Classification of Acceleration Data for Biometric Gait Recognition on Mobile Devices

Claudia Nickel, Holger Brandt, Christoph Busch
Hochschule Darmstadt - CASED, Darmstadt, Germany,
c.nickel@fbi.h-da.de

Abstract: Ubiquitous mobile devices like smartphones and tablets are often not secured against unauthorized access as the users tend to not use passwords because of convenience reasons. Therefore, this study proposes an alternative user authentication method for mobile devices based on gait biometrics. The gait characteristics are captured using the built-in accelerometer of a smartphone. Various features are extracted from the measured accelerations and utilized to train a support vector machine (SVM). Among the extracted features are the Mel- and Bark-frequency cepstral coefficients (MFCC, BFCC) which are commonly used in speech and speaker recognition and have not been used for gait recognition previously. The proposed approach showed competitive recognition performance, yielding 5.9% FMR at 6.3% FNMR in a mixed-day scenario.

1 Introduction

Phone theft is becoming a serious security problem in many countries. For example in the United Kingdom about 228 mobile phones were reported stolen every hour in 2010, which lead to a call by the Minister of Crime Prevention to the mobile phone industry to better protect device owners against theft [CJ10]. A perpetrator having access to private and business emails, contacts and social networks can easily impersonate the victim. Although well-known technical security measures like PIN, password and strong encryption would protect the sensitive information, data protection techniques and knowledge based authentication with PIN and password are often not applied by the owner of the device even though they are available at essentially no additional cost.

There have been few phones with fingerprint scanners that never really entered mass market. The reasons are probably the rather high costs for the extra sensor that is not needed by the average end-user. Other biometric modalities, namely speaker, face and gait recognition, do not have this problem as most modern phones are capable of realizing a biometric verification system using one or more of these modalities. Compared to the mentioned modalities, gait recognition has one unique advantage: it does not require explicit user interaction during the verification process as the phone does it literally on-the-go. The driving motivation for gait recognition is that the device continuously authenticates the owner, when he is on the move and thus more rarely requires an explicit user authentication.

From a technological point of view research in biometric gait recognition can be categorized into three main approaches (analog to Gafurov in [Gaf07]): Machine Vision Based, Floor Sensor Based and Wearable Sensor Based Gait Recognition. This study focuses on the last approach, wearable sensor (WS) based gait recognition.

WS-based gait recognition is the most recent form of gait recognition as lately proposed as 2005 [ALM$^+$05] and further developed by Gafurov [Gaf08]. WS-based gait signals are captured by a sensor that is attached to the body, typically an accelerometer. Accelerometers used for gait recognition generally are tri-axial i.e. they measure acceleration in three spatial dimensions (backward-forward, sideways and vertical). The acceleration signals acquired are the result of the acceleration of the person's body, gravity, external forces like vibration of the accelerometer device and sensor noise.

When research in WS-based gait recognition started, dedicated sensors were used for data collection. But the interesting aspect of WS-based gait recognition is that accelerometers are in the meantime a standard component of mobile devices like smartphones and tablets (e.g. iPad). Because of the ubiquitous nature of mobile smartphones with accelerometers an ideal "off-the-shelf" hardware platform for gait analysis and recognition is already available. The strong advantage is that only new software (yet another "app") needs to be developed with no additional hardware cost and customizations opposed to biometric-only sensors like fingerprint readers. Therefore, recently researchers started using smartphone equipment [DNBB10, FMP10, KWM10, NBRM11, SZ09] for data collection.

2 Data Set

For the experiments the data set used in [DNBB10] and [NBRM11] was employed. The acceleration data was collected using a G1 smartphone with a customized application to access the accelerometer measurements and to output the data from the sensor to a file (40-50 data points per second for each of the three directions x, y and z). While recording the gait data the phone has been placed in a pouch attached to the belt of the subject (see figure 1). In total, data of 48 healthy subjects was successfully recorded on two sessions at two different days with the subjects wearing their usual shoes and walking at normal pace. Subjects were walking straight on flat floor for around 37 meters, turned around at the end and walked back (resulting in two data sets called "walk"). Age and gender distribution are given in table 1.

