A Cumulant-Based Method for Gait Identification Using Accelerometer Data with Principal Component Analysis and Support Vector Machine

SEBASTIJAN SPRAGER, DAMJAN ZAZULA
System Software Laboratory
University of Maribor, Faculty of Electrical Engineering and Computer Science
Smetanova ulica 17, 2000 Maribor
SLOVENIA
sebastijan.sprager@uni-mb.si, zazula@uni-mb.si http://storm.uni-mb.si

Abstract: - In this paper a cumulant-based method for identification of gait using accelerometer data is presented. Acceleration data of three different walking speeds (slow, normal and fast) for each subject was acquired by the accelerometer embedded in cell phone which was attached to the person's hip. Data analysis was based on gait cycles that were detected first. Cumulants of order from 1 to 4 with different number of lags were calculated. Feature vectors for classification were built using dimension reduction on calculated cumulants by principal component analysis (PCA). The classification was accomplished by support vector machines (SVM) with radial basis kernel. According to portion of variance covered in the calculated principal components, different lengths of feature vectors were tested. Six healthy young subjects participated in the experiment. The average person recognition rate based on gait classification was 90.3±3.2%. A similarity measure for discerning different walking types of the same subject was also introduced using dimension reduction on accelerometer data by PCA.

Key-Words: - Gait Identification, Gait Recognition, Body Sensor, Accelerometer, Pattern Recognition, High-Order Statistics, Cumulants

1 Introduction

Rapid development of body sensors provides a number of innovations in the area of biomedicine. Such sensors are interconnected and form a special network, called Body Area Network (BAN). BANs collect data that can be observed as parameters providing the information about the user's health state. Important components of today's BANs are sensors of acceleration, i.e. accelerometers. Accelerometers have recently been introduced in more complex and advanced technological commercial products, mostly in cell phones. Cell phones today represent multipurpose devices and can directly be implemented as parts of BANs.

The purpose of our experiments was to determine whether and how accurate it is possible recognize identity of the user from the accelerometer data acquired by the cell-phone accelerometer. Our purpose was also to determine the efficiency of recognition of different types of walking with the same user and how similar is to the walking of other users. Every person has their own style of walking. This means that the identification of the observed person can be recognized according to their walking style, and, consequently, it is possible to determine the similarity of walking styles of several observed persons.

A sample of gait data is obtained by an accelerometer for each person and stored in a database. If, at some later time, the walking pattern of the same observed person changes, this is probably due to health problems. Therefore, our future research aims at upgrading the proposed method, so that we will be able to recognize personal movement disorders.

The problem of the gait analysis can be divided into two parts: biometric gait identification and gait identification for biomedical purposes. As we will see later, the majority of existing works is related to the gait identification in terms of biometry. Nevertheless, due to analogy of the problem, the gait analysis can be the same in both cases.

The existent gait identification methods can be grouped into three categories: machine vision (MV) based, floor sensor (FS) based and wearable sensor (WS) based [7]. By MV-based gait recognition, gait is captured using a video camera from the distance [10]. Video and image processing techniques are then employed to extract gait features for recognition purposes. By FS-based gait recognition, a set of force plates are installed on the floor [8]. Such force sensors measure gait related features, when a person walks over them. By WS-based gait recognition, including our proposed method, the information is collected using body worn motion recording sensors, mostly accelerometers [1-6]. Sensors can be located at different positions on the observed person, such as pocket [1, 6], leg [2, 3, 18], waist [4], belt [5], hip, etc. The acquired acceleration signal of the gait is then used for gait

identification.

Feature extraction from detected gait is crucial for the efficient gait identification. In the related works, many different features were used, such as absolute distance [1], correlation [1, 5], histogram similarity [1, 2, 5], high-order moments [1], cycle length [2], estimated walking speed and distance [4], FFT coefficients [5], wavelet decomposition [6, 8], DTW [18], and other regular features, such as mean, median, standard deviation, RMS, maximum and minimum value and amplitude [3].

This paper describes gait identification and similarity measurement using an accelerometer embedded in the cell phone, which was implemented attached to the hip. Signal cycles were detected from the accelerometer data and cumulants of orders from 1 to 4 with lags 3, 5, 10, 20 and 30 were calculated for each order as features from those cycles. Classification was performed using support vector machines. A similarity measure of subjects' gait was devised using principal component analysis (PCA).

Section 2 describes the experimental protocol. Section 3 explains the method for gait identification and similarity measurement. Results are presented in Section 4, while Section 5 concludes the paper.

2 Experiment protocol

We prepared experimental protocol that induces collection of data containing the information on the individual characteristics of a person's gait. The experiment was performed in a 50 m long corridor with the surface made of stone plates. Each subject was asked to walk across the corridor with their normal walking speed. After a few seconds of rest the subject walked back to the starting point with same walking speed. In the second part of the experiment, the same procedure was repeated, but now with a faster walking speed, while the last part of the experiment required slow walking speed. Thus, we collected 6 segments of acceleration signals, two for normal, two for fast and two for slow walking.

Six healthy male subjects were tested. Their average age was 30.2 years with standard deviation 4.02 years. The average height of the subjects was 179 cm with standard deviation 3.2 cm.

3 Gait identification

Our method for gait identification follows the signal processing flow, as shown in Figure 1. Subsection 3.1 explains acquisition of accelerometer signal and its preprocessing. Subsection 3.2 explains the extraction of gait cycles from the acquired accelerometer signals. Subsection 3.3 describes our implementation of cumulants and Subsection 3.4 their role in feature extraction from the detected gait cycles. Subsection 3.5 reveals the classification results.

3.1 Accelerometer signal acquisition

Acceleration data was acquired using a cell phone with a built-in accelerometer. The accelerometers measure accelerations up to three different directions, regarding to the type of accelerometer (one-, two- or three-axis accelerometer). Acquired multichannel signals contain the magnitude and the direction of the acceleration. In our experiments, the cell phone Nokia N95 was used for data acquisition. This cell phone contains 3-axis accelerometer SM LIS302DL which measures accelerations between ±2g with a resolution of 10 bits. Due to power saving function of the cell phone the sampling frequency is not constant. Therefore we had to interpolate the acquired signal using linear interpolation. The obtained average sampling frequency was 37 Hz.

