Short Term Template Aging Effects on Biometric Dynamic Handwriting Authentication Performance

Tobias Scheidat, Karl Kümmel, Claus Vielhauer

Short Term Template Aging Effects on Biometric Dynamic Handwriting Authentication Performance

Tobias Scheidat[1,2], Karl Kümmel[1,2], Claus Vielhauer[1]

[1]Brandenburg University of Applied Sciences, Germany
[2]Otto-von-Guericke University of Magdeburg, Germany
{tobias.scheidat,karl.kuemmel,claus.vielhauer}@fh-brandenburg.de

Abstract. In biometrics the variance between data acquired from the same user and same trait is not only based on different sensors or user's form of the day, but it also depends on an aging factor. Over time the biological characteristics of a human body changes. This leads to physical and mental alternations, which may have significant influence on the biometric authentication process. In order to parameterize a biometric system, the study of the degree of aging's influence is an important step. In this paper we provide an experimental evaluation on the influence of changes of handwriting biometrics by acquiring data from writers in three sessions with a time difference of one month each. The aim is to analyze the potential impact of aging processes on different written content within a biometric handwriting system in terms of authentication performance. In the worst case, the equal error rate determined on verification data acquired two month after the reference data ($EER = 0.162$) is four times higher than the equal error rate calculated based on reference and verification data from the first session ($EER = 0.041$).

Keywords: biometrics, handwriting, template aging, verification

1 Introduction

In order to increase the security of IT systems, the need for secure automatic user authentication methods is growing steady. The main goal is to protect information and/or property against theft, unauthorized manipulation and destruction. In order to do so, there are three main methods for user authentication: secret knowledge, personal possession and biometrics. On one side, for secret knowledge and personal possession, authentication object (e.g. password and/or smart card) can be lost, stolen or handed over to other unauthorized persons. On the other side, an advantage is that there is no fuzziness of validity of the authentication object, i.e. either it is the right one or it is not. The authentication object in biometric systems is a physiological characteristic of a person (static: e.g. face, fingerprint) or a trait caused by a person's behavior (dynamic: e.g. handwriting, speech). Thus, it is directly linked with the body or the behavior of a person and theft, loss or hand over is not possible in an easy way. Because of the variability of data from the same person (intra-class variability) or similarities between data of different persons (inter-class similarity), biometric systems lack of false recognition probabilities. Another problem that causes intra-class variability is the change of a human body and mind due to biological aging. For this reason, it can contribute to problems caused by poorer representations of biometric characteristics or by difficulties in use of biometric systems.

There are two main reasons to study the influence of biological aging effects on the verification performance of biometric systems. Firstly, aging affects the human body and therefore most of the physical and/or mental characteristics. Those changes are caused by growth and biological aging processes, which could have an impact on both, acquisition of a biometric trait and the recognition based on it. Additionally, the biological aging of a person influences also the mental skills that may decrease the abilities to handle technical devices such as a biometric system. Secondly, the lifespan of human beings grows in nearly each industrialized county worldwide. For example, from the observation of the European demographic deployment in the last 60 years and the forecast for the next 40 years [1] it can be seen that a demographic change is in progress that leads to an aging population. In 1950, proportion of humans having an age of 65 and higher, amounts 8.22%. Up today (2010) this proportion is enhanced to 16.18%. For year 2050 the demographic forecast predicts an increase to 26.92%. Based on these calculations, a threefold increase in the proportion of older people in the European population within 100 ($1950 - 2050$) years is expected.

In this paper we focus on an experimental evaluation of the influence of aging effects on dynamic handwriting biometrics. The evaluation is carried out on handwriting data, which is acquired in three different monthly sessions. In our experiments, we also evaluate alternative written contents (so-called semantics) in relation to handwritig based verification which are a *Given* and a *Secret PIN*, a *Pseudonym*, a free chosen *Symbol* and *Place* as the answer to the question 'Where are you from?'. Based on these initial evaluations we do not only show the impact of aging on verification performance, but also a tendency: In any tested case the equal error rate decreases significantly, if the enrollment data from session one is tested with verification data acquired in session two and session three, accordingly.

This paper is structured as follows: The next section concludes some general aging effects on biometric modality of handwriting. In section three the Biometric Hash algorithm as reference method is described as well as the evaluation methodology and the test setup. In the same section the evaluation results are presented and discussed. The fourth section concludes the paper and gives a short overview on future work in this research area.