	< 20	$20-24$	$25-30$	> 30	unknown
male	1	1	25	9	2
female	0	5	4	0	1
total	1	6	29	9	3

Table 1: Age and gender distribution of data subjects.

Figure 1: Position of phone during data collection.

3 Feature Extraction

When deriving gait features from the captured signal there are two possible approaches: cycle-based and non-cycle-based feature extraction. A gait cycle physically corresponds to two consecutive steps that the subject has taken, i.e. the period after one foot touches the ground until the same foot touches the ground again. Cycle-based features are computed by identifying gait cycles in time-series data representing a walking person. Then the feature extraction is conducted on identified cycles and the resulting features are used for biometric template creation and sample comparison. Currently this approach for representing gait is the predominantly used method in gait recognition literature. As an alternative approach, a non-cycle-based gait representation, was used for this study. Here, features are extracted from the times-series data from a selected time window without prior identifying the contained gait cycles.

The collected gait samples are preprocessed such that the feature extraction algorithm works with consistent and portioned data. The first step is a linear interpolation to a fixed sampling rate as this is not given with the collected raw data. The average sampling rate of the raw data is about 40-50 data points per second. After interpolation the signals s_x, s_y, s_z are normalized by the mean acceleration μ_a of the respective acceleration direction s_a: $\bar{s}_a(t) = s_a(t) - \mu_a, a \in x, y, z$.

The normalized acceleration signals are then separated into parts of several seconds using a sliding window approach with overlapping rectangular windows. This means that the original signal of length l is splitted into *segments* of length t and distance d between consecutive segments. The remaining part is dropped and not used any further. The segmentation is done for all normalized signals. Note that actually three normalized acceleration signals $\bar{s}_x, \bar{s}_y, \bar{s}_z$ are segmented.

The segments at this stage are still represented as time series. As the intention was to benefit from the well-performing classification capabilities of SVMs a transformation to a fixed length vector of discrete values has to be conducted. For each segment one feature vector is created. As a starting point statistical features were calculated for the acceleration signals: mean, maximum, minimum, binned distribution (relative histogram distribution in linear spaced bins between the minimum and the maximum acceleration in the segment), root mean squared acceleration (rms, the square root of the mean of the squares of the acceleration values of the segment), standard deviation and zero crossings (number of sign changes in the segment).

Further used features were the Mel-frequency cepstral coefficients (MFCC) and Bark-frequency cepstral coefficients (BFCC), which belong to the most widely used spectral representations of audio signals for automatic speech recognition and speaker verification [GFK05]. The general workflow for creating MFCC is laid out in figure 2. A more elaborate discussion can be found in [RJ93]. BFCC are created similar to MFCC, the only difference is that instead of the Mel-scale the Bark-scale is applied.

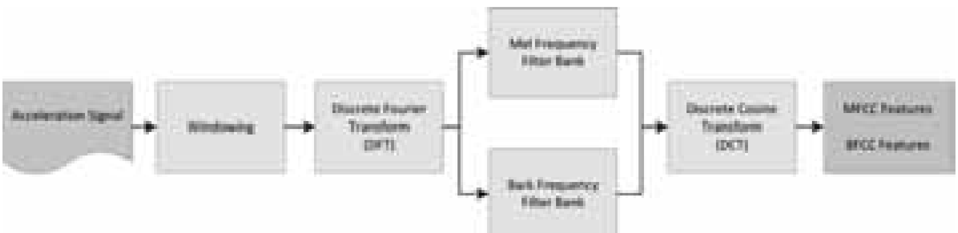

Figure 2: MFCC and BFCC feature creation

4 Preliminary Evaluation

We used SVMs as classifiers, which were introduced by Vapnik in 1982 [Vap82] as a supervised learning method based on the theory of structural risk minimization. A SVM is a classifier which is inherently a solution for two class problems. The basic idea of the SVM is to construct a hyperplane as the decision plane, which separates the patterns of the two classes with the largest margin. The experiments were carried out using the SVM implementation LIBSVM [CL01]. The Radial Basis Function (RBF) kernel was chosen and a parameter selection (grid search [HCL03]) was done as part of the optimization process.