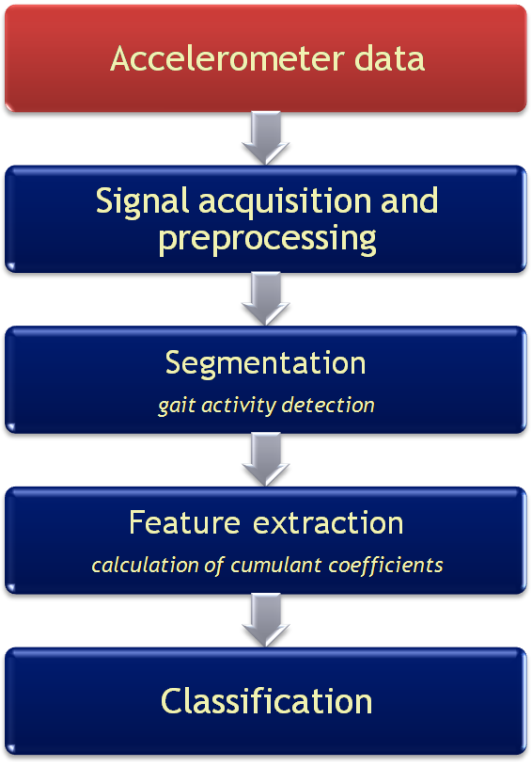

Fig. 1: Signal processing flow of the method for gait identification.

We paid particular attention to the position of the accelerometer. During our experiment the cell phone was attached to the right hip of the subject, as shown in Figure 2. This position turned out to be the most appropriate for the cell phone bearer. We also avoided the possibility of noise-generating oscillations or movements, for example unwanted bouncing during the

walk, if the cell phone were located in the subject's pocket.

The most important task was to eliminate the influence of the position and orientation of the cell phone on the accelerometer data, which could later deteriorate the classification accuracy. Because the cell phone we used doesn't have built-in gyroscopes, we had to calibrate the accelerometer data to the upright posture of each subject. Before carrying out the trial, we asked each subject to stand still in the upright position. The transformation matrix was calculated based on the gravity component of accelerometer's signal and applied to each sample acquired during the subsequent walking. This ensures that the small difference in orientation and phone position, related to the person's hip, does not influence the accelerometer signal significantly. However, it is very important that for each person we achieve more or less same placement and direction of the cell phone. Different placements of the device on the same subject would cause false recognition of subject's walk.

Fig. 2: Mounting position of cell phone and coordinate system of the accelerometer.

Acquisition results in 3-component vectors of samples stored in the matrix **A**:

$$\mathbf{A}_{\mathrm{i}} = [\mathbf{x} \quad \mathbf{y} \quad \mathbf{z}], \tag{1}$$

where i represents an ID of the subject, while $\mathbf{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ represent vectors of acquired samples for each spatial direction.

Experimental protocol ends by labelling the acquired signals. The acquired raw accelerometer data stored in **A** were then automatically segmented to the epochs of walking. The ID of the subject and type of walking were labelled for each of selected segments.

According to the experiment protocol, six segmented signal sets were collected per each trial–the first two represent accelerations during normal walk, the next two fast walk, and the last two slow walk accelerations. Each set contains 3 signals, one for each spatial direction.

3.2 Gait analysis

To determine walking characteristics, a unified approach is necessary for all test cases. We decided to base our further recognition on individual gait cycles. We assume that we are dealing with periodic signals in which every gait cycle represents one period. In fact, the periodicity of gait signals is not strict, which necessitates processing of all gait cycles from each signal set.

Visual inspection of acceleration signals discovered the cycles are clearly visible and the bounds between them are seen as the prominent peaks showing the vertical and horizontal acceleration. We applied the extraction of gait cycles using a modified peak-detection method based on combined dual-axial signal, as described in [9]. Block diagram of the method is shown in Figure 3. We neglected the side acceleration signals and used only vertical and horizontal accelerations. The gravity component was removed from the acceleration signals. Afterwards, the magnitude calculation followed for the two acceleration signals. We squared the signal magnitude in order to emphasise larger values more than smaller values. The squared signals were smoothed by 5 samples long moving average window. Finally, gait cycles for each signal set were extracted by using peak detection on the processed signal. For each segment from $\mathbf{A}_{\mathrm{i}}$ a vector of cycles was extracted:

$$\mathbf{A}_{\mathrm{i}} \rightarrow \mathbf{C} = \left[\mathbf{C}_1^{(\mathrm{t})} \quad \ldots \quad \mathbf{C}_{\mathrm{n}}^{(\mathrm{t})}\right], \tag{2}$$

where $\mathbf{C}_{\mathrm{k}}^{(\mathrm{t})}$ represents a gait cycle, n represents the number of cycles for each segment, and $t \in \{1,2,3\}$ represents walking type. Gait cycles

$$\mathbf{C}_{\mathrm{k}}^{(\mathrm{t})} = [\mathbf{x} \quad \mathbf{y} \quad \mathbf{z}], \tag{3}$$

comprise $\mathbf{x_k y_k z_k}$ vectors of samples for the extracted *k*-th cycle, one for each spatial direction, $\mathbf{x_k}$, $\mathbf{y_k}$, and $\mathbf{z_k}$.

