A Comparison of PCA, KPCA and LDA for Feature Extraction to Recognize Affect in Gait Kinematics

Michelle Karg[1], Robert Jenke[1], Wolfgang Seiberl[2],
Kolja Kühnlenz[1], Ansgar Schwirtz[2], Martin Buss[1]
Institute of Automatic Control Engineering[1], Institute of Biomechanics in Sports[2],
TU München, 80 333 Munich
`http://www.lsr.ei.tum.de`[1], `http://www.sp.tum.de`[2]

Abstract

This study investigates recognition of affect in human walking as daily motion, in order to provide a means for affect recognition at distance. For this purpose, a data base of affective gait patterns from non-professional actors has been recorded with optical motion tracking. Principal Component Analysis (PCA), Kernel PCA (KPCA) and Linear Discriminant Analysis (LDA) are applied to kinematic parameters and compared for feature extraction. LDA in combination with Naive Bayes leads to an accuracy of 91% for person-dependent recognition of four discrete affective states based on observation of barely a single stride. Extra-success comparing to inter-individual recognition is twice as much. Furthermore, affective states which differ in arousal or dominance are better recognizable in walking. Though primary task of gait is locomotion, cues about a walker's affective state are recognizable with techniques from machine learning.

1. Introduction

Humans interact socially with computers and robots. It is supposed that integration of emotions enhances Human-Computer- and Human-Robot-Interaction [16, 17]. Lazarus defines emotion as the combination of physiological disturbance, action tendencies, which are not necessarily acted out, and affect [12]. Whereas affect itself is the subjective experience during an emotion. Detection of affect is generally based on observing facial expressions, linguistic as well as acoustic features in speech, physiological parameters, gesture and body motions [16]. To enhance recognition of affect, combination of different modalities is investigated. Body motions are considered as additional modality to estimate the affective state of a human from distance. Since walking is a day-to-day motion, this work concentrates on the recognition of affect in gait patterns. Psychological studies support that emotions are expressed in walking.

This work is motivated by results of a previous study [10]. Inter-individual recognition of emotions in gait patterns is above chance level, however recognition rate is highly influenced by individual walking styles. Also the affective states sad and angry are better recognizable than neutral and happy. This has risen the questions what performance can be achieved if emotions are recognized person-dependent and if differences in arousal are better recognizable in gait patterns. For this purpose a data base has been recorded. The contribution of this work is to compare PCA, KPCA, LDA, and GDA to extract relevant features from kinematic parameters. Secondly, inter-individual as well as person-dependent recognition rate is determined based on observing a single stride using 1 Nearest Neighbor, Naive Bayes and Support Vector Machine for classification. Thirdly, recognition of discrete affective states is compared with recognition of extremes on the affective dimensions pleasure, arousal and dominance.

The outline of the paper is as follows. Section 2 summarizes related work about recognition of affect in walking. Section 3 describes the data base, data preprocessing, feature extraction and classification. Results are presented in section 4 and discussed in section 5. The paper ends with a conclusion in section 6.

2. Related Work

Various models to categorize emotions exist in psychology. Ekman's basic emotions, which are anger, disgust, fear, joy, sadness, and surprise, and the dimensional pleasure-arousal-dominance (PAD) model are widely used in automatic emotion recognition [8, 14]. The PAD model spans a 3-dimensional space with the independent and bipolar axes pleasure, arousal and dominance. An affective state is described as a point within this state space.

Early studies based on black displays illuminating only

978-1-4244-4799-2/09/$25.00

the joints of the body, demonstrated that human observers can recognize gender or a familiar person in walking [6,11]. Montepare *et al.'s* psychological study indicated that observers can also identify emotions from variations in walking styles [15]. Furthermore, the emotions sadness and anger are easier to recognize than pride for human observers. Crane and Gross showed that bodily expression of felt and recognized emotions are also associated with emotion-specific changes in gait kinematics and not solely depend on gesticulatory behavior [5]. They identified velocity, cadence, head orientation, shoulder and elbow range of motion as significant parameters which are affected by emotions.