In a first step the discrimination capabilities of single features and of combined features were investigated. The acceleration directions were also used separately to study their different contribution to the classification result.

The acceleration signals collected on the first day were used for training and the signals collected on the second day were used for calculation of the recognition performance (cross-day scenario). First of all the data was interpolated to 100 Hz. The segment length t was 7500 ms, the segment distance d was 3750 ms (corresponding to an overlap of 50%). Further interpolation rates and segment lengths are tested later on. Last, the feature extraction was conducted for various feature types. To determine the biometric performance, cross-validation was conducted calculating the false match rate (FMR) and the false non-match rate (FNMR).

Five configurations were tested for each feature type using different data sources, namely the x-, y-, z-acceleration, the resulting acceleration, and all of them combined. The re-

Feature	X		Y		Z		Magn.		Combined	
	FMR	FNMR	FMR	FNMR	FMR	FNMR	FMR	FNMR	FMR	FNMR
max.	0.5%	96.9%	0.5%	98.2%	1.9%	96.7%	1.0%	97.5%	4.2%	82.1%
min.	0.5%	99.3%	0.9%	97.5%	0.4%	97.8%	1.2%	98.2%	4.7%	77.4%
mean	0.0%	99.8%	1.1%	99.1%	0.2%	99.5%	0.5%	98.4%	3.6%	89.4%
std. dev.	0.3%	97.5%	0.6%	96.0%	0.3%	100.0%	0.2%	98.7%	4.2%	81.0%
rms	0.9%	96.2%	0.6%	95.4%	0.3%	100.0%	0.5%	97.4%	4.6%	86.7%
zero cross.	0.2%	99.3%	0.1%	99.1%	0.6%	99.5%	0.0%	100.0%	5.3%	85.8%
bin dist.	6.2%	87.2%	5.7%	90.0%	4.7%	91.1%	4.5%	91.4%	5.2%	76.8%
MFCC	6.1%	76.3%	3.6%	81.4%	4.1%	82.1%	4.2%	78.8%	1.5%	72.1%
BFCC	6.6%	74.5%	4.6%	77.9%	3.5%	82.9%	3.8%	76.3%	1.5%	67.9%

Table 2: Evaluation of discrimination capabilities of single features

Set	Used Features	Length	FMR	FNMR
1	BFCC, MFCC	104	1.3%	69.3%
2	BFCC, MFCC, bin dist., min.	128	1.2%	69.0%
3	BFCC, MFCC, bin dist., min., std. dev.	132	1.2%	69.0%
4	BFCC, MFCC, bin dist., min., std. dev., max.	136	1.3%	68.8%
5	BFCC, bin dist., min.	76	1.3%	67.0%
6	BFCC, bin dist., min., std. dev	80	1.5%	66.2%
7	BFCC, bin dist., min., std. dev., max.	84	1.5%	65.3%
8	BFCC, mean, max., min., std. dev., rms, zero cross., bin dist.	96	1.2%	66.8%
9	BFCC, mean, max., min., std. dev., rms, zero cross.	76	0.9%	67.5%
10	BFCC, mean, max., min., std. dev., rms	72	1.4%	66.6%
11	BFCC, max., min., std. dev., rms	68	1.7%	66.2%

Table 3: Evaluation of combined features' discrimination capabilities.

sulting vector is also referred to as the magnitude vector and is calculated as follows: $\bar{s}_{res}(t) = \sqrt{\bar{s}_x(t)^2 + \bar{s}_y(t)^2 + \bar{s}_z(t)^2}$. Here $\bar{s}_x(t), \bar{s}_y(t), \bar{s}_z(t)$ are the interpolated and normalized accelerations measured in the corresponding directions at time t. The results are given in table 2.

All feature types have a resulting vector length of one. Exceptions are the binned distribution, were five bins were used and thus a vector of length five is created. For MFCC and BFCC the common number of 13 coefficients was generated for each segment. Of course, when the features are combined (generated for each orientation and the magnitude as well) the length is four times this number.

