Study of long-term quality of online signature verification systems

Conference Paper · November 2016

DOI: 10.1109/CCIntelS.2016.7878206

Tobias Kutzner
Brandenburg University of Technology Cottbus - Senftenberg
11 PUBLICATIONS **30** CITATIONS
SEE PROFILE

Ingrid Bönninger
Brandenburg University of Technology Cottbus - Senftenberg
13 PUBLICATIONS **38** CITATIONS
SEE PROFILE

Carlos M. Travieso
Universidad de Las Palmas de Gran Canaria
374 PUBLICATIONS **2,788** CITATIONS
SEE PROFILE

Anushikha Singh
Amity University
54 PUBLICATIONS **336** CITATIONS
SEE PROFILE

Study of long-term quality of online signature verification systems

Tobias Kutzner[1], Ingrid Bönninger[1]
[1]Institute of Medical Technology
Brandenburg University of Technology Cottbus – Senftenberg, GERMANY

Carlos Travieso[2], Malay K. Dutta[3]
[2]Departamento de Señales y Comunicaciones, IDeTIC, Universidad de las Palmas de Gran Canaria, SPAIN.
[3]Amity School of Engineering & Technology, Amity University, Noida, INDIA.

Abstract— **Real handwriting authentication systems need a robust writer identification over a long time period.**

In this paper we work with signature sessions of the ATV-Signature Long Term Database (ATV- SLT DB).

The database contains 6 sessions generated by 27 users over 15 month. We examine the quality change of the verification results over a period of 15 month. We extract 64 static and dynamic biometric features from the ATV-SLT DB sessions and use 3 different classifiers.

For the impostor test we add a 7th session, the impostor session, with 6 signatures for each user.

The best result of 99.17% success rate for a correct classification, we reached with the k-Nearest Neighbor classifier. The best result of 2.47% false accepted rate is reached with Naïve Bayes classifier. (***Abstract***)

Keywords—online handwrting; signatur; writer verification; sessions; data mining; mobile devices; artificial intelligence (key words)

I. Introduction

With increasing number of mobile devices the need of security systems today is growing. Online signature verification is one of the biometric methods to achieve more safety for transactions on mobile devices. Most prior research has concentrated on feature selection and classifier evaluation. The notation that real authentication systems need a high writer identification quality over a long period of time appears unattended.

In [1][2][3] we have concentrated our research to select features and algorithms to identify writers by their handwritten passwords.

In this paper we test our features and algorithm (k-Nearest Neighbor, Naïve Bayes, and Bayes-Nets) with the public database ATV-SLT. The DB in [4] with the handwriting session data collected over long time is the basis for our experiments. For the experiments we enlarge the dataset by one more session, the impostor session written on a mobile device. We have to transform the handwriting data files to our format to make it compatible to our system. An important fact is that a standard mobile device has no touch pressure sensor for the display. Therefore we want to work without the dynamic pressure values. We only use x, y coordinates and time for the feature generation. This makes it more complicated, but it is closer to the reality for security systems development for mobile devices without display pressure sensors.

II. Related Work

In previous handwriting analysis little attention has been paid to aging of handwriting.

The effect of the user age are analyzed in [5][6]. Methods for classifying of three age ranges are recommended in [5]. Handwriting of 405 Persons is analyzed. Three age ranges are classified with a precision about 70%.

An age prediction model has been developed in [6] with 30 signatures of each person (210 persons, age range of 1-73) in 2 sessions. Three age ranges are classified with 75% by using only feature of handwriting. In combination with iris feature the prediction systems achieves up to 90% accuracy.

A database of handwritings of 400 users (from 16 to 90 old) collected in 4 sessions in a time span of 4 month is used in [7]. The experiments show a loss of writing speed in later life (60 and above). The probability of someone being incorrectly verified as someone else (FAR) depends on age. Handwriting of older writer is easier to falsify than younger ones. But the intra-personal variability does not seem to be significantly dependent on age. All age ranges appear to have equal FFRs.