Walking is a complex movement in which the emotional state is covered by the primary task to perform locomotion. To the authors knowledge, only one work studies recognition of emotions in walking with techniques from machine learning [9]. Janssen *et al*. investigated the recognition of four emotional states by means of artificial neural nets. Person-dependent recognition reaches 83.7% in average based on kinetic data measured with a force plate in the ground. However, inter-individual recognition rate remains around chance level. Using kinematic data leads to 79% correct classification of gait patterns during listening calming, excitatory or no music.

A general overview of analytical techniques for clinical and biomechanical gait analysis is given in [4]. It mainly refers to classification of clinical disorders, though the methods for feature extraction can also be taken for psychological gait analysis. Advanced techniques for dimension reduction, such as Kernel Principal Component Analysis (KPCA) improves recognition of age in walking [21]. Performance of Principal Component Analysis (PCA) and KPCA is compared in [2]. Usually, it is believed that considering class affiliation in dimension reduction, such as Linear Discriminant Analysis (LDA) enhances classification. Martinez and Kak showed that if training sets are small comparing to feature dimension, PCA can outperform LDA [13].

Focus of this work is to extract relevant features from a kinematic data set to improve recognition. A comparison of the capability to recognize differences in pleasure, arousal and dominance might explain why sad and angry are generally better recognizable by an algorithm or a human observer in independent studies. The assumption is that arousal is easier to recognize and thus expressed than pleasure in walking.

3. Methods

3.1. Data Base

To investigate recognition of affect in gait patterns, a data base has been recorded at TU München. The gait of 13 male non-professional actors (mean age: 26) was recorded with a VICON optical tracking system (240 Hz). 35 passive markers were affixed to the participant's skin, where anatomic points define the marker position. Based on the Plug-in-Gait Model, the VICON software provides marker position, joint centres and joint angles over time for further data analysis.

Due to a highly artificial environment and frequent repetition of each affective state, successful induction of affect is challenging and it has been decided to pose affect. Each recording contains at least two strides. Participants walked straight in the laboratory hall. Participants were asked to feel angry, happy, neutral or sad, and to imagine a situation in which they feel a particular affect. Each trial was repeated 10 times. In addition, extremes of the dimensions of the pleasure-arousal-dominance (PAD) model were recorded, so that the gait data base contains also 10 walks of each participant expressing the affective states displeased, content, bored, excited, obedient and dominant.

3.2. Data Preprocessing

For further feature extraction, a single stride is extracted from each gait. As the characteristics of the marker affixed to the heel is similar among all participants, this marker is chosen for extracting a single stride from a complete recording. Figure 1 shows the z-coordinate of the marker affixed to the right heel over time. A minimum indicates that the right heel touches the ground and is followed by a nearly constant position of the heel during the stance phase of the right leg. During the swing phase of the right leg, the foot leaves the ground and the height of the heel increases.

An algorithm based on detecting these minima calculates the starting time of all strides in each recording. The algorithm first selects time intervals, in which height of the marker at the heel is less than 30% of maximum height, then disregards intervals which last less than 125ms and finally calculates the minimum in each interval. The complete recording of one gait is cut down to a single stride for all marker positions, joint centres and joint angles. Following feature extraction is based on observation of one stride.

3.3. Feature Extraction

Crane and Gross indicated the parameters velocity (m/s), cadence (steps/second), head orientation and shoulder range of motion as significant factors which are influenced by the walker's affective state [5]. In addition to these factors, forward tilt of upper body and stride length are investigated. Table 1 lists the parameters calculated for each factor. As a single stride is clipped from the complete data set based on the trajectory of the right heel, velocity, cadence and step length are computed consistently on the forwards motion of this marker. To describe the range of motion of the shoulder