It is apparent that the best performances are yielded when all sensor orientations and the magnitude as well is used. Now, various combinations of the features are tested. The results are presented in table 3. The first four feature sets consist of combinations of the best performing single features. One can see that the results are basically the same. When the MFCC coefficients are removed from these feature sets (resulting in sets 5 to 7), the FNMR decreases while the FMR stays on the same level. Therefore, further combinations of BFCC and several statistical features were tested (set 8 to 11). The obtained results are

	5000		7500		10000	
Set	FMR	FNMR	FMR	FNMR	FMR	FNMR
5	1.7%	70.3%	1.3%	67.0%	1.2%	64.7%
6	1.8%	70.5%	1.5%	66.2%	1.1%	65.0%
7	1.7%	69.6%	1.5%	65.3%	1.1%	64.7%
8	0.9%	67.8%	1.2%	66.8%	0.9%	64.2%
9	1.1%	66.3%	0.9%	67.5%	1.3%	59.9%
10	1.3%	69.1%	1.4%	66.6%	1.4%	63.6%
11	1.2%	69.3%	1.7%	66.2%	1.3%	62.3%

Table 4: Evaluation results for segments of length 5,000, 7,500 and 10,000 and an overlap of 50% (frequency = 100).

	50		100		200	
Set	FMR	FNMR	FMR	FNMR	FMR	FNMR
5	1.0%	67.3%	1.2%	64.7%	1.1%	64.4%
6	1.2%	65.7%	1.1%	65.0%	1.1%	64.7%
7	1.0%	65.2%	1.1%	64.7%	1.2%	64.4%
8	0.9%	65.4%	0.9%	64.2%	0.8%	63.9%
9	1.1%	63.0%	1.3%	59.9%	1.2%	61.5%
10	1.4%	64.3%	1.4%	63.6%	1.5%	63.6%
11	1.3%	63.3%	1.3%	62.3%	1.5%	62.8%

Table 5: Evaluation results for interpolation rates of 50, 100 and 200 samples per second and a segment size of 10000 ms.

similar to the ones when combining BFCC with the best performing statistical features, but feature set 7 remains the best one.

For the seven best performing features sets (set 5 to 11), we evaluated the influence of the segment size. The overlap of the segments is set to 50%, as before. Table 4 gives the results for the new segment lengths 5,000 and 10,000, and the previously used length of 7,500 ms. One can see that the segment length of 10,000 performs best. No further segment lengths were evaluated as the length is limited by the duration of one walk. 10,000 ms is the maximal segment length to assure that at least three segments can be extracted from each walk such that enough testing data is available.

For the same feature sets, the interpolation frequency was varied from 100 to 50 and 200 samples per second. The results are given in table 5. The results of frequency 100 and 200 are nearly the same and slightly better than the ones of frequency 50. The best result is a FNMR of 59.9% at a FMR of 1.3% obtained when using feature set 9, a segment size of 10,000 ms and an interpolation rate of 100 samples per second.

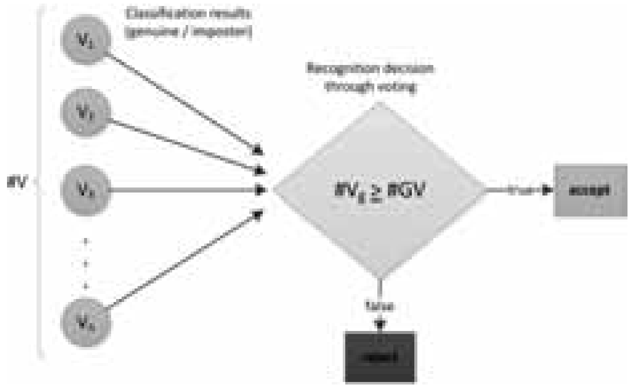

Figure 3: Quorum voting scheme

5 Introduction of Voting Scheme

With the yielded recognition performance it is very likely that a genuine is incorrectly rejected. An imposter on the other side is rarely mis-classified and thus seldom falsely accepted. A post-processing approach that reduces the number of false rejects is to use multiple classifications for one recognition decision while incorporating a different confidence in the classification correctness.

