The influence of template ageing on a dynamic signature verification system

Joanna Putz-Leszczynska[1]

Abstract. *This paper addresses template ageing in automatic signature verification systems. Handwritten signatures are a behavioral biometric sensitive to the passage of time. The experiments in this paper utilized a database that contains signature realizations captured in three sessions. The last session was captured seven years after the first one. The results presented in this paper show a potential risk of using an automatic handwriting verification system without including template ageing*

Reference: Joanna Putz-Leszczynska, The Influence of Ageing on a Dynamic Verification System. J. Forensic Document Examination, Vol. 24, pp. 47-52.

Keywords: Signature verification, Template ageing, Hidden signature

1. Introduction

Research on automatic signature verification started in 1977 when Herbst and Liu (N.M. Herbst, N., Liu, C., 1977) published the first article in this research area. They applied automatic methods to distinguish a genuine signature from a forgery/simulation. The idea was based on the assumption that a handwriting signature is a result of a signing process that is individual for a person, namely behavioural biometrics. Since then, for over 30 years, hundreds of ideas and methods have been proposed for handwritten signature recognition.

Recently, biometrics have become an integral part of many identity and travel documents (i.e. passports), and because of that *template ageing* (ISO/IEC 19795-1:2006) has become an important issue in biometrics. Each of the aforementioned documents is valid for a number of years, and it means that the biometric template included on a document needs to be valid for the same number of years as well. The question is what if new samples match the template less accurately as time passes? This question opens a new research area named *template ageing*. The two main questions are: what is the effect of template ageing on verification errors, and how is the scientist going to address this effect?

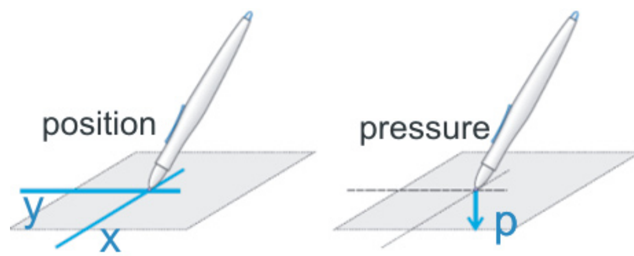

Figure 1: Handwritten signature captured on-line

Figure 2: Digitizing tablet

In the biometric literature, warnings have already been published for physiological biometrics for iris identification (Fenker S., Bowyer K., 2012), for face identification (Uludag U., Ross A., A. Jain A., 2004, and for fingerprints (Ryu J. et al., 2007). If this phenomenon is present for physiological biometrics, it will affect behavioral biometrics. Physiological

1. Warsaw University of Technology
Nowowiejska 15/19, 00663 Warsaw, POLAND
Email: jputz@elka.pw.edu.pl, http://zbum.ia.pw.edu.pl/

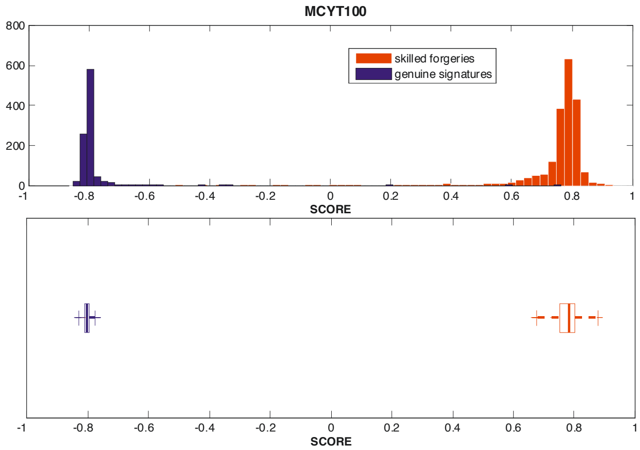

Figure 3: Scores histograms (top) and distribution box plots (bottom) for MCYT database.

group	EER[%]		
	Session 1	Session 2	Session 3
time	0 day	after 1 month	after 7 years

Table 1: The EER calculated independently for each session.

biometric sample intrasession variability is also higher than for the physiological one.

One of the most popular behavioral biometrics is the handwritten signature captured on-line, often called the dynamic signature. This signature is something more than a 2D image. It is a multi-dimensional sequence in time that stores information about the pen tip position, and pen tip pressure for each moment of the signing process. (Figure 1). These signatures are captured with the use of digitizing tablets (Figure 2).

A signature represented as a sequence of vectors opens up a great number of opportunities to draw from the signal processing methods, what was already presented in several surveys (Impedovo D., Pirlo G., 2008)

During the last ten years, a number of signature verification competitions have been organized (SVC 2004 (Dit-Yan Yeung, 2003), ESRA 2011 (N. Houmani et al., 2011)). The verification error rates of the best participating algorithms were in the 2-4 % range. Papers (Guest R., 2007; M. Erbilek M., Fairhurst M., 2012) show a strong age dependency in handwritten dynamic signature verification systems. For example, the errors in a group of young people is higher than in a group of older ones, whose signing process is more automatic. The differences in error rates were also noticed between the genders.