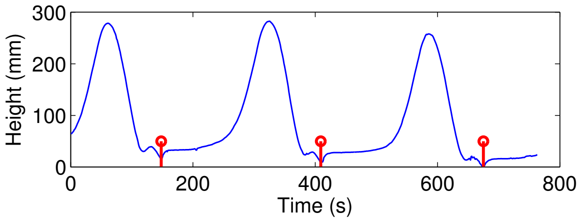

Figure 1. The periodic characteristic of the marker affixed to the heel is an example of the 105 recorded trajectories. Here, a minimum, marked with a red bar, indicates the start of one stride.

angles, forward tilt as well as abduction, neck tilt and thorax tilt, the parameters minimum, mean and maximum for the trajectories of the corresponding angles are calculated. Hence, the complete kinematic feature vector $\mathbf{x}_{kin}$ contains $M = 15$ parameters. Mean and standard deviation of the feature vector $\mathbf{x}_{kin,norm}$ is normalized for PCA and KPCA.

Transforming the feature space can improve classification. With this purpose, PCA, KPCA, LDA and GDA are applied [1, 7, 20]. PCA, also known as Karhunen-Lòeve Transformation, transforms the original data space to an orthogonal set of Principal Components (PCs). Principal components $\mathbf{u}_i$ with highest Eigenvalues λ_i represent the vectors with maximum variance in the data set. The Eigenvalue problem to solve is defined as

$$\begin{aligned}(\frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_{kin,norm}\mathbf{x}_{kin,norm}^{T})\mathbf{u}_i &= \lambda_i\mathbf{u}_i \qquad (1)\\ \text{with } i &= 1, ..., M\end{aligned}$$

with N observations of $\mathbf{x}_{kin,norm}$. The term within the brackets in Equ. 1 is the covariance matrix. Original data is mapped on up to a maximum of M principal components.

A non-linear extension of PCA is KPCA [19]. Its advantages are nonlinearity of eigenvectors and higher number of eigenvectors. KPCA maps the original data vector $\mathbf{x}_{kin,norm}$ into a feature space F using a non-linear map Φ.

$$\Phi : R^n \rightarrow F, \quad \mathbf{x}_{kin,norm} \mapsto \mathbf{X}_{kin,norm} \qquad (2)$$

Then, it performs linear PCA in the high-dimensional space F which corresponds to a non-linear PCA in the original data space. The covariance matrix $\mathbf{C}$ is given by

$$\mathbf{C} = \frac{1}{N}\sum_{j=1}^{N}\Phi(\mathbf{x}_{kin,norm,j})\Phi^T(\mathbf{x}_{kin,norm,j})\ . \qquad (3)$$

Applying the kernel trick, the eigenvalue problem becomes

$$\begin{aligned}N\lambda\mathbf{a} &= \mathbf{Ka} \qquad (4)\\ \text{with } K_{ij} &:= \Phi(\mathbf{x}_{kin,norm,i})^T\Phi(\mathbf{x}_{kin,norm,j}),\end{aligned}$$

where the scalar product of Φ can be substituted by a kernel function $K(\mathbf{x}, \mathbf{y})$. Within this study, a polynomial kernel $K = (\mathbf{x}^T\mathbf{y})^d$ and a Gaussian kernel $K = \exp((-\|\mathbf{x} - \mathbf{y}\|^2)(2\sigma^2)^{-1})$ are used. In contrast to PCA, KPCA can find up to N, in general number of instances, eigenvectors.

Factor	Parameter
Stride Length	Length of One Stride
Cadence	Time for One Stride
Velocity	Cadence/(Stride Length)
Neck Angle (Forward Tilt)	Min, Mean, Max
Shoulder Angle (Flexion)	Min, Mean, Max
Shoulder Angle (Abduction)	Min, Mean, Max
Thorax Angle (Forward Tilt)	Min, Mean, Max

Table 1. Kinematic features.

