Learning Robust Features for Gait Recognition by Maximum Margin Criterion

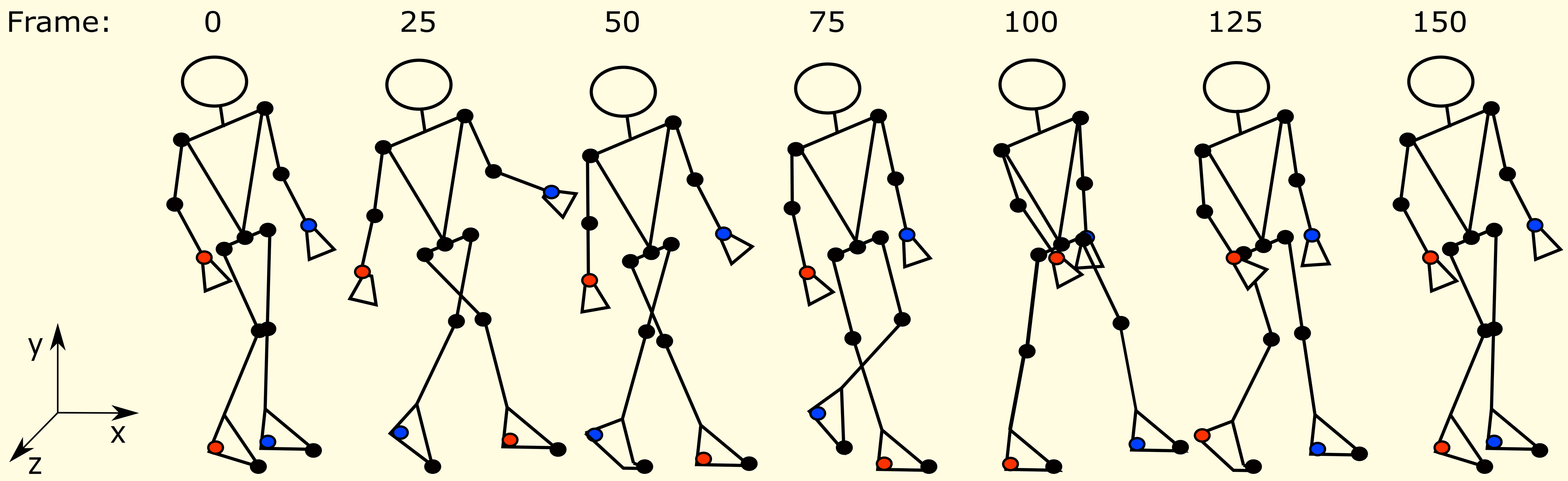
Michal Balazia · Petr Sojka

Masaryk University, Faculty of Informatics, Botanická 68a, 602 00 Brno, Czech Republic

https://gait.fi.muni.cz

Motion Capture Data

Motion capture (MoCap) technology provides video clips of individuals walking which contain **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. These MoCap data can be collected online by a system of multiple cameras (Vicon) or a depth camera (Microsoft Kinect). To visualize the MoCap data, a simplified **stick figure** representing the human skeleton (graph of joints and bones) can be recovered from body point spatial coordinates in time. Recent rapid improvement in MoCap sensor accuracy has brought affordable MoCap technology to assist **human identification** in such applications as access control and video surveillance.



Learning Gait Features

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

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

is a **gait sample** (one gait cycle) in which $\gamma_j(t) \in \mathbb{R}^3$ are 3D coordinates of a joint $j \in \{1, \dots, J\}$ at time $t \in \{1, \dots, T\}$ normalized to the person's position and walk direction. Each learning sample falls strictly into one of the **identity classes** $\{\mathcal{I}_c\}_{c=1}^{C}$. Class $\mathcal{I}_c$ has N_c samples and *a priori* probability p_c.