However, there is a lack of publications that show the differences between the intrasession and intersession signatures variability, and the effect of passing time – how the variability changes after one month, two months, or a year. A question arises whether we can tell that an observed increasing variability is a session effect or an aging effect. Signature ageing in time is an important aspect of the presented tests, since it is known that a genuine realization of a particular signature is more similar when written at the same session than when done at different sessions, separated in time. Here, the resent results were obtained for a database containing signature realizations that were captured in three sessions. The second session took place after one month, and the third one after seven years.

2. Database, algorithm

The verification scores were evaluated in estimation conditions. The results presented in this paper are calculated with the use of the signature verification system at the NASK/PW Biometric Laboratory the Warsaw University of Technology (Putz-Leszczynska J., Pacut A. 2012, Putz-Leszczynska J., Kudelski M. 2010). The algorithm used in the experiments placed in the top of the ESRA 2011 competition. This automatic method uses signatures captured on-line to create a person's signature template – an artificial signature, called the hidden signature. The hidden signature is estimated with the use of a number of signatures captured during the registration process. Then, during the verification, the system makes a decision (acceptance or rejection) based on a verification score - the result of comparison between person's template (hidden signature) and verified signature. Two signatures are aligned with the use of dynamic time warping (DTW) in position coordinates space; however, the pressure value and dynamics also impact the score rate.

Additionally, the tests presented in this article were conducted with the use of a database that contains data belonging to 85 signature users (office workers and students), which is a property of the NASK/PW Biometric Laboratory and not publicly available because of personal data protection regulations. The signatures were collected under the guidance of a supervisor, and the collection took place in a controlled standard office environment. The samples were captured using a graphical tablet with the same controlled environment and technical equipment. It is important to notice that there were two groups. The first group contained 50 signature classes. For each signature class, 20 genuine signatures and 20 skilled forgeries/simulations were collected in the first session, and the same procedure was repeated after a month; also 20 genuine signatures and 20 skilled forgeries/simulations were collected for each class. The second group contained 35 signature classes – only six genuine signatures and six skilled forgeries/simulations were collected for each class. The collection procedure for the same people was repeated seven years after the first one. Again, six genuine signatures and six skilled forgeries/simulations were collected. After combining both groups the resulting database can be described by:

Session 1: 85 signature classes
Session 2: 50 signature classes / after 1 month
Session 3: 35 signature classes /after 7 years

For each signature class five genuine realizations were used for template creation; the remaining genuine signatures

group	EER[%]		
	Session 1	Session 2	Session 3
time	0 day	after 1 month	after 7 years
intrasession	2.8	3.2	2.9
intersession	2.8	10.2	27.94

Table 2: The EER calculated independently for each session.

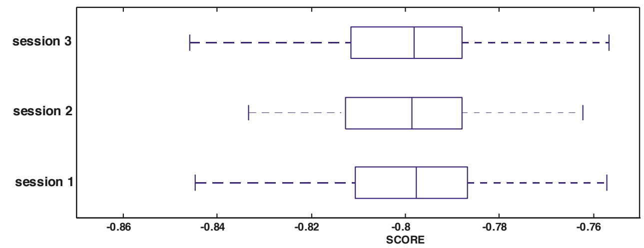

Figure 4: Genuine signature scores distribution in each session.

and skilled forgeries were used for verification. It gave the one genuine score values and six skilled forgeries/simulations score value for each of 35 classes in Sessions 1 and 3 and 15 genuine score values and 20 skilled forgeries/simulations score value for each of 35 classes in Sessions 1 and 2. Obtained scores were used for the estimation of EER (Equal Error Rate), the value that equalizes the False Acceptance Rate (FAR) and False Rejection Rate (FRR) of signatures.

3. Intersession vs. ageing

First, it was important to demonstrate the algorithm's potential. The EER estimated on the MCYT database (Ortega-Garcia J., Fierrez-Aguilar J. et al, 2003) (five signatures for template) gives a result of EER=2.4%. The genuine signatures and skilled forgeries histogram of the scores (Figure 3) can be represented as well in the form of box plot that displays differences between them without making any assumptions of the underlying statistical distribution. The one presented in this article contains also outliers.

1.1. Intrasession results

Independently, for each session in the database presented in section 2, the same experiment which was carried out for the MCYT100 signature database. It means that the template was enrolled and compared with signatures from the same session, independently for each session. The obtained error rates (Table 1) were comparable and similar to the one calculated for a MCYT database.