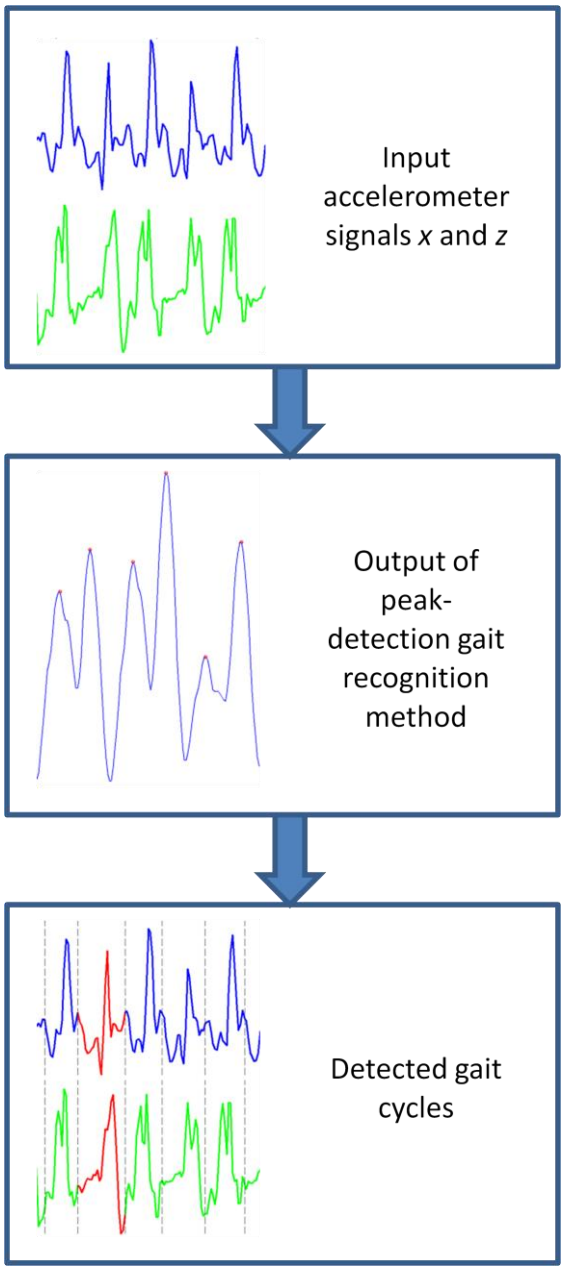

Fig. 3: Block diagram of the method for detection of gait cycles.

3.3 Cumulants as features

Higher-order statistics are very useful when we are dealing with non-Gaussian signals and many real-world applications are truly non-Gaussian, what is also true for the acceleration signals we are dealing with. If the random signal is Gaussian, the first and second order statistics are enough to describe its properties. The third and fourth-order statistics carry additional information about the observed non-Gaussian random signal. Actually, using the higher-order statistics we can identify processes beyond the random signal.

Moments are statistical measures that are used to characterize properties of the observed signal. Although the moments provide all of the needed information for higher-order statistical analysis, it is preferable to work with related quantities called cumulants. Several properties make cumulants mathematically more convenient than moments [11]. Cumulants also measure the departure of a random signal from a Gaussian random signal.

If $x(n)$ is a non-Gaussian random process and $x'(n)$ is a Gaussian random process with same mean and correlation as $x(n)$, then the first two cumulants of $x(n)$ are equal to the mean and covariance of $x(n)$. Cumulants of order 3 and 4 are then defined as [12, 13]:

$$Cum_x^{(k)}(l_1,\dots,l_k) = M_x^{(k)}(l_1,\dots,l_k) - M_{x'}^{(k)}(l_1,\dots,l_k), \quad (4)$$

where k represents the order of cumulant. $M_x^{(k)}$ stand for the k-order moments defined as in the following equations [12]:

$$M_x^{(1)} = E\{x(n)\} = mean_x, \quad (5)$$

$$M_x^{(2)}(l_1) = E\{x^T(n)x(n+l_1)\} = corr_x(l_1), \quad (6)$$

$$M_x^{(3)}(l_1,l_2) = E\{x^T(n)x(n+l_1)x(n+l_2)\}, \quad (7)$$

$$M_x^{(4)}(l_1,l_2,l_3) = E\{x^T(n)x(n+l_1)x(n+l_2)x(n+l_3)\}, \quad (8)$$

where $x(n)$ represents a random signal, l_i a lag in the i-th dimension, and E mathematical expectation $E\{x\} = \sum_i x_i p_i$.

If we look at cumulants as the functions of multiple variables, the first-order cumulant (mean) is a constant, the second-order cumulant is a function of one variable, the third cumulant is a function of two variables and the fourth cumulant is a function of three variables.

Considering the symmetry of cumulants, we focused only on one region from the codomain that carries the same information as the remaining parts of the codomain. For the first order cumulant (mean) the symmetry property is trivial. For second order cumulant (covariance function) we take a half of the vector. The symmetry property for third-order cumulants is shown in Figure 4. The codomain is split in the six regions of symmetry [12]. We chose the first octant (shaded area in Figure 4. The symmetry property of fourth-order cumulant is an extension of the third-order cumulant symmetry property. Instead of triangle shaped area (shown shaded in Figure 4) in this case the regions are prismoids.

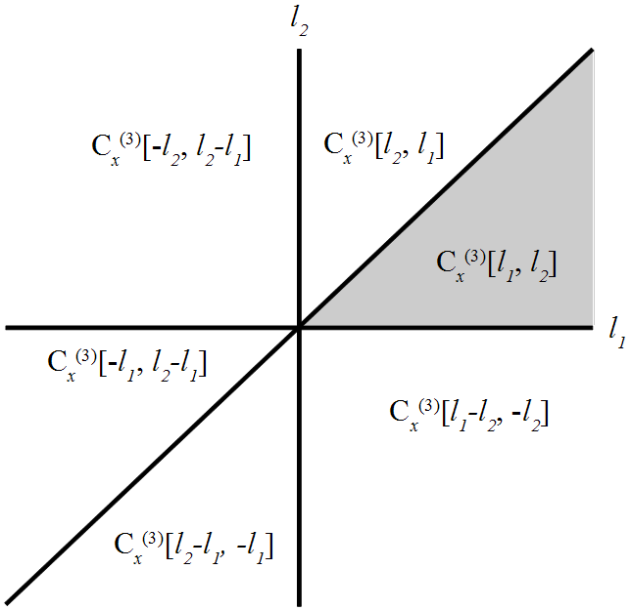

Fig. 4: Six regions of symmetry for third-order cumulant of a real random signal.