For the whole labeled data, we denote the **b**etween- and **w**ithin-class and **t**otal **scatter matrices**

$$\Sigma_{\mathrm{B}} = \sum_{c=1}^{C} p_c (\mu_c - \mu)(\mu_c - \mu)^\top$$

$$\Sigma_{\mathrm{W}} = \sum_{c=1}^{C} p_c \sum_{n=1}^{N_c} \left(\mathbf{g}_n^{(c)} - \mu_c\right)\left(\mathbf{g}_n^{(c)} - \mu_c\right)^\top$$

$$\Sigma_{\mathrm{T}} = \sum_{c=1}^{C} p_c \sum_{n=1}^{N_c} \left(\mathbf{g}_n^{(c)} - \mu\right)\left(\mathbf{g}_n^{(c)} - \mu\right)^\top = \Sigma_{\mathrm{B}} + \Sigma_{\mathrm{W}}$$

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 $\mu = \frac{1}{N_L}\sum_{n=1}^{N_L} \mathbf{g}_n$.

Margin of two classes is the Euclidean distance of their means minus variances $\Sigma_c = \sum_{n=1}^{N_c} p_c \left(\mathbf{g}_n^{(c)} - \mu_c\right)\left(\mathbf{g}_n^{(c)} - \mu_c\right)^\top$.

We measure **class separability** of a given feature space by a representation of the Maximum Margin Criterion (MMC) used by the Vapnik's Support Vector Machines (SVM)

$$\mathcal{J} = \frac{1}{2}\sum_{c,c'=1}^{C} p_c p_{c'} \left((\mu_c - \mu_{c'})^\top (\mu_c - \mu_{c'}) - \mathrm{tr}(\Sigma_c + \Sigma_{c'})\right)$$

$$= \ldots = \mathrm{tr}(\Sigma_{\mathrm{B}} - \Sigma_{\mathrm{W}}).$$

Feature extraction is given by **feature matrix** $\mathbf{\Phi} \in \mathbb{R}^{D \times \widehat{D}}$ from a D-dimensional measurement space $\{\mathbf{g}_n\}_{n=1}^{N}$ to a target $\widehat{D}$-dimensional **feature space** $\{\widehat{\mathbf{g}}_n\}_{n=1}^{N}$ with $\widehat{D} < D$ and gait samples $\mathbf{g}_n$ are transformed into **gait templates** $\widehat{\mathbf{g}}_n$ by $\widehat{\mathbf{g}}_n = \mathbf{\Phi}^\top \mathbf{g}_n$. The objective is to learn a transform $\mathbf{\Phi}$ that **maximizes MMC** in the feature space

$$\mathcal{J}(\mathbf{\Phi}) = \mathrm{tr}\left(\mathbf{\Phi}^\top (\Sigma_{\mathrm{B}} - \Sigma_{\mathrm{W}}) \mathbf{\Phi}\right).$$

A **query** sample can be assigned to a walker using the same transformation and comparing against the **gallery** templates with a distance measure such as the **Mahalanobis distance**

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

Measured data can be alternatively projected to a lower dimensional space using **Principal Component Analysis** (PCA) and **Linear Discriminant Analysis** (LDA), resulting in the two-stage PCA+LDA feature extraction technique

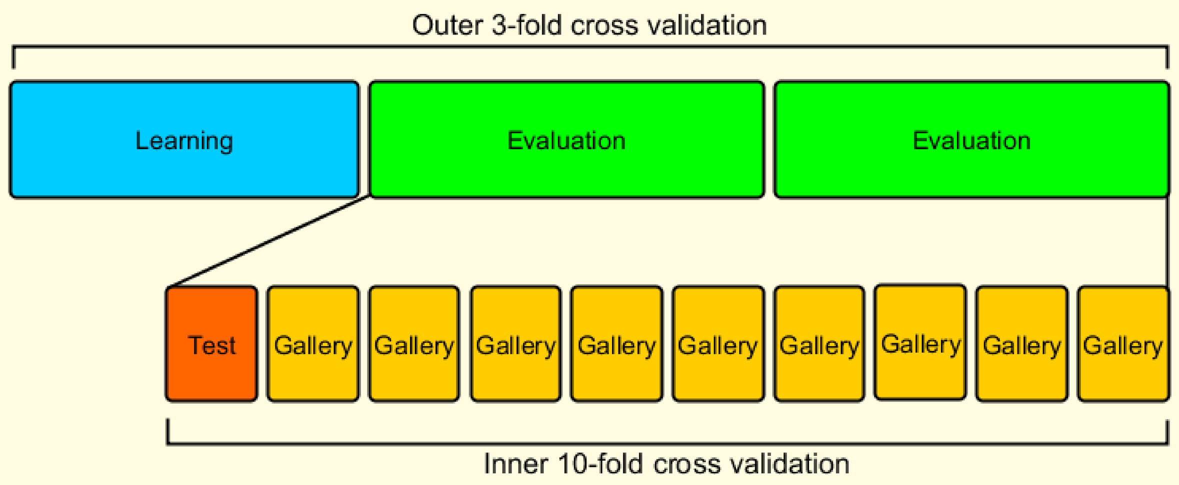
$$\mathcal{J}(\mathbf{\Phi}_{\mathrm{PCA}}) = \mathrm{tr}\left(\mathbf{\Phi}_{\mathrm{PCA}}^\top \Sigma_{\mathrm{T}} \mathbf{\Phi}_{\mathrm{PCA}}\right)$$