Simultaneously, the genuine signature scores distribution (Figure 4) looks similar, and zero hypothesis of equality of mean value between the sessions cannot be statistically rejected. This result was

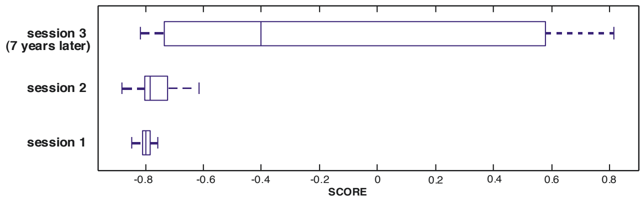

Figure 5: Genuine signature scores distribution in each session for one common template created in session 1.

obtained with the use of one-way analysis of variance (ANOVA1, α=5%). In short, I can say that samples/ signatures in these three sessions were compared with the templates with the same mean values.

1.2. Intersession results

To obtain the intersession results, the procedure of template creation was changed. In the intersession experiments, a template was enrolled once, with the use of five signatures from the first session. Then the template was compared with signatures from the first, second and third session.

The obtained error rates differ from the ones from the intrasession experiment (Table 2). The first difference that is visible for the Session 2 shows the well-known *intersession effect*. It is more visible for the behavioral biometrics. For this case also the young age (20 to 25) of the population has an impact on the obtained results (more in Section 3.3). The more interesting results are the ones obtained after seven years that involve the expected significant increase of error.

Additionally, when we observe the genuine scores (Figure 5) distributions, we notice that they are visually different, especially between Session 3 and the other sessions. This visual observation was confirmed by statistically rejecting a zero hypothesis of equality of mean value between the sessions. This result was obtained with a use of one-way analysis of variance (α=5%).

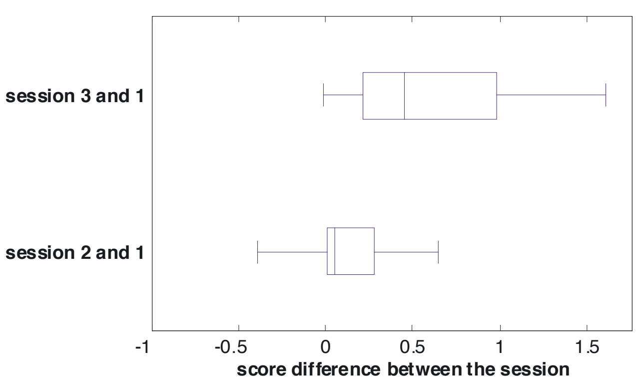

Figure 6: Differences between the scores of the same persons in between the sessions.

However, the database has a certain weakness in its construction, namely the intersection class population of Session 1 and Session 2 is Session 2, the intersection class population of Session 1 and Session 3 is Session 3, and the intersection of Session 2 and Session 3 is empty. As a result, we cannot easily conclude if the observed difference between Session 3 and other sessions is the result of template ageing. The differences observed in Figure 5 may be only the intersession effect, which is different for the population in Sessions 2 and 3. In order to check it, one more experiment was performed.

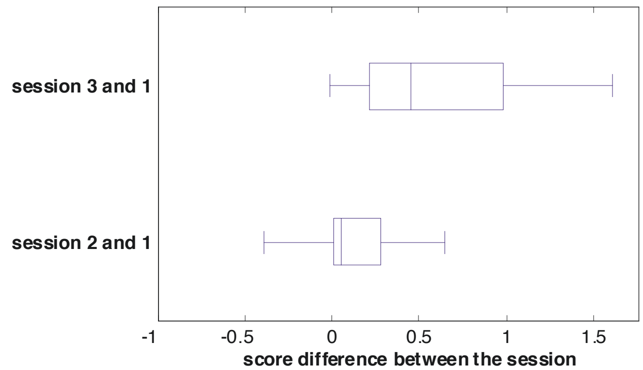

Figure 6: Differences between the scores of the same persons in between the sessions.

1.3. Template ageing

To get the answer and omit the problem of two subsets in our database, the scores between the sessions were not analyzed, but the differences between the scores of the same person in Session 1 and Session 2 independently from the differences between the scores of the same persons in Session 1 and Session 3 (see Figure 6). The hypothesis backing this step was that, if in both Sessions 2 and 3 the observed increasing of the EER was only the intersession effect, the corresponding distributions of the differences (Figure 6) would be similar. Again, this hypothesis was tested using the one-way analysis of variance (α=5%).

The zero hypothesis of equality of means was strongly rejected (p= 6e-007), thereby confirming the visual differences between the distributions in Fig.6. This fact suggests that the greater dispersion observed in session 3 is not only a result of the intersession effect but mainly a result of template ageing.