3.4 Feature extraction

The crucial step in the proposed method for gait identification is to select features that lead to the best classification results. We experimented with cumulants of orders from 1 to 4. We calculated cumulant coefficients from zero-lag to lag *d* (for second, third and fourth order) for each gait cycle.

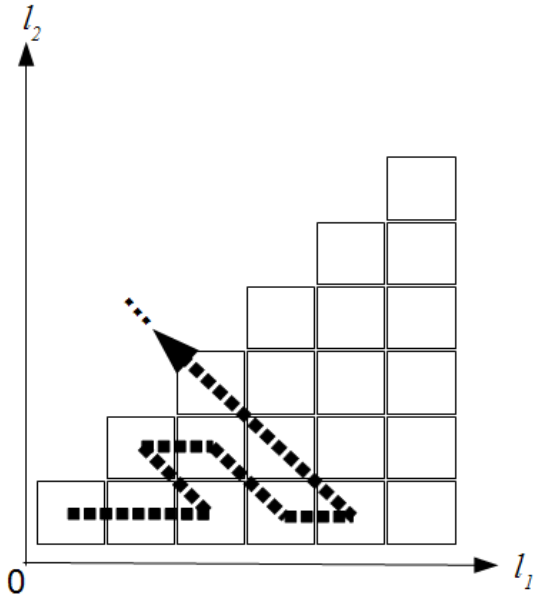

Fig. 5: Vectorization of third-order cumulants using zig-zag procedure.

Cumulant coefficients were calculated for each gait cycle, in the selected cumulant region [11, 12, 13]. Afterwards, non-vector regions were vectorized (for third- and fourth-order cumulants) using a zig-zag procedure, as shown in Figure 5. The first *d* values were taken from each vector as features. We experimented with five different values of *d* as follows: 3, 5, 10, 20 and 30. The dependence of recognition accuracy versus *d* is explained in Section 4. Thus, for each gait cycle we calculated feature vectors in *x*, *y* and *z* directions as follows:

$$\mathbf{f}_k = \left[\mathrm{Cum}_{\mathbf{x}_k}^{(r)}(l) \quad \mathrm{Cum}_{\mathbf{y}_k}^{(r)}(l) \quad \mathrm{Cum}_{\mathbf{z}_k}^{(r)}(l)\right]^{\mathrm{T}}, \tag{9}$$

where r represents the cumulant order and l represents lags. A feature vector matrix was generated from all feature vectors:

$$\mathbf{F} = [\mathbf{f}_1 \quad \ldots \quad \mathbf{f}_m]^{\mathrm{T}}, \tag{10}$$

where m is the number of all calculated feature vectors.

In order to reduce the length of feature vectors and preserve the recognition accuracy, a dimension reduction was introduced using principal component analysis (PCA) [17]. PCA implemented singular value decomposition (SVD) on the matrix **F** from Eq. (11) based on the Karhunen-Loéve transformation:

$$\mathbf{F} = \mathbf{W}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}, \tag{11}$$

and the dimension reduction was done using projection of **F** into the reduced space defined by only first *j* singular vectors, called *j*-th principal components, $\mathbf{W}_j$:

$$\mathbf{F}_j = \mathbf{W}_j^{\mathrm{T}}\mathbf{F} = \mathbf{\Sigma}_j\mathbf{V}_j^{\mathrm{T}}, \tag{12}$$

where $\mathbf{W}_j^{\mathrm{T}} = [\mathbf{w}_1 \quad \ldots \quad \mathbf{w}_m]$.

The dimension *j* was defined implicitly according to the chosen number of lags *d* and the portion of variance of data covered by the calculation of feature vectors (cumulants). We experimented with three different portions of the covered variance: 90%, 95% and 98%. The dimension *j* and portions for each selected value of *d* are shown in Figure 6.

3.5 Classification

Block diagram of the classification task is shown in Figure 7. First, the identification of gait is done by support vector machine (SVM) [14, 16]. The classification is performed by a special tool suite for machine learning, called WEKA [15]. In the SVM classification we used a Gaussian radial basis kernel function:

$$K(\vec{u}, \vec{v}) = e^{-(\vec{u}-\vec{v})^2/(2\sigma^2)}. \tag{13}$$

The classification results of our experiments are

presented in Subsection 4.1.

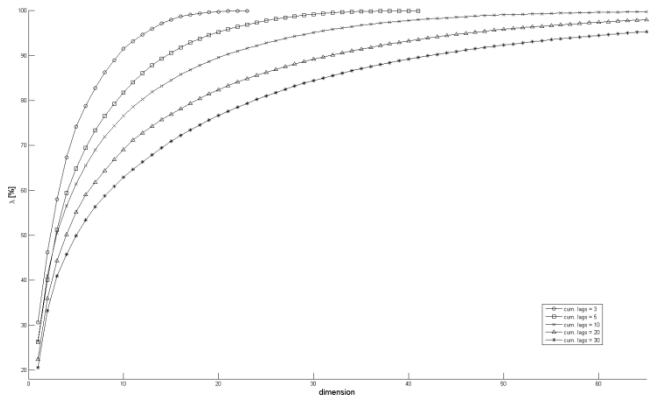

Fig. 6: Sum of the first j feature-vector components cover λ portion of data variance for different lags d.