In contrast to algorithms based on PCA, LDA considers class membership for dimension reduction. Key idea of LDA is to separate class means of the projected directions well while achieving a small variance around these means. Alike PCA, the derived features of LDA are linear combinations of the original data. For a c-class problem, the maximum number of eigenvectors $\mathbf{w}_k$ is $(c-1)$, in our case 3. Based on the total mean $\mathbf{m} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{kin,i}$, the mean $\mathbf{m}_j$ for samples $\mathbf{x}_{kin,i}^j$ of each class j, and the number of samples n_j for each class, the between-class scatter matrix S_B

$$S_B := \sum_{j=1}^{c} n_j(\mathbf{m}_j - \mathbf{m})(\mathbf{m}_j - \mathbf{m})^T \qquad (5)$$

and the within-class scatter matrix S_W

$$S_W := \sum_{j=1}^{c}\sum_{i=1}^{l_j}(\mathbf{x}_{kin,i}^j - \mathbf{m}_j)(\mathbf{x}_{kin,i}^j - \mathbf{m}_j)^T \qquad (6)$$

are defined. Maximizing the between-class measure while minimizing the within-class measure leads to the following eigenvalue problem

$$(S_B - \lambda_k S_W)\mathbf{w}_k = 0 \text{ with } k = 1, 2, 3. \qquad (7)$$

The kernel trick allows GDA to compute LDA in a high-dimensional feature space without ever having to explicitly map into it. Polynomial as well as Gaussian kernel are applied in our applications.

3.4. Classification

For recognition several standard classifiers are compared. Naive Bayes is a classifier, which estimates probabilities for the membership to each class. It is robust to irrelevant features. However, features which are not conditional independent or not Gaussian distributed decrease accuracy. 1 Nearest-Neighbor (1NN) using Euclidean distance is a lazy classifier. It can produce arbitrarily shaped decision boundaries and is susceptible to noise. Besides Nearest Neighbor and Naive Bayes, Support Vector Machines afford good results in emotion recognition. A L1 soft-margin Support Vector Machine (SVM) with a radial basis function as kernel is used as well for classification [3]. The SVM learning problem is a convex optimization problem, so that the optimal solution is calculated in contrast to Neural Networks which can get stuck in local minima. Standard SVM is limited to two-class problems. Common extension to multi-class problems is the one-against-one method.

As the number of samples in the data sets is small, the recognition rate is calculated using leave-one-out cross validation.

4. Experimental Results

4.1. Person-Dependent Recognition

Following results are based on 13 data sets each with samples of one walker expressing all affective states sad, neutral, happy and angry. Recognition rate is calculated for the data set of each walker separately. For evaluation, mean accuracy of all walkers is compared. Results for all classifiers are shown in Table 2. Recognition based solely on velocity, stride length and cadence (vts) reaches 84% averaged over all walkers for 1NN. Mean accuracy increases to 91 % for classification based on the complete feature vector, including 15 kinematic parameters (all). Classification based barely on velocity, stride length and speed gives lower results as based on the complete feature vector. Recognition results based on a feature vector excluding velocity, stride length and speed (all w/o vts) is comparable to results based only on vts. Hence, the complete data sets contain redundant information.

Table 2 lists mean accuracy for applying PCA, KPCA and LDA to the complete feature vector $\mathbf{x}_{kin}$ (all). Dimension reduction with PCA to eight dimensions leads to mean accuracy of 88%. The nonlinear extension KPCA with a Gaussian kernel reaches less accuracy for all classifiers compared to dimension reduction with PCA. Resulting feature representation is even less suitable than the original feature space. Changing the kernel to a polynomial kernel does not improve recognition.

LDA reduces the feature vector to 3 dimensions. It performs best, if the feature vector $\mathbf{x}_{kin,LDA}$ excludes mean and maximum of the thorax angle. In this case, mean accuracy is 91% for all classifiers. Individual recognition rates for each walker range between 80% and 100%. Standard deviation is 6%. The confusion matrix for this case, see Table 3, shows that the affective states sad, neutral and angry are better recognizable than happy.