If we look closer at the distributions of differences in Figure 6, we can notice that some of the values are negative. Negative values indicate that EER has not changed for all signature classes. Some signature realizations fit better to the template after a month (they are from session 2) than after a minute (session 1).

For better presentation of this issue, the histograms of individual (calculated independently for each signature class, individual threshold for each class) EER in all sessions are presented (Figure 7).

We can observe that in Session 1 almost all individual EER are equal to 0, while for Session 2, 60% of classes and for Session 3 only 20% of classes are equal to zero. The most disturbing thing is that for almost 30% of the classes in Session 3, it is impossible to distinguish between the genuine signatures and skilled forgeries/simulations using a template created in Session 1 – the person's signatures have changed dramatically, and it is impossible to recognize these signatures with the use of the old (ageing) template. On the other hand, these signatures (from Session 3) were collected after seven years; so one can say that the EER is only 30%.

4. Conclusions

Template ageing is now one of the most important issues in biometrics. Because a handwritten signature is a behavioral biometrics, it is sensitive to the passage of time. In this paper, the ageing of a signature template exists has a strong impact on the increasing of the error rates. It is an extension of the conclusions presented in M. Fairhurst publication from 2012 (M. Erbilek, M. Fairhurst, 2012). Their experimental results show that it is very hard to categorically identify clear trends when considering the relationship between error rate performances in signature biometrics as a function of age groupings with a user population. The authors extrapolate their findings to conclude that the need for frequent template updating with physical ageing might be a less sensitive issue than might have been imagined. This current research expands these conclusions.

Forensic handwriting experts also notice the ageing problem. That is why examining signatures in wills are sometimes very difficult where there are few contemporary samples of the decedent's writing or samples were written many years before the will was dated. However, human experts cope easier with this problem. For automatic verification/identification there is a need for mechanisms that renew the template to reduce the error rate resulting from aging.

References

Albert A.M., Ricanek K., (2010), Implications of Adult Facial Aging Factors on Biometrics, In Biometrics, Vienna, Austria, InTech Publishing, 2010.

Erbilek M., Fairhurst M., (2012), Framework for managing ageing effects in signature biometrics, IET Biometrics, Volume 1, (2), 136–147 2012.

Fenker S.P., Bowyer K.W., (2012), Analysis of Template Aging in Iris Biometrics, Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, Rhode Island: 45–51, 2012.

Guest R., (2006) Age dependency in handwritten dynamic signature verification systems. Pattern Recognition Letters, 27(10):1098–1104, 2006

Herbst N.M., Liu . C.N. (1977) Automatic signature verification based on accelerometry. In IBM J.Res. Develop: 245-253.

Impedovo D., Pirlo G., (2008) Automatic Signature Verification: The State of the Art, IEEE Transactions On Systems, Male, And Cybernetics. Part C, Applications And Reviews, 38 (5): 609-635.

Houmani N. Garcia-Salicetti, S. ; Dorizzi, B. ; Montalvao, J. ; Canuto, J.C. ; Andrade, M.V. ; Qiao, Y. ; Wang, X. ;Scheidat, T. ; Makrushin, A. ; Muramatsu, D. ; Putz-Leszczynska, J. ; Kudelski, M. ; Faundez-Zanuy, M.; Pascual-Gaspar, J.M. ; Cardenoso-Payo, V. ; Vivaracho-Pascual, C. ; Argones Riia, E. ; Alba-Castro, J.L. ; Kholmatov, A. ; Yanikoglu, B.., (2011), BioSecure Signature Evaluation Campaign (ESRA'2011): Evaluating Systems on Quality-based categories of Skilled Forgeries, Proc. IEEE IJCB 2011, Washington, D.C., USA, 2011ISO/IEC 19795-1:2006, (2006), "Information technology - Biometric performance testing and reporting…", Section 6.4.6.

Ortega-Garcia J., Fierrez-Aguilar J. et al., (2003) MCYT baseline corpus: a bimodal biometric database. Vision, Image and Signal Processing, IEE Proceedings, 150, 395-401, 2003.

Putz-Leszczynska, J., Pacut, A. (2012) Model approach to DTW signature verification using error signals", Journal of Forensic Document Examination (JFDE) 22: 41-54

Ryu J., Jang J., Kim H., 2007 Analysis of Effect of Fingerprint Sample Quality in Template Ageing, NIST Biometric Quality Workshop , 2007

Uludag, U., Ross A., Jain A., (2004) Biometric template selection and update: a case study in fingerprints, Pattern Recognition 37, 1533–1542, 2004.

Dit-Yan Yeung, Hong Chang, Yimin Xiong, George S., Kashi R., Matsumoto T., and Rigoll G.. (2004) SVC2004: First International Signature Verification Competition. Proceedings of the International Conference on Biometric Authentication (ICBA), Hong Kong, 15-17 July 2004.