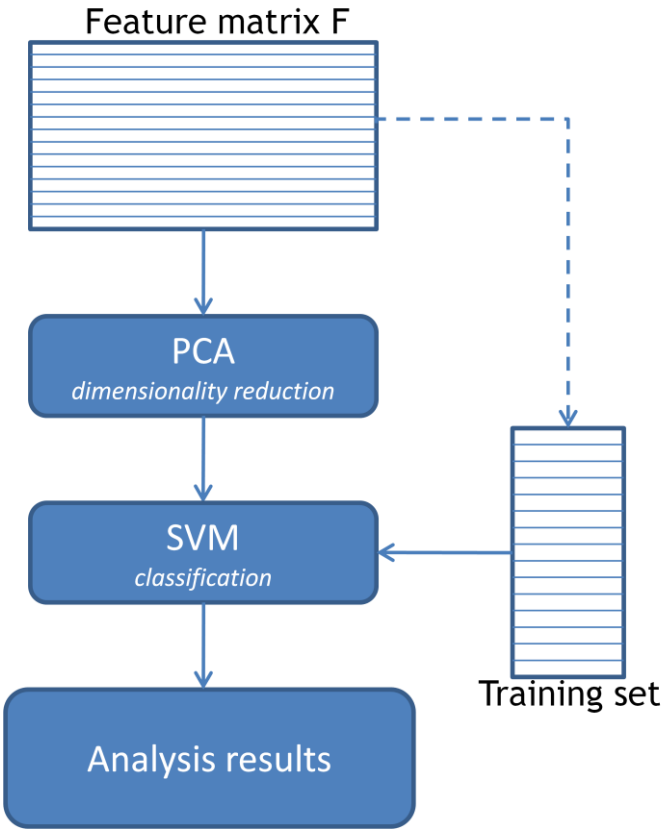

Fig. 7: Block diagram of the classification procedure using PCA and SVM.

In the second set of experiments we tried to find out how similar are walking patterns of different persons. We reused already calculated principal components (Eq. (12)). Only first three principal components were taken in to account, which can be represented as 3D points. Those points can be easily interpreted when they are plotted in 3D coordinate system as ellipsoids (results in Subsection 4.2).

The dimension reduction was done using a projection of **F** into the reduced space defined by only first 3 singular vectors, $\mathbf{W}_3$:

$$\mathbf{F}_3 = \mathbf{W}_3^{\mathrm{T}}\mathbf{F} = \mathbf{\Sigma}_3\mathbf{V}_3^{\mathrm{T}}, \tag{15}$$

where $\mathbf{W}_3^{\mathrm{T}} = [\mathbf{w}_1 \quad \dots \quad \mathbf{w}_{\mathrm{m}}]$.

As we mentioned, each feature vector is now represented as a point $\boldsymbol{w}_k$ in a 3D space. In our experiments, the points that represent the same class of gait appeared grouped together. Therefore, only the mean of these points was considered for each class for better representation:

$$\mathbf{p}_i^{(t)} = \frac{1}{\mathrm{N}_i^{(t)}} \sum_{\mathrm{k}=1}^{\mathrm{m}} \mathbf{w}_{\mathrm{k}}^{(t)}, \tag{16}$$

where $\mathrm{N}_i^{(t)}$ represents the number of points belonging to the same class (subject i and walking type t).

Distances between the calculated class centroids $\mathbf{p}_i^{(t)}$ measure the similarity between the subjects' walking patterns. The shorter the distance between two centroids, the higher is the similarity and vice versa. The standard deviation for each centroid is also calculated. Thus, each class is represented as an ellipsoid in 3D space. A class centroid defines the ellipsoid centre and each principal component standard deviation corresponds to one axis of the ellipsoid.

Using the first three principal components we cover up to 60% of the variance, as seen in Figure 6.

The results of gait similarity measurement for our experiments are presented in Subsection 4.2.

4 Results

4.1 Gait identification using SVM

Each data set prepared for classification contained 1641 feature vectors for all subjects and types of walking. A 10-fold cross-validation of the recognition accuracy was performed by WEKA for classification with SVM. The results of the classification for the cumulants calculated to lag *d=10* are shown in Tables 1 and 2. The six tested persons are labelled by capital letters *A*, *B*, *C*, *D*, *E*, and *F*, whereas their walking types are designated by attached lower-case letters *n*, *f*, and *s* for normal, fast, and slow, respectively. In Table 1, true positive (TP) and false positive (FP) rate of classification are presented. In Table 2, the obtained results are illustrated by the confusion matrix. The recognised classes are ordered vertically in comparison to the reference classes in the horizontal direction. The overall classification accuracy is 90.8±5.4%.

Class label	Number of collected cycles	TP Rate	FP Rate
An	98	0.969	0.005
Af	80	0.963	0.001
As	114	0.807	0.012
Bn	97	0.897	0.005
Bf	87	0.954	0.003
Bs	107	0.907	0.005
Cn	85	0.871	0.004
Cf	75	0.893	0.001
Cs	97	0.969	0.008
Dn	89	0.966	0.002
Df	75	0.933	0.002
Ds	101	0.802	0.018
En	81	0.951	0.004
Ef	82	0.939	0.001
Es	108	0.926	0.003
Fn	86	0.930	0.010
Ff	77	0.857	0.002
Fs	102	0.853	0.011

Table 1: Results of the classification using SVM for the cumulants calculated to lag *d=10*. Class labels are denoted by capital letters corresponding to subjects' IDs, while low-case letters which correspond to the type of walking (*n* for normal, *f* for fast, and *s* for slow walking pace).

	An	*Af*	*As*	*Bn*	*Bf*	*Bs*	*Cn*	*Cf*	*Cs*	*Dn*	*Df*	*Ds*	*En*	*Ef*	*Es*	*Fn*	*Ff*	*Fs*
An	**95**	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
Af	3	**77**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
As	0	0	**92**	0	0	0	0	0	0	0	0	22	0	0	0	0	0	0
Bn	0	0	0	**87**	0	7	0	0	0	0	0	0	0	0	0	0	0	3
Bf	0	0	0	3	**83**	0	0	0	0	0	0	0	0	1	0	0	0	0
Bs	0	0	0	4	0	**97**	0	0	0	0	0	3	0	0	0	0	0	3
Cn	2	0	0	1	0	0	**74**	1	5	0	0	0	0	0	0	0	0	2
Cf	0	0	0	0	0	0	6	**67**	0	0	2	0	0	0	0	0	0	0
Cs	0	0	0	0	0	0	0	0	**94**	0	0	2	0	0	0	0	0	1
Dn	0	0	1	0	0	0	0	0	1	**86**	0	0	0	0	0	0	0	1
Df	0	0	0	0	3	0	1	0	0	0	**70**	0	0	0	0	0	1	0
Ds	0	0	18	0	0	0	0	0	2	0	0	**81**	0	0	0	0	0	0
En	0	0	0	0	0	0	0	0	0	0	0	0	**77**	0	4	0	0	0
Ef	0	0	0	0	1	0	0	0	0	0	0	0	4	**77**	0	0	0	0
Es	1	0	0	0	0	0	0	0	0	0	0	0	3	0	**100**	3	0	1
Fn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**80**	2	4
Ff	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	6	**66**	2
Fs	2	0	0	0	0	1	0	0	3	1	0	1	0	0	0	7	0	**87**

Table 2: Confusion matrix of the classification using SVM for the cumulants calculated to lag *d=10*: recognized classes are depicted vertically, reference classes horizontally.