As LDA reduces the data efficiently onto a low-dimensional space, it is suited for graphical representation of the data sets. Figure 2 and Fig. 3 show exemplary the data set of walker 7, whose affective walks are recognized 100% correctly, and of walker 4. The 2-dimensional representation of affective gaits are clearly separable in Fig. 2. The clusters for the gaits of each affective walk are closer to each other in Fig. 3, which leads to lower accuracy (78%). Mapping of LDA is calculated based on the training set, so that analogous plots for each leave-one-out cycle deviate slightly from each other. Still, the plots for each subject differ less within one subject than in-between subjects. Overall accuracy for reducing the feature vector to two dimensions with LDA applied to $\mathbf{x}_{kin,LDA}$ is 88% for Naive Bayes and SVM. GDA performs poorly on this data set. Recognition rates are around chance level for all classifiers.

Features	1NN	Naive Bayes	SVM
vts	84	83	76
all w/o vts	83	86	88
all	83	**91**	88
PCA	88	87	87
KPCA	77	68	85
LDA	**91**	**91**	**91**

Table 2. Mean accuracy is above chance (25%) for distinguishing four affective states. LDA transforms the original data sets most efficiently.

affect	neutral	happy	sad	angry	acc
neutral	122	6	2	0	94%
happy	13	100	14	3	77%
sad	1	10	118	1	91%
angry	0	1	5	124	95%

Table 3. Confusion matrix shows that the affective state happy is often misclassified as neutral or sad.

4.2. Inter-Individual Recognition

Inter-individual recognition rate for all mentioned feature extraction and classification methods have been calculated for this data base. However, training the classifier on the data sets of 12 walkers and recognition of the affective states of the remaining walker, leads to lower recognition rates, see Table 4. This is explained by inter-individual differences in expressing affect and individual walking styles.

Best result is achieved for applying LDA on the complete feature vector $\mathbf{x}_{kin}$ with 58% accuracy. Extra success comparing to a random predictor is 44% for inter-individual recognition. In comparison, person-dependent extra success is twice as much with 88% for LDA. Results of GDA are around chance level. Also, KPCA with a Gaussian kernel and reduction to 2 dimensions does not perform better than PCA. In this case, PCA projects the original data set on 10 dimensions.

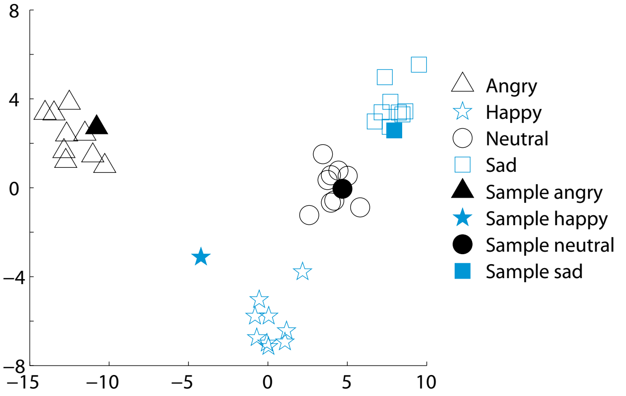

Figure 2. Mapping the complete data set of walker 7 on 2 dimensions using LDA shows that the cluster of angry trials is clearly separable from the others.

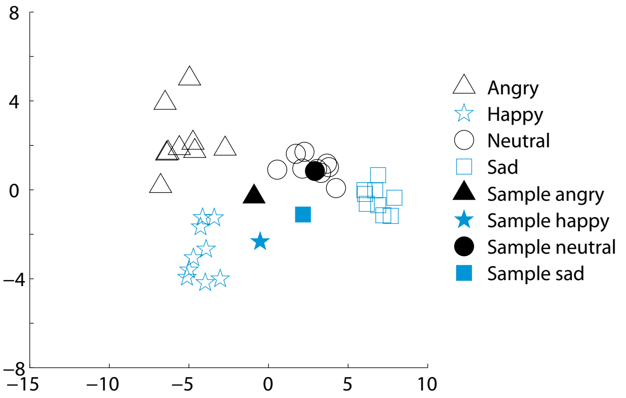

Figure 3. Clusters for each affective walking style of walker 4 are spatially less separated, so that misclassifications occur. Within this leave-one-out cycle, the angry and sad sample are misclassified as neutral.