2 Aging and biometrics

Most biometric evaluations regarding aging affects were published in context to the evaluation of the biometric passport (e.g. see [2], [3]). Often for this documents the leading biometrics fingerprint, face and iris were tested and used for authentication. Reported results are not or only hard to compare due to inconsistent test setups. Additionally to different sensors, software and authentication performance measurements; almost every evaluation study uses their own biometric database with varying number of users and different aging levels. For that reasons, the following overview of different studies on handwriting from the perspective of aging effects can only show a tendency.

The handwriting is changing with increasing age. Birren and Botwinick [4] observed age-related changes in the writing velocity. For people with an age of mid to late $50's$ the speed of the writing process decreases. Further, elderly need more time for the writing process [6], write with lesser speed and pressure [5], and produced more unstable movements compared to younger persons [6]. On the other hand, younger people write with a higher speed and have a smoother writing process, which is characterized by less change in the pen velocity [7]. In [8] Guest investigates a possible age dependency of biometric signature verification systems based on dynamic handwriting features. The author reports that there are no significant differences between the age groups tested in the ability to enroll and to verify based on a biometric handwriting system. Clear differences between age groups are found in those handwriting characteristics that affect the execution time and the pen dynamics (such as velocity and acceleration). With increasing age of the test subjects the pen velocity and acceleration decrease. In contrast, the writing time and the number of pen up events increase with the age. In addition, Guest states that the reproducibility of the signature did not vary significantly with the age. According to Michel [9], even the signature is more resistant to disturbances than other written content. The degree of the influence of age and disease specific degradation phenomena is also smaller for the signature. Michel states that this is based on the frequently execution of the signature, which automates its writing more than the rest of the writing processes.

3 Experimental evaluation

To show the influence of aging described in this paper the evaluation data was acquired in three sessions with a time difference of one month. This subsection describes the function of the Biometric Hash algorithm for dynamic handwriting from [10], used as reference method, as well as the methodology and the setup used for evaluation of the verification performance from the aging effects point of view. The evaluation results are presented and discussed in the fourth subsection.

3.1 Biometric Hash algorithm for handwriting

Basis of the evaluation carried out in this paper is the Biometric Hash algorithm, which is described by Vielhauer in [10], originally developed to generate unique hashes from dynamic handwriting data. Generally, sensors for dynamic handwriting biometrics (e.g. tablet PC, signature tablet) provide time dependent signals for the horizontal and vertical pen position, pen tip pressure and partly pen orientation angles azimuth and altitude. From these data the Biometric Hash algorithm extracts 131 statistical features. Each statistical feature is individually mapped onto a hash value to create a k-dimensional ($k = 131$) hash vector. For the verification, the reference hash vector is compared to the hash vector, calculated based on the currently presented verification data using a distance function. For the generation of reference and verification vectors, the same Biometric Hash algorithm parameterization is used, which can be estimated for each user individually as well as global based on all registered users or a disjoint set of users.

3.2 Methodology

In order to study also the influence of alternative written content, we use five different so-called semantics. Such semantic can be based on given or individual (secret) information as well as on writer's creativity. While, for the semantic of the *Given PIN* all writers use the same combination of 5 digits (77993), a combination of five individually chosen numbers is used for semantic *Secret PIN*. The third semantic is a *Pseudonym* as a replacement for the signature. Here the donors were asked to train a freely chosen name before starting the data acquisition of the Pseudonym. The semantic *Symbol* holds individual, creative characteristics and consists of secret knowledge based components in terms of the sketched object and the order of single strokes. Finally, the semantic *Place* is the handwritten individual answer to the question 'Where are you from?'.

The evaluation is executed based on the Biometric Hash algorithm, as shortly introduced in section 3.1. In order to determine the verification performance of the aging scenarios, biometric error rates are applied. Since it is not possible to measure these error rates from the system directly, they have to be determined empirically. In order to do so, for each threshold, the numbers of acceptances or rejections for authorized and non-authorized persons are determined experimentally. The *false rejection rate (FRR)* describes the ratio between the number of false rejections of authentic persons and the total number of tests. On the other hand, the *false acceptance rate (FAR)* is the ratio between number of false acceptances of non-authentic persons and the entire number of authentication attempts.

For a comparative analysis of authentication performance, the *equal error rate (EER)* is a common measurement in biometrics. EER denotes the point in error characteristics, where FRR and FAR yield identical value. However, the EER is not to be interpreted as the optimal operating point of a biometric system, it is mainly used as a normalized reference point for comparisons between biometric evaluations, algorithms, etc.