Tables 3 and 4 summarize the classification results according to the different lags *d* and different number of principal components. As we foretold, cumulants were calculated up to the lags *d* equal to 3, 5, 10, 20 and 30 and the portion of covered variance by PCA set at 90%, 95% and 98%. In Table 3, classification with the first *j* principal components is shown versus lag *d*. It is obvious that the number of principal components grows when increasing *d* or the portion of covered variance. In Table 4, the average recognition accuracy and its standard deviations for all subjects and their walking styles are depicted. The accuracy and its standard deviation improve, when increasing *d* and the portion of covered variance, as seen in Table 4. The same results are also illustrated in Figure 8.

The proposed method yields 90.8% recognition accuracy with feature vectors containing only 30 features. Compared to other related gait recognition methods, we can conclude that our method performs considerably well. The recognition rates of compared methods [1-6] are between 71% and 98%.

Our method was additionally tested by control gaits. A week after the first experiment we repeated the same trials with the same subjects in order to compare their gait cycles. Sets of features from the first experiment were used as training sets. Newly acquired cycles were used as test sets. The overall recognition rate was 89.7%. This means that the recognition rate, on the subject's identification basis, actually remained the same. The recognition rate of walking types for each subject was a bit lower, i.e. from 69% to 85%, due to inconsistencies in subjects' walking styles, such as a different pacing during the same walking style. We are going to tackle this issue in further investigations.

First *j* principal components		*Variance covered*		
		90%	**95%**	**98%**
Lags *d* (number of calculated cumulant coefficients)	**3 (31)**	10	13	15
	5 (49)	15	20	26
	10 (94)	21	30	41
	20 (184)	32	47	65
	30 (274)	43	63	90

Table 3: The number of the first *j* principal components in classification using SVM versus lag *d* and the portion of covered variance.

Average accuracy rates		*Variance covered*		
		90%	**95%**	**98%**
Lags used for cumulant calculation (number of calculated cumulant coefficients)	**3 (31)**	83.4±10.1%	86.4±9.8%	87.6±11.6%
	5 (49)	86.6±5.3%	89.9±6.3%	90.8±5.3%
	10 (94)	88.5±6.9%	90.8±5.4%	93.0±4.3%
	20 (184)	90.9±4.1%	93.6±3.5%	94.4±3.3%
	30 (274)	91.2±4.0%	93.0±3.9%	94.2±3.2%

Table 4: Average accuracy rate and its standard deviations of classification using SVM versus lag *d* and the portion of covered variance.

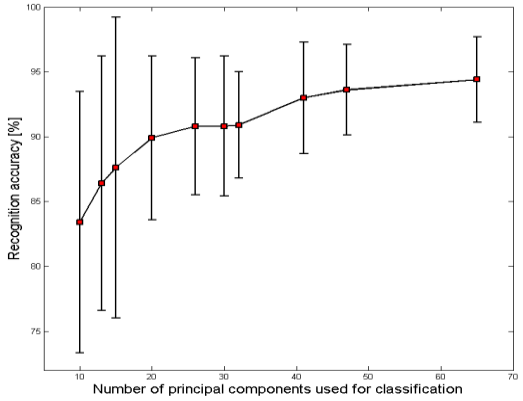

Fig 8: Recognition accuracy and its standard deviations versus the number of principal components used for classification.

4.1 Gait similarity using PCA

As explained, the distances between the gait class centroids correspond to the similarity of gait patterns. A similarity study of walking styles of all subjects is shown in Table 5. The table presents class centroids and their standard deviations of the first three principal components (designated by PC1, PC2, and PC3) in dependence of the tested subjects and their walking styles. The same results are also illustrated by class ellipsoids, as depicted in Figures 9 and 10. Both presentations confirm that similar walking patterns can be recognized when using PCA, and also the individuals can be identified referring to their gait patterns.