$$\mathcal{J}(\mathbf{\Phi}_{\mathrm{LDA}}) = \mathrm{tr}\left(\frac{\mathbf{\Phi}_{\mathrm{LDA}}^\top \mathbf{\Phi}_{\mathrm{PCA}}^\top \Sigma_{\mathrm{B}} \mathbf{\Phi}_{\mathrm{PCA}} \mathbf{\Phi}_{\mathrm{LDA}}}{\mathbf{\Phi}_{\mathrm{LDA}}^\top \mathbf{\Phi}_{\mathrm{PCA}}^\top \Sigma_{\mathrm{W}} \mathbf{\Phi}_{\mathrm{PCA}} \mathbf{\Phi}_{\mathrm{LDA}}}\right)$$

Evaluation Set-Up

For the evaluation purposes we have extracted a large number of samples from the general **CMU MoCap database** as a well-known and recognized database of structural human motion data. The extracted database has **54 walking subjects** that performed **3,843 samples** in total, which makes an average of about 71 samples per subject. Motions are recorded with the optical marker-based Vicon system with tracking space that creates a video surveillance environment.

All results are estimated with **nested cross-validation**. In the outer 3-fold cross-validation loop, templates in one fold are used for learning the features and the remaining two folds for evaluations. Feature space is first evaluated on class separability coefficients. Evaluation of classification based metrics advances to the inner 10-fold cross-validation loop taking one dis-labeled fold as testing set and rest as gallery. Test templates are classified by the winner-takes-all strategy.