More specifically one uses $\#V$ segments of a probe gait signal instead of only one segment for the recognition. For each segment the classification is carried out as usually. Then, the $\#V$ results are combined. An imaginable straightforward approach is majority voting, but it is not likely to perform well as there is such a large difference between the two error rates. Therefore a *quorum* voting for a genuine is implemented, which is inspired by a petition quorum.

This quorum requires that at least $\#GV$ positive classification results are obtained for an accept, otherwise the probe signal is rejected. The described concept is visualized in figure 3. Note that $\#V_g$ is the number of results that classify the respective segment as stemming from the genuine, in other words $\#V_g$ is the number of *votes* for genuine.

Of course, while the FNMR is decreased by this approach the number of false accepts and thus the FMR increases. We conducted a series of experiments with the intention to find a balanced setting where both error rates are in the same range.

6 Voting Results

The voting results presented in table 6 are reported separated by three experiment setups:

Cross-day Data of the first session is used for training, data from the second session is used for testing (as in preliminary evaluation)

Same-day Half of the data of one session is used for training, the other half from the same session is used for testing

Mixed Half of the data from both sessions is used for training, the other half of the data from both sessions is used for testing

	Cross-day		Same-day		Mixed	
Set	FMR	FNMR	FMR	FNMR	FMR	FNMR
5	4.7%	35.4%	0.3%	29.2%	3.6%	4.2%
6	4.1%	37.5%	0.5%	14.6%	2.6%	4.2%
7	4.1%	37.5%	0.3%	14.6%	4.6%	4.2%
8	3.2%	37.5%	1.0%	10.4%	4.0%	4.2%
9	5.0%	25.0%	1.0%	16.7%	5.9%	6.3%
10	4.9%	27.1%	1.3%	20.8%	7.0%	6.3%
11	4.7%	25.0%	0.8%	20.8%	6.8%	6.3%

Table 6: Gait recognition performance (with voting)

Various settings were evaluated to identify the optimal combination of votes $\#V$ and genuine votes $\#GV$. Best results have been obtained when using $\#GV = 1$ and the maximal amount of segments for $\#V$. For the cross-day and mixed setting this results in $\#V = 6$, whereas for the same-day setting $\#V$ was only 3 because of the reduced amount of data available.

The reason for analysing the three different setups is to get an impression of the impact of time on the recognition results. The most relevant setup for a practical application is the cross-day performance, as in general the enrolment will take place on a different day than the authentication. We consider feature set 9 as the best setting, as it performed well in all evaluations and provides one of the best cross-day performances of 5.0% FMR and 25.0% FNMR. One can see that the error rates greatly decrease when no separation between the days of collection is made. For the mixed-day setup we obtained a FNMR of 6.3% at a FMR of 5.9% which shows the suitability of SVMs for classification of gait data. The difference to the same-day scenario indicates that having a larger training database with a greater intra-class variability results in better trained SVMs and hence in better recognition rates.

The mixed-day results can be compared to some extend to our previous results using the same database. Unfortunately the partition into training and testing data has not been the same in the three evaluations. In [DNBB10] a cycle extraction method was applied to the data. Reference cycles were extracted from the first walk collected on the first day, probe cycles were extracted from the remaining three walks. In this mixed-day scenario we obtained an equal error rate (EER) of 20.1%, which is much higher than the SVM results obtained in this paper. The second evaluation on that database used Hidden Markov Models for classification [NBRM11]. All data from the first session and parts of the second session were used for training, the remaining parts of the second session were used for testing. This resulted in an EER of approximately 10%. These results underline the great performance of the SVMs for classification of biometric gait data. Nevertheless, a fair comparison of the stated methods is necessary.

7 Conclusion and Future Work

Biometric gait recognition using wearable sensors like accelerometers is a young research area with promising future applications. It offers the unique advantage of a truly unobtrusive capturing of a biometric characteristic and is especially interesting for mobile devices that nowadays are sold with a suitable sensor already embedded.

With future applications in mind the experiments in this study carefully distinguish between the recognition performance on the same day and the recognition performance cross-days. That this distinction is necessary becomes obvious by the good results with mixed and same-day tests compared to a rather weak performance of cross-day tests. Unfortunately this has not been sufficiently addressed in most of the other gait recognition studies using wearable sensors. Therefore it is questionable if the reported performances were capable of providing a good cross-day performance.