That means handwriting changes on age.

The question is now, how reliable are real authentication systems that have to work with aging signatures.

A multi-session database (180 users, 6 sessions, and 12 - 96 hours) is used in [8] to evaluate an effective histogram based feature extraction algorithm. Although the sessions are distributed to 12-96 hours, the results of inner-sessions verification are better than inter-session verification and inter-session results are better when the classifier is trained with the preceding session.

In [4] the aging problem is analyzed by Dynamic Time Warping (DTW), and Hidden Markov Model (HMM). They used the ATV-SLT Database with six handwriting sessions over a period of 15 month. Their conclusions are: Aging seems not depend on the type of signature but on the signer, dynamic features are less stable than static features, DTW as a classifier is more robust than HMM for long-term writer identification.

We use the same ATV-SLT Database like [4]. Our objectives are to test:

- whether the forgery resistance is changing over a long time period.
- whether we can confirm the results of [4] with our features and algorithms
- whether we can recommend the use of displays without pressure sensors in authentication systems
- whether the authentication system is forgery resistant, if skilled the impostor sees all original signatures during the forgery (see fig. 3).

The basis for our experiments is the Signature Long Term DB [4], we enlarge the DB by one session, the Impostor Session (get fig. 1).

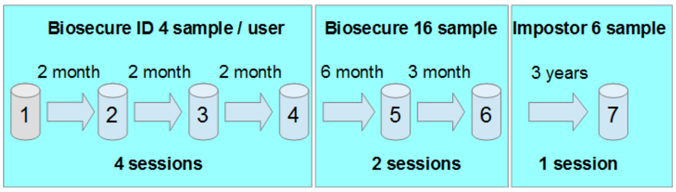
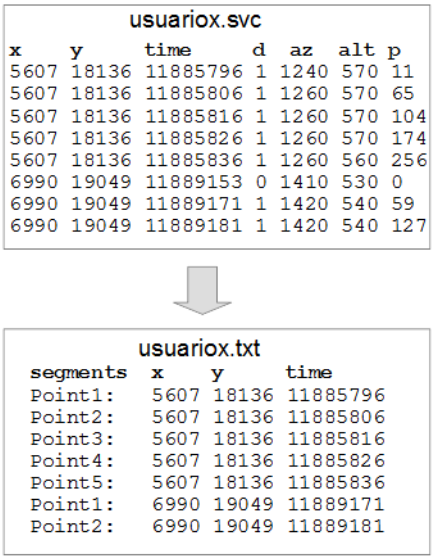


Fig. 1 Description Signature Long Term DB and Impostor Session (detailed description session 1 - 6 see [4]).

III. Feature Datafiles Generation

A. Transform ATV-SLT Datafiles

To work with the data from ATV-SLT DB in our system we have to transform the given SVC Files to our data format. With a Java program we do the following transformation (see fig. 2).



usuariox.svc

```
x    y     time     d  az   alt  p
5607 18136 11885796 1 1240 570 11
5607 18136 11885806 1 1260 570 65
5607 18136 11885816 1 1260 570 104
5607 18136 11885826 1 1260 570 174
5607 18136 11885836 1 1260 560 256
6990 19049 11889153 0 1410 530 0
6990 19049 11889171 1 1420 540 59
6990 19049 11889181 1 1420 540 127
```

usuariox.txt

```
segments  x    y     time
Point1:   5607 18136 11885796
Point2:   5607 18136 11885806
Point3:   5607 18136 11885816
Point4:   5607 18136 11885826
Point5:   5607 18136 11885836
Point1:   6990 19049 11889171
Point2:   6990 19049 11889181
```

Fig. 2 SVC to TXT transformation

For the impostor test we need the images of the signatures. We generated these images as PNG graphic from the SVC files with a Java program too (see fig. 3.