3.3 Test setup

The evaluation is based on data acquired within three sessions, where the semantics Given PIN, Secret PIN, Pseudonym, Symbol and Place are used. The temporal distance between the individual sessions amounts one month each. The acquisition of additional biometric data from the same persons after a longer time turns out as very difficult. This is founded by the fact, that most of the test persons are unavailable after a longer period of time. In general, biometric authentication systems, for example login applications, are used in more or less periodical short intervals. Thus, the time distance of one month suggested in this paper is chosen empirically as initial value. However, based on this evaluation setup tendencies can be shown to motivate further studies.

During the three acquisition sessions, every test subject was asked to donate ten samples for each of the five semantic classes. The database holds data from 53 individuals. In order to generate homogeneous test sets, each data acquisition session was carried out using identical hardware in the same laboratory at similar acquisition time as in the sessions before under guidance of the same supervisor.

To create the reference data R_i for any of the i sessions ($i = 1, 2, 3$) and the five semantics, we take the first five samples to generate the necessary parameters and the corresponding reference BioHash. The remaining five samples are used to determine five hashes for verification attempts. In order to determine the false rejection rate (FRR, see 3.2), the reference data R_i of each person is compared with the verification data V_i of the same person depending on session i and semantic class. The false acceptance rate (FAR, see 3.2) is calculated using the reference data of each person compared to the verification data of each other users. Thus, in this closed environment no influence of external attackers are studied.

Based on this test setup an evaluation is carried out to find possible time dependent influences of biometric handwriting data. The results of the corresponding individual evaluations are presented and discussed in the next subsection.

3.4 Evaluation results

The experimental evaluation is twofold: First we test the reference data R_i with the corresponding verification data V_i of the same session ($i = 1, 2, 3$). In this way we can observe if there are outliers that cause unexpected results already during the age independent verification. In the second step we evaluate the reference data R_1 and R_2 with verification data acquired in following sessions (V_2 and V_3 vs. R_1, V_3 vs. R_2) in order to find dependencies related to template aging. The results of the both evaluation parts are shown and discussed in this subsection.

Aging independent evaluation. The results of the evaluation of reference data (R_i) and verification data (V_i) are shown in table 1 where the rows present the values of the combinations of the reference and verification data acquired in the same session based on the different semantic classes: R_1 *vs.* V_1, R_2 *vs.* V_2 and R_3 *vs.* V_3.

The differences of the verification results for the individual semantics are very small. The highest difference of approximately 0.011 is determined for semantic Symbol between the first two sessions. The best verification performance is calculated also for semantic Symbol in the first session where the EER amounts 0.024.

Table 1. EERs of reference and verification data of the same session in dependency on different semantic classes

	EER		
Semantic	$R_1 vs. V_1$	$R_2 vs. V_2$	$R_3 vs. V_3$
Given PIN	0.063	0.062	0.058
Secret PIN	0.041	0.047	0.051
Pseudonym	0.040	0.037	0.033
Symbol	0.024	0.035	0.038
Place	0.039	0.029	0.033

The best results of the other two sessions are determined for Place ($EER = 0.029$ in session 2, $EER = 0.033$ in session 3). An important observation is the fact that there is no session which contains all best or all worse results for all five semantics. In this way we can exclude in this test setup some kind of habituation of the users after first or second session. The results are also shown in figure 1 to provide a imagination of the relations between them.

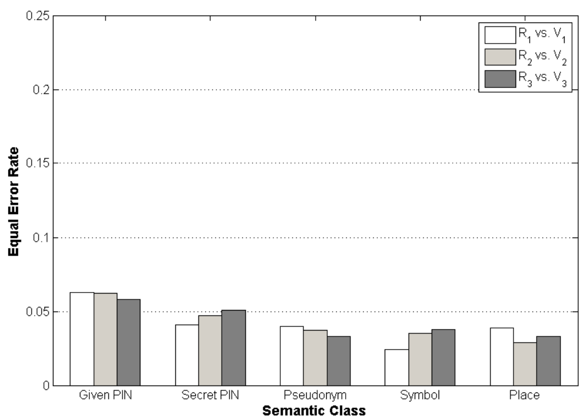

Fig. 1. EERs in dependency on semantic classes without aging between enrollment and verification process.

Age dependent evaluation. While table 2 and figure 2 show the result of the verification based on references of session 1 (R_1) and verification data of all three sessions (V_1, V_2, V_3), table 3 and figure 3 refer to the verification results of second session's reference data (R_2) and verification data of sessions 2 (V_2) and 3 (V_3).