In regard to the work carried out for this study the next task will be the porting of the developed program to the Android platform. One limitation of the current approach is that the training of the SVM requires feature vector instances of imposters which implies that a complete product would need to ship with a database with the feature instances. Although this is not impractical, from a privacy point of view it would be preferable to have a pre-trained SVM that is only further trained with the instances of the genuine user during enrolment. Therefore incremental SVM learning approaches (e.g. [WZC08]) should be evaluated as well. I addition, a practical solution should incorporate activity recognition to use only walking data for classification.

In future we will analyze the performance of SVMs on a larger database, allowing a better training of the classifier. Another open question is whether the training of the SVMs can yield a better generalization in terms of a tolerance against inter-day gait variability when data from several days is used for training. To answer this question a multi-day gait database is needed. Preferably this database will also contain different walking conditions (shoes, underground etc.) to analyse their influence on the recognition performance. In addition, we will focus on a fair comparison of the performance of SVMs, HMMs and cycle extraction methods.

8 Acknowledgments

This work was kindly supported by CASED (www.cased.de). The authors thank the numerous volunteers that participated in the gait data collection.

References

[ALM$^+$05] H.J. Ailisto, M. Lindholm, J. Mäntyjärvi, E. Vildjiounaite, and S.-M. Mäkelä. Identifying people from gait pattern with accelerometers. In A. K. Jain and N. K. Ratha, editors, *Biometric Technology for Human Identification II*, volume 5779, pages 7–14. SPIE, 2005.

[CJ10] R. Cellan-Jones. Government calls for action on mobile phone crime. BBC News, February 2010. Online; http://news.bbc.co.uk/2/hi/technology/8509299.stm; last accessed 30th May 2011.

[CL01] 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.

[DNBB10] M.O. Derawi, C. Nickel, P. Bours, and C. Busch. Unobtrusive User-Authentication on Mobile Phones using Biometric Gait Recognition. In *Proceedings of the Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2010.

[FMP10] J. Frank, S. Mannor, and D. Precup. Activity and Gait Recognition with Time-Delay Embeddings. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, 2010.

[Gaf07] D. Gafurov. A Survey of Biometric Gait Recognition: Approaches, Security and Challenges. In *Proceedings of Annual Norwegian Computer Science Conference*, 2007.

[Gaf08] D. Gafurov. *Performance and security analysis of gait-based user authentication*. PhD thesis, Faculty of Mathematics and Natural Sciences, University of Oslo, 2008.

[GFK05] T. Ganchev, N. Fakotakis, and G. Kokkinakis. Comparative evaluation of various MFCC implementations on the speaker verification task. In *Proceedings of the International Conference on Speech and Computer*, 2005.

[HCL03] C. W. Hsu, C. C. Chang, and C. J. Lin. *A practical guide to support vector classification*. Department of Computer Science, National Taiwan University, Taipei, Taiwan, 2003.

[KWM10] J. R. Kwapisz, G. M. Weiss, and S. A. Moore. Cell Phone-Based Biometric Identification. In *Proceedings of the IEEE Fourth International Conference on Biometrics: Theory, Applications and Systems*, 2010.

[NBRM11] C. Nickel, C. Busch, S. Rangarajan, and M. Möbius. Using Hidden Markov Models for Accelerometer-Based Biometric Gait Recognition. In *7th International Colloquium on Signal Processing & Its Applications*, 2011.

[RJ93] L. Rabiner and B.-H. Juang. *Fundamentals of speech recognition*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[SZ09] S. Sprager and D. Zazula. A cumulant-based method for gait identification using accelerometer data with principal component analysis and support vector machine. *WSEAS Transactions on Signal Processing*, 2009.

[Vap82] V. Vapnik. *Estimation of Dependences Based on Empirical Data: Springer Series in Statistics (Springer Series in Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1982.

[WZC08] Y. Wang, F. Zhang, and L. Chen. An Approach to Incremental SVM Learning Algorithm. In *ISECS International Colloquium on Computing, Communication, Control, and Management*, 2008.