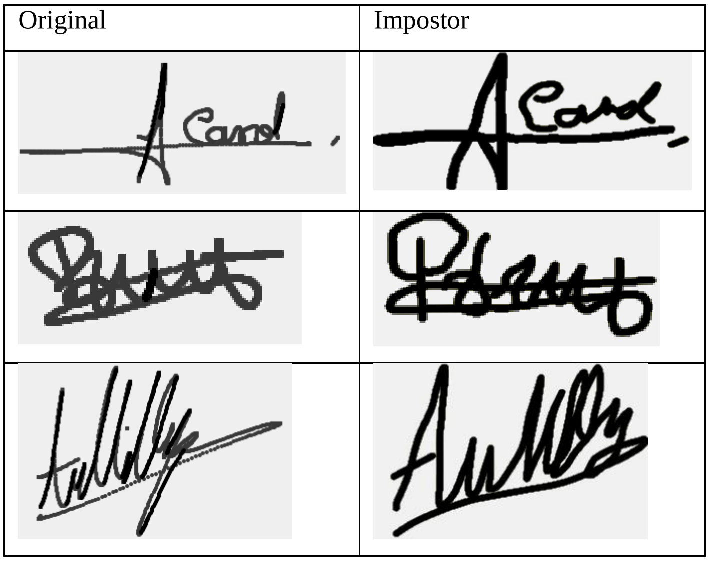


Fig. 3 Samples: PNG from SVC and Impostor PNG

B. Generate Feature Datafiles

After transformation we generate the feature data files with the segmentation, x, y coordinates and time. From this data files we generate the parameters. Beside some statistical parameters we use primarily time, speed and relation parameter. In [1][2][3] you find the detailed description of the parameter we extract from the online signatures.

IV. Experiments

We run four experiments: first we use the first four sessions separately and use one session for training and one for testing. Each session contains four samples. In the second experiment we combine two sessions' pairs for training and one for testing. In the third experiment we use all 6 sessions and split percentage into train and test data for the classification. In the final experiment we run the impostor test with all six sessions for training and the seventh session for testing. The following part figure out the results of the experiments:

First experiment

The results of the first experiment training and testing all first four sessions separately and classify with three classifiers Naïve Bayes, Bayes Net, and KNN are shown in table I, II and III:

TABLE I. Test Naïve Bayes Classifier

test session train	**first**	**second**	**third**	**fourth**
first	100%	83,33%	63,89%	40,74%
second	77,78%	100%	66,67	37,04%
third	51,85%	53,70%	100%	75,00%
fourth	47,22%	45,37%	86,81%	100%

TABLE II. TEST BAYES NET CLASSIFIER

test session train	first	second	third	fourth
first	100%	96,30%	89,81%	78,70%
second	89,15%	100%	89,81%	77,78%
third	94,44%	87,04%	100%	83,33%
fourth	82,41%	74,07%	89,81%	100%

TABLE III. TEST KNN CLASSIFIER

test session train	first	second	third	fourth
first	100%	94,44%	67,50%	56,48%
second	92,59%	100%	68,52%	51,85%
third	66,67%	66,67%	100%	82,41%
fourth	58,33%	53,70%	87,04%	100%

We see the first and second session are relatively similar and achieve good results, the third and fourth session achieve worse results compared with the first and second session. The best result of 96.30% correctly classified delivers the Bayes Net classifier with the first session for training and the second session for testing.

Second experiment

The results of the second experiment that combines session pairs as training sets and use one session as test set are shown in table IV, V and VII:

TABLE IV. TEST NAÏVE BAYES CLASSIFIER

test session train	first	second	third	fourth
first_second	100%	100%	53,70%	35,19%
first_third	100%	99,07%	100%	78,70%
first_fourth	100%	94,44%	88,89%	100%
second_third	92,60%	100%	100%	75,00%
second_fourth	96,30%	100%	93,52%	100%
third_fourth	67,69%	67,59%	100%	100%