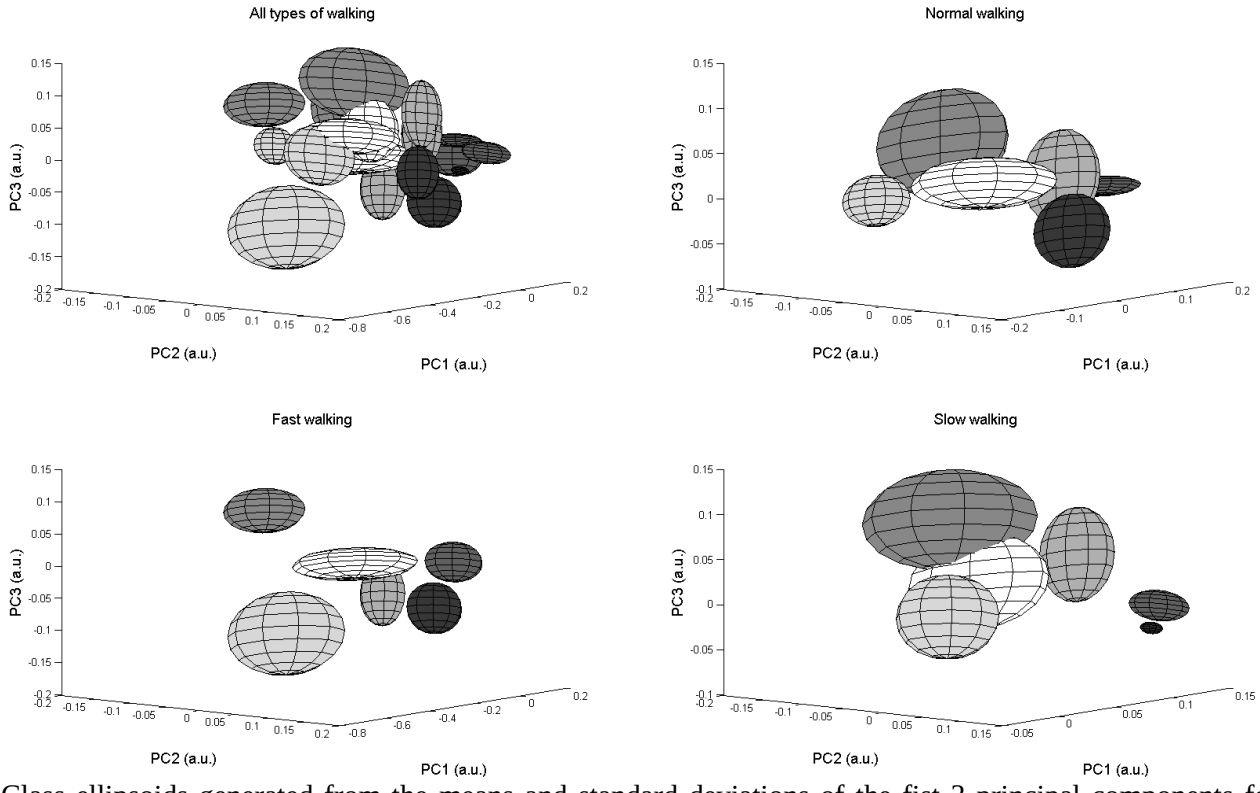

Fig. 9: Class ellipsoids generated from the means and standard deviations of the fist 3 principal components for all walking types and for normal, fast and slow walking separately.

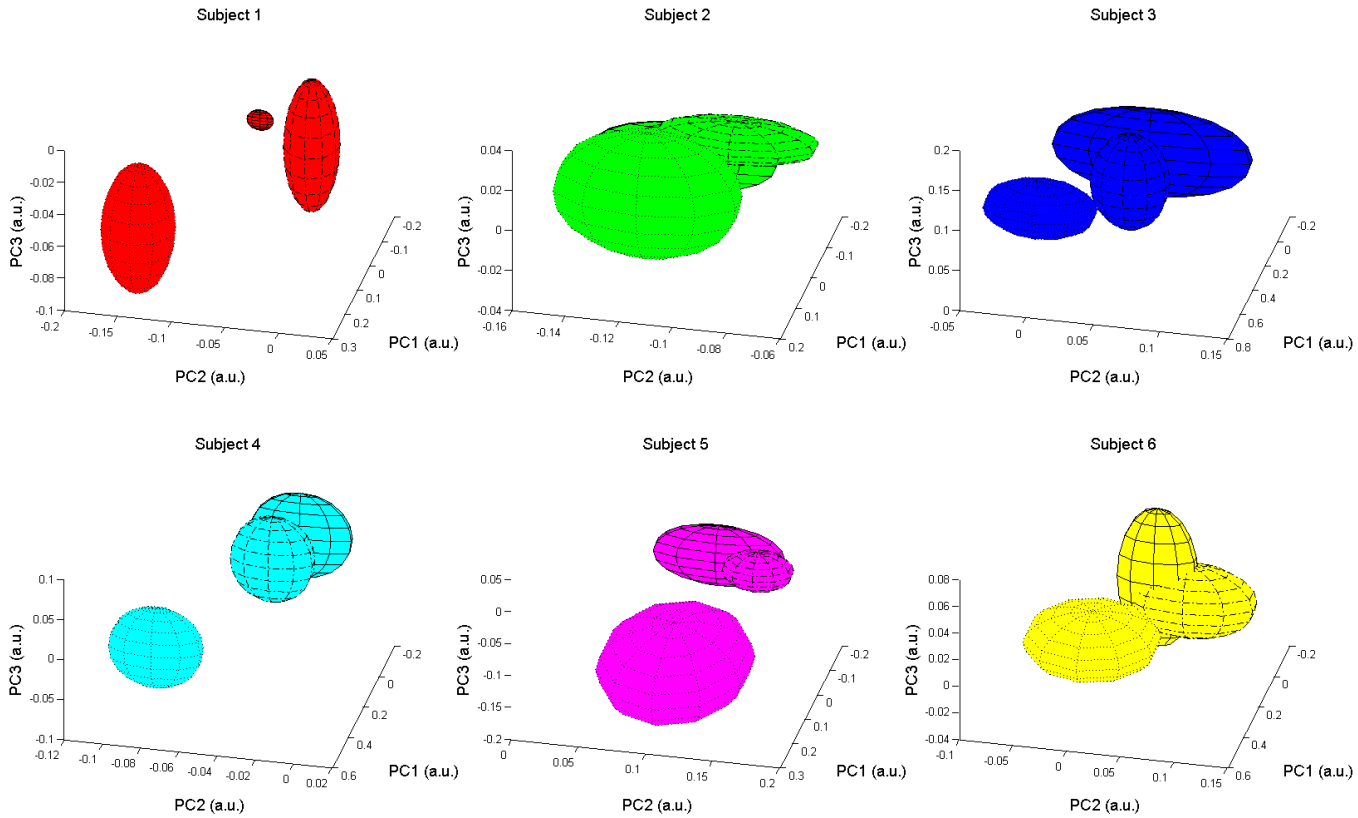

Fig. 10: Class elliposids generated from the means and standard deviations of the first 3 principal components, all three walking types for each subject. Line-bordered ellipsoids represent slow walking, dash-dotted bordered ellipsoids normal walking, and dotted bordered ellipsoids fast walking.

