AN APPROACH TO DATA COLLECTION IN AN ONLINE SIGNATURE VERIFICATION SYSTEM

Andreea Salinca, Sorin Mircea Rusu, Ştefan Diaconescu
Research & Development, SOFTWIN, Bucharest, Romania
asalinca@softwin.ro, srusu@softwin.ro, sdiaconescu@softwin.ro

Keywords: Data Collection, Database, Biometry, Dynamic Signature, Online Signature, Authentication System

Abstract: Online handwritten signature verification is one of the branches of behavioral biometry that is gaining popularity in protecting sensitive information. Our paper addresses a key issue in evaluating performances of online signature authentication systems: data collection. Acquiring a real dataset with handwritten signatures is a major step in the system verification. We will present our collecting techniques in the process of acquisition of dynamic handwritten signatures (more than 5000 genuine signatures and more than 2000 skilled forgeries have been collected from a total of 113 people) useful in the improvement of evaluation results for the authentication system.

1 INTRODUCTION

Biometric authentication has been widely used as a trusted security solution in protecting sensitive asset. These systems are not based on what a person possesses (password, PIN, token), but on the basis of what the person "is". Unlike physiologic biometry, the authentication with holographic signature is non intrusive. For hundreds of years, the signature has been approved by all cultures and civilizations with a major social implication, and it has been accepted as legal evidence.

This paper presents several signatures databases acquisitions (PHILIPS, SVC'2004 Development Set, MCYST Signature Subcorpus, BIOMET Signature Subcorpus, BioSecure Signature Subcorpus DS2 and DS3) and includes a survey on acquisition devices, procedures of acquiring genuine signatures and several types of forgeries, and the main results obtained in their evaluation and in international evaluations like SVC'2004 - First International Signature Verification Competition (Bernadette, Chollet, and Petrovska-Delacrétaz, 2009).

The signatures database collected for SVC'2004 contains samples from only 60 people, and for privacy reasons, the signatures are not "real". There were two sets of signatures: one containing only coordinate information and the other containing also pen pressure and orientation. The team from Sabanci University of Turkey obtained the best result in the evaluation of a Dynamic Time Warping based system for both sets (the first set had an ERR of 2.84% and the second set had an ERR of 2.89%) (Yeung et al., 2004). A recent article presents evaluation results of online signature obtained in BioSecures Signature Evaluation Campaign (BSEC'2009), depending on the signatures quality captured on fixed (DS2) or mobile platforms (DS3) from a total of 382 people. The main task was to evaluate the algorithms' results in different acquisition conditions of signatures as well as the complexity of information contained in signatures. The results revealed that system robustness depends on the quality of signatures (on DS2 an ERR of 2.2% for skilled forgeries and an ERR of 0.51% for random forgeries, and also, on DS3 an ERR of 4.97% for skilled forgeries and an ERR of 0.55% for random forgeries) (Houmani et. al, 2011).

Comparing to previous works in collecting signatures databases which have focused on improving the evaluation methods, we will focus on post-processing the data already collected in order to improve the evaluation performance of a system.

The first section briefly describes our data acquisition system and the acquisition application. Further on, we present the procedures we used to accomplish the acquisition process and our strategies for post-processing the dataset. Finally, we present experimental results proving the impact of data collecting strategies over the system performances on evaluating the authentication system.

2 ONLINE SIGNATURE ACQUISITION SYSTEM

The online acquisition system contains three main sub-systems: the acquisition device, the acquisition application and the signatures database (Figure 1).

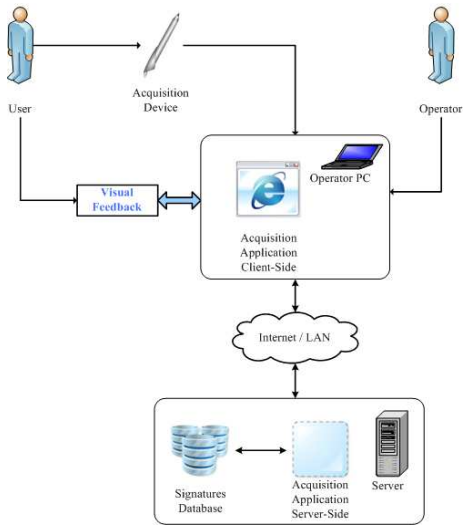

Figure 1: Acquisition System Architecture.