Features	1NN	Naive Bayes	SVM
vts	52	45	45
all	56	40	40
PCA	50	55	47
KPCA	41	52	42
LDA	**54**	**58**	**54**

Table 4. Recognition of the affective states of an unknown walker reaches 58% accuracy.

4.3. Dimensional Categorization of Affect

The PAD model considers the three dimensions pleasure, arousal and dominance to describe an affective state. Walkers expressing extremes of each dimension and one neutral data set have been recorded. Each dimension is divided in high, neutral and low. As difference in accuracy varies little among classifiers for recognition of affect in gait patterns, see Table 2, results are presented based on Naive Bayes in this subsection. Table 5 shows mean person-dependent recognition rates separately for each dimension. In all cases, recognition rate for distinguishing expressive levels of pleasure is lower than for arousal and dominance.

Features	Pleasure	Arousal	Dominance
vts	71	92	80
all	**84**	**96**	**93**
PCA	79	92	90
KPCA	60	68	75
LDA	80	96	92

Table 5. Recognition rates for distinguishing three levels of arousal or dominance are higher than for pleasure, regardless of which feature mapping is applied to the original feature space.

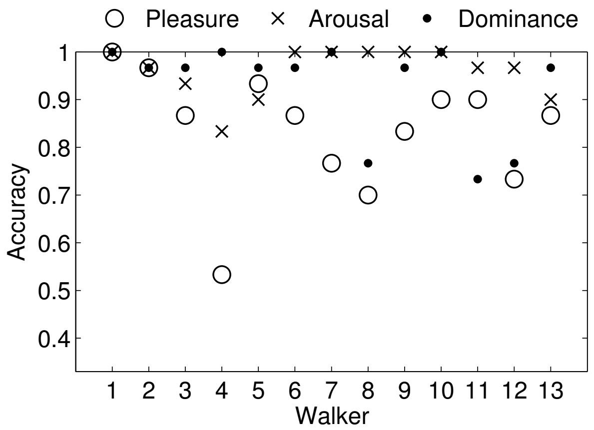

Figure 4. Person-dependent recognition for three levels of arousal and dominance reaches 100% for individual walkers, whereas pleasure is lowest recognizable.

For this three-class problem, chance level is 33%. LDA is applied to the feature vector containing stride length, cadence, velocity, neck angle and minimum of shoulder flexion and maps the data sets onto two dimensions. Recognition rate is comparable to using all features. PCA reduces the number of dimension to five, though accuracy is not improved. Recognition rates for KPCA using a Gaussian kernel are also listed in Tab 5.

Figure 4 shows separately the accuracy for the data sets of each walker. In accordance to Table 5, lowest accuracy is achieved for most walkers for the dimension pleasure. Depending on the walker, different levels of arousal or dominance are best recognizable. Mean accuracy is 84% for pleasure, 96% for arousal and 93% for dominance.

5. Discussion

Within this work inter-individual as well as person-dependent recognition above chance has been accomplished for four acted affective states during walking. In comparison to [9], accuracy is increased by selecting significant kinematic parameters. Recognition of affective states of a known walker leads to twice as much extra success as of an unknown walker. Due to individual walking styles, recognition of affective states barely on gait is highly sensitive to individuals. Further improvement can be achieved by normalization of one's walking pattern.

Affective states which differ in arousal or dominance are better recognizable in walking than affective states with different pleasure levels. This explains that in this work and in [10] the affective state happy is harder recognizable. This result is in accordance with similar findings for the body motions knocking and drinking [18], for which it has been shown that activation is highly correlated with kinematics.