TABLE V. TEST BAYES NET CLASSIFIER

test session train	first	second	third	fourth
first_second	100%	100%	88,89%	78,70%
first_third	100%	97,22%	100%	89,81%
first_fourth	100%	95,37%	96,30%	100%
second_third	95,37%	100%	100%	92,60%
second_fourth	97,22%	100%	98,15	100%
third_fourth	89,81%	89,81%	100%	100%

TABLE VI. TEST KNN CLASSIFIER

test session train	first	second	third	fourth
first_second	100%	100%	65,74%	60,19%
first_third	100%	95,37%	100%	83,33%
first_fourth	100%	93,52%	97,22%	100%
second_third	97,22%	100%	100%	83,33%
second_fourth	98,15%	100%	96,30%	100%
third_fourth	69,44%	66,67%	100%	100%

In this experiment, two sessions with eight training samples and four test samples per user where summarized. The best result of 98.15% correctly classified deliver the Bayes Net and the k-Nearest Neighbor classifier with the second_fourth session for training, Bayes Net with the third session for testing and k-Nearest Neighbor with the first session for testing.

Third experiment

For the third experiment we split all samples beginning from first up to sixth session randomly into training- and test pairs. After that, we had altogether 1296 sets ((4 samples x 4 sessions + 16 samples x 2 sessions) x 27 writer) to run the third experiment. At last, Bayes-Nets, Naïve Bayes and k-Nearest Neighbor classifiers were used to classify each pair simultaneously. The results of the third experiment are shown in table VII and fig. 4:

TABLE VII. TEST SPLIT ALL CLASSIFIER

classifiers N%split	Bayes-Nets	Naive Bayes	KNN
50%	98,02%	97,36%	96,54%
60%	98,35%	97,33%	97,53%
70%	98,46%	96,98%	97,80%
80%	98,35%	97,12%	97,12%
90%	99,17%	97,52%	98,35%
Cross F 20	98,85%	97,86%	97,86%
AVG Time in s	0.35	0.06	0.01

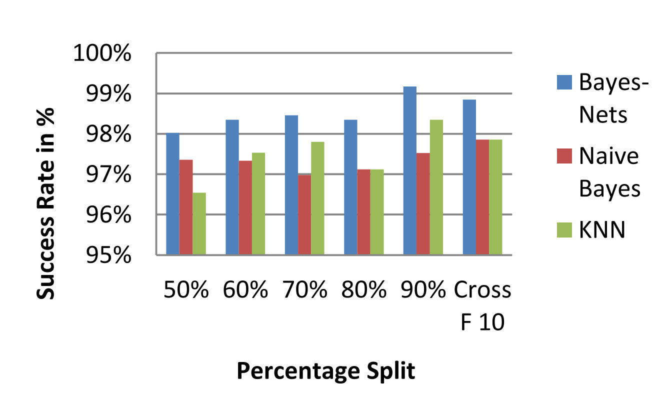


Fig. 4 Bar chart of the third experiment

Fourth experiment

For the fourth experiment we used all six sessions for training and the seventh session with six skilled forgeries per user for testing. We had all together 162 forgeries for the experiment. First we trained all sessions separately and secondly all sessions together. Finally the average time for the classification process was calculated. The results of the fourth experiment are shown in table VIII and fig. 5:

TABLE VIII. IMPOSTOR TEST

classifiers session	Bayes-Nets	Naive Bayes	KNN
first	10,49%	7,41%	10,49%
second	11,11%	8,02%	11,73%
third	8,64%	4,32%	4,93%
fourth	8,02%	7,41%	4,32%
fifth	9,64%	9,88%	9,88%
sixth	7,41%	10,49%	10,49%
all sessions	11,73%	2,47%	11,11%
AVG Time in s	0.39	0.02	0.05