Considering the verifications based on the reference data acquired in the first session (see table 2) the best verification performance is determined for semantic Symbol with an EER of 0.024 using reference and verification data acquired at session one. Also the combination of first session enrollment data and verification data acquired in session two and three, semantic Symbol provides the best verification results holding an EER of 0.069 and 0.071, respectively. The highest degradation between EER determinations of session one and two is shown by semantic class Secret PIN. As shown in the third row of table 2, the EER determined for Secret PIN based on verification data acquired at session two is approximately three times higher than those calculated for verification data acquired one month earlier. Similar degradations can be observed for the other semantics. The smallest worsening can be observed for semantic Place where $EER(R_1\ vs.\ V_2)$ is 2.4 times higher than $EER(R_1\ vs.\ V_1)$.

The EERs determined based on first session's reference data (R_1) and last session's verification data (V_3) show a further degradation for each semantic class. The smallest change is based on semantic Place again. Here the magnification factor amounts approximately 2.6 with an EER of 0.101. The highest change can be observed for semantic Secret PIN where $EER(R_1\ vs.\ V_3) = 0.162$ is four times higher than $EER(R_1\ vs.\ V_1) = 0.041$.

On one side, based on this information, the semantic Place could be an alternative because its stability over the observed time is higher than those of the other semantics. On the other side, the EER of semantic Place is 1.66 times higher than the Symbol's EER (see table 2: $EER(R_1\ vs.\ V_1) = 0.039$ for Place, $EER(R_1\ vs.\ V_1) = 0.024$ for Symbol). This relation between the two semantics is quite similar for $EER(R_1\ vs.\ V_2)$ with 1.35 and $EER(R_1\ vs.\ V_3)$ with 1.43.

Table 2. EERs of reference data from first session and verification data of all three sessions in dependency on different semantic classes

	EER		
Semantic	$R_1 vs. V_1$	$R_1 vs. V_2$	$R_1 vs. V_3$
Given PIN	0.063	0.172	0.224
Secret PIN	0.041	0.124	0.162
Pseudonym	0.040	0.120	0.132
Symbol	0.024	0.069	0.071
Place	0.039	0.093	0.101

Table 3 shows the verification results of reference data acquired in session two (R_2) and verification data of the second (V_2) and third (V_3) session. The results indicate an increase of the EER in all semantics between verification data of session two and verification data of session three. Semantic Place achieves the lowest EER ($EER = 0.029$) within session two ($R_2\ vs.\ V_2$) but also reaches the highest EER increase factor of 3.31. The highest EER was determined for semantic Given PIN in both sessions ($EER = 0.062$ in session two and $EER =$

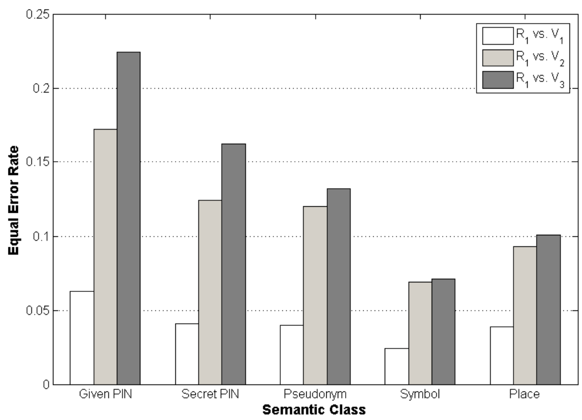

Fig. 2. EERs of reference data from first session and verification data of all three sessions in dependency on different semantic classes

0.152 in session three). The lowest EER increase factor was achieved by the semantic Symbol (2.37). Therefore, the lowest aging affect within session two and three is indicated by semantic Symbol, whereas the most aging impact shows semantic Place. Figure 3 shows the graphical results of the different EER within session two and session three of all semantics.

Table 3. EERs of reference data from second session and verification data of sessions 2 and 3 in dependency on different semantic classes

	EER	
Semantic	$R_2 vs. V_2$	$R_2 vs. V_3$
Given PIN	0.062	0.152
Secret PIN	0.047	0.133
Pseudonym	0.037	0.099
Symbol	0.035	0.083
Place	0.029	0.096

4 Conclusions and future work

In this paper a time dependent experimental evaluation is presented to study the influence of aging on the verification results of a biometric handwriting recognition system. The evaluation is carried out on data, which was acquired in three sessions with one month between each session and under similar conditions. All results show a significant decrease of the verification performance for each of the