Essential characteristic for recognizing affect in walking is velocity. Person-dependent recognition reaches 84% barely on measuring velocity, cadence and length of a single stride. Using a set of 15 features, accuracy increases to 91%. Optimal projection is achieved with LDA onto three dimensions. As number of training samples is sufficiently larger than number of features, the advantage of considering class affiliation outperforms PCA. Nonlinear extensions, like KPCA and GDA, do not improve accuracy. As mentioned in [20], nonlinear techniques do not necessarily perform as well in real-world tasks as in selected artificial tasks.

6. Conclusion and Outlook

This study investigates recognition of posed affective states in human gait patterns. For this purpose, we recorded a gait data base. Based on a single stride, person-dependent recognition rates have been calculated. Comparing recognition of discrete affective states and distinguishing extremes on the dimensions pleasure, arousal and dominance of the PAD-model leads to the conclusion that gait is more capable to reveal affective states which differ in arousal and dominance than in pleasure. It is concluded that affect is recognizable above chance in gait patterns, however further investigation is required how well recognition of affective states based on gait patterns performs in real-world scenarios. In this context, a potential side aspect is how strong influences like gender, age, weight or complaints interfere with recognition of affect.

Acknowledgements

This work is supported within the DFG excellence initiative research cluster ”Cognition for Technical Systems - CoTeSys”, see also www.cotesys.org.

References

[1] C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[2] L. Cao, K. Chua, W. Chong, H. Lee, and Q. Gu. A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine. *Neurocomp.*, 55:321–336, 2003.

[3] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. 2001. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[4] T. Chau. A review of analytical techniques for gait data: Part 1 and 2. *Gait and Posture*, 13:49-66, 102:120, 2001.

[5] E. Crane and M.Gross. Motion capture and emotion: Affect detection in whole body movement. In *ACII*, 2007.

[6] J. Cutting and L. Kozlowski. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of Psychonomic Society*, 9:353–356, 1977.

[7] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, 2001.

[8] P. Ekman and W. V. Friesen. A new pan-cultural facial expression of emotion. *Motivation & Emo.*, 10:159–168, 1986.

[9] D. Janssen, W. Schöllhorn, J. Lubienetzki, K. Fölling, H. Kokenge, and K. Davids. Recognition of emotions in gait patterns by means of artificial neural nets. *Nonverbal Behaviour*, 32:79–92, 2008.

[10] M. Karg, R. Jenke, K. Kühnlenz, and M. Buss. A two-fold PCA-approach for inter-individual recognition of emotions in natural walking. In *Int. Conf. on MLDM*, 2009.

[11] L. Kozlowski and J. Cutting. Recognizing the sex of a walker from a dynamic point-light display. *Perception and Psychophysics*, 21:575–580, 1977.

[12] R. Lazarus. *Emotion & Adaptation*. Oxford Univ. Press, 1991.

[13] A. Martinez and A. Kak. PCA versus LDA. *IEEE Transactions on PA and MI*, 23:228–233, 2001.

[14] A. Mehrabian and J. Russel. Evidence for a three-factor theory of emotions. *J. of Research in Personality*, 11:273–94, 1977.

[15] J. Montepare, S. Goldstein, and A. Clausen. The identification of emotions from gait information. *Nonverbal Behavior*, 11(1):33–42, 1987.

[16] C. Peter and R. Beale. *Affect and Emotion in Human-Computer Interaction*. Springer, 2008.

[17] R. Picard. *Affective Computing*. MIT Press, 2002.

[18] F. Pollick, H. Paterson, A. Bruderin, and A. Sanford. Perceiving affect from arm movement. *Cognition*, 82:B51–B61, 2001.

[19] B. Schölkopf, A. Smola, and K.-R. Müller. Kernel principal component analysis. *ANN*, 1327:583–588, 1997.

[20] L. van der Maaten, E. Postma, and H. van den Herik. Dimensionality reduction: A comparative review. *Submitted to Neurocognition*, 2009.

[21] J. Wu, J. Wang, and L.Liu. Feature extraction via KPCA for classification of gait patterns. *Human Movement Science*, 26:393–411, 2007.