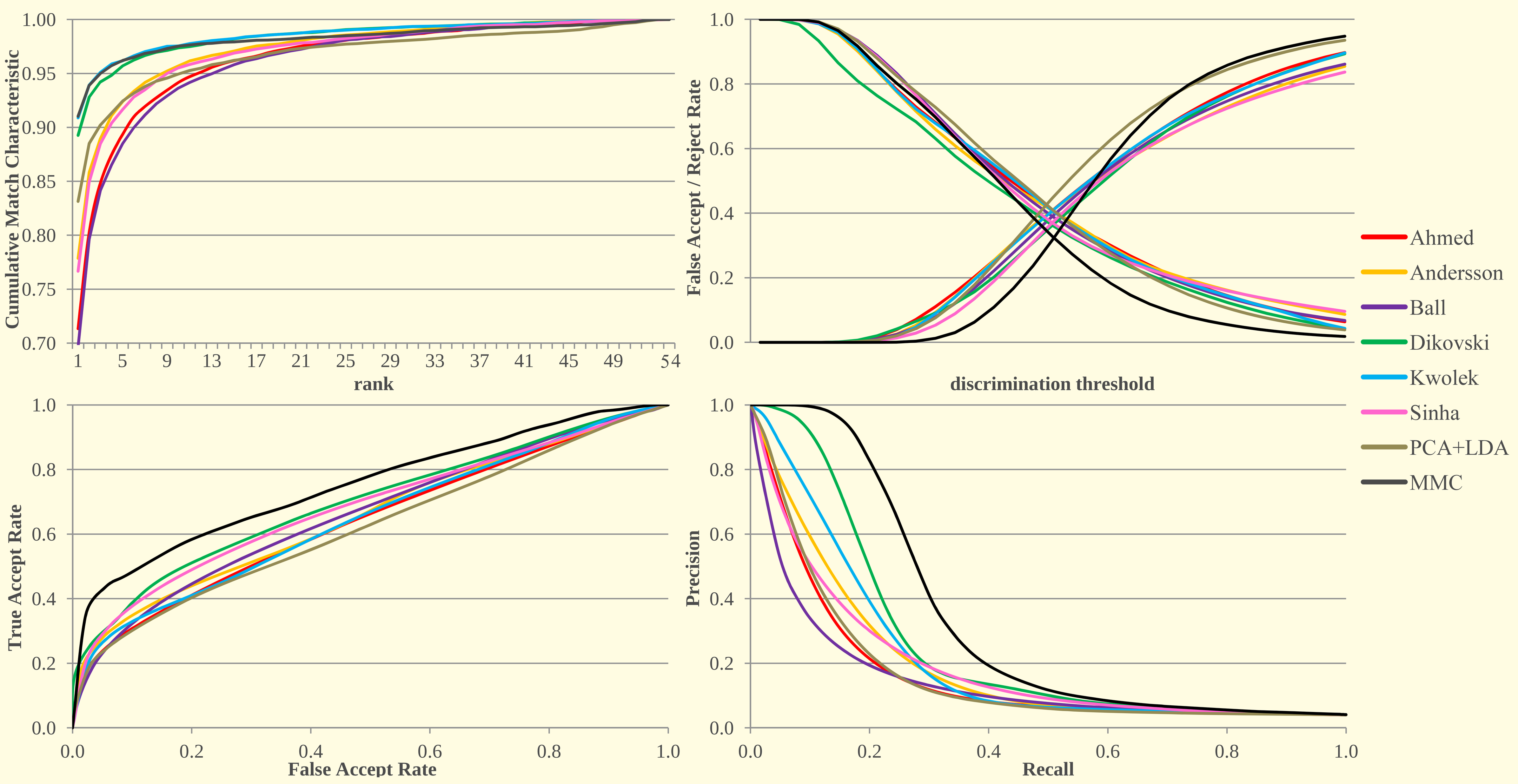


Class separability coefficients:
- Davies-Bouldin Index (DBI)
- Dunn Index (DI)
- Silhouette Coefficient (SC)
- Fisher's Discriminant Ratio (FDR)

Classification based metrics:
- Cumulative Match Characteristic
- False Accept Rate / False Reject Rate
- Receiver Operating Characteristic (ROC)
- Recall / Precision Rate
- Correct Classification Rate (CCR)
- Equal Error Rate (EER)
- Area Under ROC Curve (AUC)
- Mean Average Precision (MAP)

Results

method	template dimensionality	class separability coefficients				classification based metrics			
		DBI	DI	SC	FDR	CCR	EER	AUC	MAP
Ahmed	24	189.31	1.1823	−0.1323	0.9213	0.7134	0.411	0.6387	0.1617
Andersson	80	175.5	1.2047	−0.1117	0.9277	0.7787	0.413	0.6545	0.1926
Ball	18	200.75	1.014	−0.14	0.982	0.6963	0.3906	0.6612	0.1454
Dikovski	71	138.44	1.788	−0.0873	**1.2803**	0.8926	0.3625	0.6964	0.2582
Kwolek	660	151.28	1.2027	−0.074	0.974	0.9089	0.4072	0.6477	0.2121
Sinha	45	166.83	1.3	−0.1256	1.0413	0.7666	0.3706	0.6809	0.1858
PCA+LDA	between C and $N_L - C$	195.19	1.021	−0.084	0.8207	0.8314	0.447	0.6216	0.1643
MMC	up to $C - 1$	**96.2**	**1.8453**	**0.2227**	1.278	**0.9102**	**0.3223**	**0.7551**	**0.2996**



The MMC-based optimization technique appears to be more effective than designing geometric gait features. We interpret the high scores as a sign of robustness. Apart from 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.

Acknowledgements

Published in the Proceedings of the 23rd IEEE/IAPR **International Conference on Pattern Recognition** (ICPR), Cancún, Mexico, 2016. IEEE Xplore DL, pp. 901–906. https://doi.org/10.1109/ICPR.2016.7899750

The data used in this project was created with funding from NSF EIA-0196217 and was obtained from http://mocap.cs.cmu.edu.