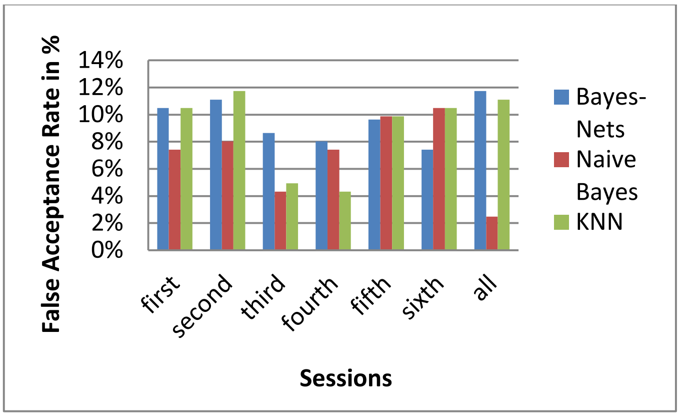


Fig. 5 Bar chart of the fourth experiment

V. CONCLUSION

In conclusion, it can be noted that we can recommend the use of displays without pressure sensors, because the writer verification results with and without pressure seem not be different. We can confirm the results of [4] using our 64 features, the classification algorithms Naïve Bayes, Bayes Nets and KNN reach better results than HMM- and Global feature-based systems [4]. But using DTW [4] reaches the best FAR and FRR results.

Aging of handwriting over a time period of 15 month has a negative influence to the correct writer verification, if we only use one session as training set. We can improve the results by using more sessions of signatures. We recommend the use of many signatures over a long time period.

Our recommended system is forgery resistant for skilled forger who can see the original signature with the FAR of 2.47%, if we use Naïve Bayes for the classification and all 48 signatures per writer of the time period of 15 month.

With better classifiers and more parameters, the system can be further improved in the future.

REFERENCES

[1] Kutzner T., Travieso C.M., Bönninger I., Alonso J.B., Vasquez J.L.: Writer identification on mobile device based on handwritten, Security Technology (ICCST), 2013 47th International Carnahan Conference, pp. 1-5, 2013.

[2] Tobias Kutzner, Fanyu Ye, Ingrid Bönninger, Carlos Travieso, Malay Kishore Dutta, Anushikha Singh, User Verification using Safe Handwritten Passwords on Smartphones, Eight International Conference on Contemporary Computing (IC3), 2015.

[3] Tobias Kutzner, Mario Dietze, Ingrid Bönninger, Carlos Travieso, Malay K. Dutta, Online Handwriting Verification with Safe Password and increasing number of features, 3rd International Conference on Signal Processing and Integrated Networks, SPIN-2016, 2016.

[4] Aging in Biometrics: An Experimental Analysis on On-Line Signature Javier Galbally*, Marcos Martinez-Diaz, Julian Fierrez Biometric Recognition Group-ATVS, Universidad Autonoma de Madrid, Madrid, Spain 2013.

[5] N. Bouadjenek, h. Nemmour, Y. Chibani, Age, Gender and Handedness Prediction from Handwriting using Gratient Features, 13th International Conference on Docment Analysis and Recognition (ICDAR), pp.116-1120, 2105.

[6] M. Erbilek, M. Fairhurst, M. Da Costa-Abreu, Improved age prediction from biometric data using multimodal configurations, Intern. Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1-7, 2014.

[7] M. Faundez-Zanuy, E. Sesa-Nogueras, J. Roure-Alcobé, On the relevance of age in handwritten biometric recognition, IEEE International Carnahan Conference on Security Technology (ICCST), pp. 105-109, 2012.

[8] N. Sae-Bae, N. Memon, Online Signature Verification on Mobile Devices, *Fellow, IEEE 2014, pp. 933-974.,2014.*

[9] J. Fierrez, J. Ortega-Garcia, D. Ramos, and J. Gonzalez-Rodriguez, “HMM-based on-line signature verification: Feature extraction and signature modeling,” *Pattern Recognit. Lett.*, vol. 28, pp. 2325–2334, Dec. 2007.

[10] H. Feng and C. C. Wah, “Online signature verification using a new extreme points warping technique,” Pattern Recognit. Lett., vol. 24, no. 16, pp. 2943–2951, 2003.

.