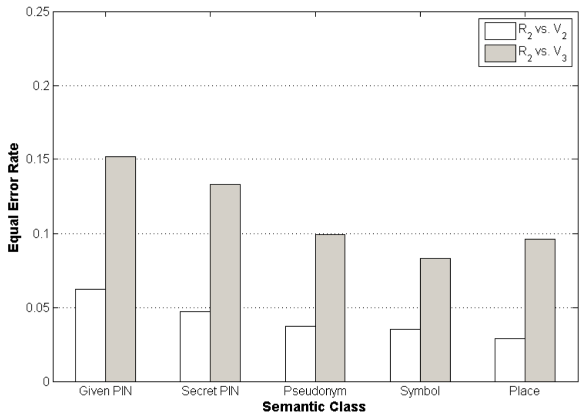

Fig. 3. EERs of reference data from second session and verification data of sessions 2 and 3 in dependency on different semantic classes

semantic used. In the worst case (semantic Secret PIN) the EER of the enrollment data from first session and verification data of third session is four times higher than EER determined based on reference and verification data from the first session. The other evaluation results are similar: Using the enrollment data from first session in combination with verification data from second session the degradation of the EER amounts a magnification from 2.36 up to 3.0 and for last session's verification data from 2.56 up to 3.52, respectively, compared to the verification results of first session's data.

One very important topic of future work is the acquisition of time dependent data from a high number of persons. The aim is the widespread analysis and evaluation of the influence of biological aging processes on authentication performance of biometric systems. For each biometric modality, effectual time distances between individual acquisition sessions have to be found out. The determination and consequently rejection of those statistical features which are more influenced by aging than other could be one possibility to improve the verification performance with respect to a long term usage of the corresponding biometric system. Therefore, adequate feature analysis and/or selection methods have to be carried out. Future research in the area of aging and aging effects in biometrics should be also engaged in multi-biometrics, for example using combination of biometric modalities or algorithms. However, a drawback of multi-biometric systems may lead to a higher complexity of appliance if more than one biometric characteristic has to present. Thus, multi-biometric systems using only one sample of a single modality for biometric fusion (e.g. multi-algorithmic fusion) have to be analyzed from the aging effects point of view on authentication performance. As another possibility to compensate aging effects on biometric data we study methods to update the reference data after each successful verification attempt.

Acknowledgments. This work is supported by the German Federal Ministry of Education and Research (BMBF), project "OptiBioHashEmbedded" under grant number 17N3109. The content of this document is under the sole responsibility of the authors. We also like to thank Prof. Jana Dittmann of the Otto-von-Guericke University Magdeburg and the StepOver GmbH for supporting the project "OptiBioHashEmbedded".

References

1. United Nations, Department of Economic and Social Affairs, Population Devision, Population Estimates and Projections Section: World Population Prospects, the 2010 Revision. Version: 2010. http://esa.un.org/unpd/wpp/index.htm, last checked: 2012-05-02
2. UK Passport Service: Biometrics Enrolment Trial. Version: 2005. http://dematerialisedid.com/PDFs/UKPSBiometrics_Enrolment_Trial_Report.pdf, last checked: 2012-05-11
3. Ministry of the Interior & Kingdom Relations - Netherlands: Evaluation Report Biometrics Trial 2b or not 2b. Version: 2005. http: //dematerialisedid.com/PDFs/88_630_file.pdf, last checked: 2012-05-11
4. Birren, J.E. and Botwinick, J.: The relation of writing speed to age and to the senile psychoses. Journal of Consulting Psychology, 15:243-249 (1951)
5. Rosenblum, S. and Werner, P.: Assessing the handwriting process in healthy elderly persons using a computerized system. Aging Clinical and Experimental Research, 18(5):433–439 (2006)
6. Slavin, M.J., Phillips, J.G. and Bradshaw, J.L.: Visual cues and the handwriting of older adults: a kinematic analysis. Psychology and aging, 11:521–526 (1996)
7. Mergl, R., Tigges, P., Schrter, A., Möller, H.J. and Hegerl, U.: Digitized analysis of handwriting and drawing movements in healthy subjects: methods, results and perspectives. Journal of neuroscience methods, 90:157–169 (1999)
8. Guest, R.: Age dependency in handwritten dynamic signature verification systems. Pattern Recognition Letters, 27:1098–1104 (2006)
9. Michel, L.: Gerichtliche Schriftvergleichung: Eine Einführung in Grundlagen, Methoden und Praxis. Walter de Gruyter, Berlin (1982) (in German)
10. Vielhauer, C.: Biometric User Authentication for IT Security: From Fundamentals to Handwriting (Advances in Information Security). Springer-Verlag New York, Inc., Secaucus, NJ, USA (2006)