi.org/10.1186/2213-7459-1-6>
- [22] Jie Hong, Jinsheng Kang, and Michael E. Price. 2014. Extraction of bodily features for gait recognition and gait attractiveness evaluation. *Multimedia Tools Appl.* 71, 3 (2014), 1999–2013. DOI : <http://dx.doi.org/10.1007/s11042-012-1319-2>
- [23] Shuming Jiang, Yufei Wang, Yuanyuan Zhang, and Jiande Sun. 2015. Real time gait recognition system based on kinect skeleton feature. In *Proceedings of the ACCV 2014 Workshops on Computer Vision*, C. V. Jawahar and Shiguang Shan (Eds.). LNCS, Vol. 9008. Springer, 46–57. DOI : [http://dx.doi.org/10.1007/978-3-319-16628-5\\_4](http://dx.doi.org/10.1007/978-3-319-16628-5_4)
- [24] Gunnar Johansson. 1973. Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* 14, 2 (1973), 201–211. DOI : <http://dx.doi.org/10.3758/BF03212378>
- [25] J. Kamruzzaman and R. K. Begg. 2006. Support vector machines and other pattern recognition approaches to the diagnosis of cerebral palsy gait. *IEEE Trans. Biomed. Eng.* 53, 12 (Dec. 2006), 2479–2490. DOI : <http://dx.doi.org/10.1109/TBME.2006.883697>

- [26] Dimitris Kastaniotis, Ilias Theodorakopoulos, Christos Theoharatos, George Economou, and Spiros Fotopoulos. 2015. A framework for gait-based recognition using kinect. *Pattern Recogn. Lett.* 68, 2 (2015), 327–335. DOI : <http://dx.doi.org/10.1016/j.patrec.2015.06.020>
- [27] András Kocsor, Kornél Kovács, and Csaba Szepesvári. 2004. *Margin Maximizing Discriminant Analysis*. Springer, Berlin, 227–238. DOI : [http://dx.doi.org/10.1007/978-3-540-30115-8\\_23](http://dx.doi.org/10.1007/978-3-540-30115-8_23)
- [28] Tomasz Krzeszowski, Adam Switonski, Bogdan Kwolek, Henryk Josinski, and Konrad Wojciechowski. 2014. DTW-based gait recognition from recovered 3-D joint angles and inter-ankle distance. In *Proceedings of the International Conference Computer Vision and Graphics (ICCVG'14)*, Leszek J. Chmielewski, Ryszard Kozera, Bok-Suk Shin, and Konrad Wojciechowski (Eds.), LNCS Vol. 8671. Springer, 356–363. DOI : [http://dx.doi.org/10.1007/978-3-319-11331-9\\_43](http://dx.doi.org/10.1007/978-3-319-11331-9_43)
- [29] M. S. Naresh Kumar and R. Venkatesh Babu. 2012. Human gait recognition using depth camera: A covariance based approach. In *Proceedings of the 8th Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP'12)*. ACM, New York, NY, Article 20, 6 pages. DOI : <http://dx.doi.org/10.1145/2425333.2425353>
- [30] Bogdan Kwolek, Tomasz Krzeszowski, Agnieszka Michalczyk, and Henryk Josinski. 2014. 3D gait recognition using spatio-temporal motion descriptors. In *Proceedings of Intelligent Information and Database Systems: 6th Asian Conference (ACIIDS'14)*. LNCS Vol. 8398. Springer, 595–604. DOI : [http://dx.doi.org/10.1007/978-3-319-05458-2\\_61](http://dx.doi.org/10.1007/978-3-319-05458-2_61)
- [31] Haifeng Li, Tao Jiang, and Keshu Zhang. 2006. Efficient and robust feature extraction by maximum margin criterion. *IEEE Trans. Neural Netw.* 17, 1 (Jan. 2006), 157–165. DOI : <http://dx.doi.org/10.1109/TNN.2005.860852>
- [32] Marco Loog, Robert P. W. Duin, and Reinhold Haeb-Umbach. 2001. Multiclass linear dimension reduction by weighted pairwise fisher criteria. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 7 (Jul. 2001), 762–766. DOI : <http://dx.doi.org/10.1109/34.935849>
- [33] Yasushi Makihara, Hidetoshi Mannami, and Yasushi Yagi. 2011. Gait analysis of gender and age using a large-scale multi-view gait database. In *Proceedings of the 10th Asian Conference on Computer Vision, Volume Part II (ACCV'10)*. Springer-Verlag, Berlin, 440–451.
- [34] M. P. Murray. 1967. Gait as a total pattern of movement: Including a bibliography on gait. *Am. J. Phys. Med. Rehabil.* 46, 1 (1967), 290.
- [35] Johannes Preis, Moritz Kessel, Martin Werner, and Claudia Linnhoff-Popien. 2012. Gait recognition with kinect. In *Proceedings of the 1st International Workshop on Kinect in Pervasive Computing*. 1–4.
- [36] Jan Sedmidubsky, Jakub Valcik, Michal Balazia, and Pavel Zezula. 2012. Gait recognition based on normalized walk cycles. In *Advances in Visual Computing*. LNCS, Vol. 7432. Springer, 11–20. DOI : [http://dx.doi.org/10.1007/978-3-642-33191-6\\_2](http://dx.doi.org/10.1007/978-3-642-33191-6_2)
- [37] Aniruddha Sinha, Kingshuk Chakravarty, and Brojeshwar Bhowmick. 2013. Person identification using skeleton information from kinect. In *Proceedings of the 6th International Conference on Advances in Computer-Human Interactions (ACHI'13)*. 101–108.
- [38] Han Su, Zhi-Wu Liao, and Guo-Yue Chen. 2009. A gait recognition method using L1-PCA and LDA. In *Proceedings of the 8th International Conference on Machine Learning and Cybernetics*, Vol. 6. 3198–3203. DOI : <http://dx.doi.org/10.1109/ICMLC.2009.5212776>
- [39] Faezeh Tafazzoli, George Bebis, Sushil J. Louis, and Muhammad Hussain. 2015. Genetic feature selection for gait recognition. *J. Electron. Imag.* 24, 1 (25 Feb. 2015), 013036. DOI : <http://dx.doi.org/10.1117/1.JEI.24.1.013036>
- [40] Vladimir N. Vapnik. 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, NY.
- [41] Y. Wahab and N. A. Bakar. 2011. Gait analysis measurement for sport application based on ultrasonic system. In *Proceedings of the 2011 IEEE 15th International Symposium on Consumer Electronics (ISCE'11)*. 20–24. DOI : <http://dx.doi.org/10.1109/ISCE.2011.5973775>
- [42] Wei Zeng and Cong Wang. 2014. View-invariant gait recognition via deterministic learning. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN'14)*. 3465–3472. DOI : <http://dx.doi.org/10.1109/IJCNN.2014.6889507>

Received February 2017; revised August 2017; accepted August 2017