Class centroids of ellipsoids (means of the first 3 principal components)							
		A	*B*	*C*	*D*	*E*	*F*
n	PC1	0.0757	0.0548	-0.0585	0.0423	0.0320	0.0119
	PC2	-0.0148	-0.0857	0.0185	-0.0338	0.1430	0.0358
	PC3	-0.0041	0.0085	-0.0351	-0.0292	-0.0134	-0.0087
f	PC1	-0.0485	-0.0787	-0.3229	-0.1490	-0.1907	-0.0792
	PC2	-0.0427	-0.1311	0.0015	-0.0468	0.1501	0.0208
	PC3	0.0508	0.0683	-0.0105	0.0223	0.0800	0.0201
s	PC1	0.1105	0.0839	0.0567	0.1036	0.0764	0.0678
	PC2	-0.0486	-0.0781	0.0176	-0.0356	0.1263	0.0293
	PC3	-0.0118	0.0155	-0.0453	-0.0322	-0.0240	-0.0206
Axes of ellipsoids (standard deviations of the first 3 principal components)							
		A	*B*	*C*	*D*	*E*	*F*
n	PC1	0.0601	0.0376	0.1199	0.0545	0.0411	0.1051
	PC2	0.0421	0.0403	0.0411	0.0300	0.0349	0.0555
	PC3	0.0477	0.0220	0.0431	0.0218	0.0253	0.0421
f	PC1	0.1537	0.1098	0.2284	0.1137	0.2160	0.1751
	PC2	0.0613	0.0308	0.0454	0.0406	0.0761	0.0520
	PC3	0.0641	0.0267	0.0402	0.0290	0.0537	0.0423
s	PC1	0.0131	0.0223	0.0489	0.0157	0.0287	0.0438
	PC2	0.0155	0.0442	0.0450	0.0164	0.0378	0.0334
	PC3	0.0157	0.0182	0.0207	0.0182	0.0258	0.0250

Table 5: Means and standard deviations of the first 3 principal components (PC1, PC2, PC3).

5 Conclusion

A method for gait identification and similarity measurement based on cumulants calculated from the accelerometer data has been proposed. Classification results with 6 test subject show that the identification of people is possible with considerably high recognition rate. The obtained rate of 90.3±3.2% is high, but needs additional validation on a larger number of test subjects. We expect that the accuracy rate will slightly fall with a bigger number of test subjects. Further investigations are possible, for example, to find the correlation between subject's characteristics and their walking styles (a possible influence of their height on the walking pattern). Our final objective, however, is to apply the proposed method to identify walking disorders of subjects, which could indicate their health problems.

References:

[1] D. Gafurov, E. Snekkes, P. Bours, Gait Authentication and Identification Using Wearable Accelerometer Sensor, *IEEE Workshop on Automatic Identification Advanced Technologies*, 2007, pp. 220-225

[2] D. Gafurov, K. Helkala, T. Soendrol, Biometric Gait Authentication Using Accelerometer Sensor, *Journal of Computers*, Vol.1, No. 7, 2006, pp. 51-59

[3] A. Annadrhorai, E. Guenterberg, J. Barnes, K. Haraga, R. Jafari, Human Identification by Gait Analysis, *Proceedings of the 2nd International Workshop on Systems and Networking Support*

for Health Care and Assisted Living Environments, No.11, 2008

[4] C. Y. Lee, J. J. Lee, Estimation of Walking Behavior Using Accelerometers in Gait Rehabilitation, *International Journal of Human-friendly Welfare Robotic Systems*, Vol.3 No.2, 2002, pp. 32-36

[5] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S. M. Makela, H. A. Ailisto, Identifying users of portable devices from gait pattern with accelerometers, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol.2, 2005, pp. 973-976

[6] T. Iso, K. Yamazaki, Gait Analyzer Based on a Cell Phone With a Single Three-axis Accelerometer, *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, 2006, pp.141-144

[7] D. Gafurov, A Survey of Biometric Gait Recognition: Approaches, Security and Challenges, *Nik-2007 Conference*, 2007

[8] A. Mostayed, S. Kim, M. M. G. Mazumder, Foot Step Based Person Identification Using Histogram Similarity and Wavelet Decomposition, *International Conference on Information Security and Assurance*, 2008, pp. 307-311

[9] H. Ying, C. Silex, A. Schnitzer, S. Leonhardt, M. Schiek, Automatic Step Detection in the Accelerometer Signal, *4th International Workshop on Wearable and Implantable Body Sensor Networks*, 2007, pp. 80–85

[10] L. Wang, T. Tan, W. Hu, H. Ning, Automatic Gait Recognition Based on Statistical Shape Analysis, *IEEE Transaction on Image Processing*, Vol. 12, No. 9, 2003, pp. 1120-1131

[11] J. M. Mendel, Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications, *Proceedings of the IEEE*, Vol. 79, No. 3, 1991, pp. 278-305

[12] C. W. Therrien, Discrete Random Signals and Statistical Signal Processing, *Prentice-Hall International*, 1992

[13] C. L. Nikias, A. P. Petropulu, Higher-Order Spectra Analysis, *PTR Prentice Hall*, 1993

[14] J. C. Platt, Fast training of support vector machines using sequential minimal optimization, *Advances in Kernel Methods: Support Vector Learning*, 1999, pp. 185-208

[15] I. H. Witten, E. Frank, Data Mining: Practical machine learning tools and techniques, 2nd Edition, *Morgan Kaufmann*, 2005

[16] J. Wu, An Advanced Hybrid Machie Learning Approach for Assessment of the Change of Gait Symmetry, *WSEAS Transactions on Computers*, Vol. 8, No. 9, 2009, 1522-1532

[17] N. Pessel, J. F. Balmat, Principal Component Analysis for Greenhouse Modelling, *WSEAS Transactions on Systems*, Vol. 7, No. 1, 2008, 24-30

[18] R. Muscillo, S. Conforto, M. Schmid, T. D'Alessio, Minimizing the Set Up for ADL Monitoring through DTW Hierarchical Classification on Accelerometer Data, *WSEAS Transactions on Biology and Biomedicine*, Vol. 5, No. 3, 2008, 47-53