The acquisition device is an electronic pen which captures the bio-kinetic information of a signature. It writes on a pattern paper and it does not need a special tablet (Rusu, Dinescu, Diaconescu, 2011). When a person signs, the acquisition device (Figure 2) captures hand movements using the embedded MEMS (micro-electro-mechanical systems) accelerations sensors into a set of 4 time-based series. Moreover, it captures the graphic form of the signature through an optical navigation sensor (ONS) located inside the biometric pen.

Figure 2: Dynamic Signature Acquisition Device.

2.1 Acquisition application

The actors of our acquisition system were subjects, operators, acquisition device, acquisition application and acquisition database. In the process of data collecting the subjects were assisted by operators.

The acquisition application is designed as a web-based application. It offers a real-time visual feedback of the signature. Besides the biometric signature acquisition and storage in the system's relational database, another functionality provided by the application was of managing a set of genuine signatures and a set of forgeries for every user. The application manages additional features for operators and users using the electronic pen for signing.

3 ACQUISITION PROCESS

The creation of a dynamic handwritten signatures database must correspond to a real environment use of the authentication system. The data collection has to be large enough to cover particular cases (e.g. signatures which are very hard to forge). The subjects were asked to use their "daily life" signatures and to try to sign in the most natural way. Our signatures database was built from 113 persons of different ages and genders. The number of collected signatures was over 7000. The acquisition process of the online handwritten signatures database contains three phases:

- Phase 1 – This is the preparation phase before the proper acquisition phase. The operator enters the names of the signers in the acquisition application and prints the signing forms. We use two types of signing forms: one for genuine signatures and one for forgeries.
- Phase 2 – Each of the subjects has to give a set of 50 samples of genuine signatures and 20 skilled forgeries for another user. The signatures are collected in 6 different sessions which are presented below. The subjects providing signatures are assisted by multiple operators. Every operator has to assist an equal number of subjects.
- Phase 3 – Post-processing of the database occurs in this phase. Several procedures are applied to the dataset in order to eliminate some signatures that could be affected by different types of errors. We will prove the significant role of this phase in the evaluation of the authentication system.

3.1 Acquisition sessions

The acquisition sessions from the second phase of the acquisition process represent the acquisition itself of the set of dynamic signatures. In 6 different sessions, each signer supplied a total of 50 samples of genuine signatures and 20 skilled forgeries. There

was a few days distance in between the sessions, to make the process more real. In this purpose, the subjects giving genuine signatures were encouraged by the operators to give their real signature that expresses their individualities. Also, subjects giving forgeries were advised to practice the signature before signing not only in terms of graphics, but also in terms of speed and accelerations. They have been trained to reproduce the dynamics of the signature even by the person that they tried to forge.

Table 1 summarizes the sessions in the acquisition process of the signatures database. In the acquisition process, several measures were taken in order to enhance data collecting such as day breaks between acquisition sessions or practicing forgeries.

Table 1: Sessions of collecting the database.

Dataset	Number of signatures	Signatures Type	No. of Samples/ Subject
Session 1	1130	Genuine	10
Session 2	1130	Genuine	10
Session 3	1130	Genuine	10
Session 4	1130	Genuine	10
Session 5	2260	Genuine & Forgery	20
Session 6	1130	Forgery	10

3.2 Post-processing data collection

During the acquisition process, we have found two types of errors that could alter the signatures from that database: acquisition errors and operating errors.

The acquisition errors are caused by the improper handling of the electronic pen, for instance by holding it at an improper angle or spin related to its own axis. The visual feedback provided by the acquisition application represents the first step in order to minimize this type of error. Also, at the end of the acquisition process, these errors were removed by visual inspection of each set of signatures images belonging to a subject. The operating errors are caused by mixing the subject IDs and signatures. This type of errors was eliminated by defining validation rules in the acquisition application and in the database. By using generated forms, customized for each subject, for handwritten signatures in the acquisition process, we have eliminated operator's errors that can occur by mixing subject IDs.

After removing the errors, we used z-score or standard score which is a common statistical method for data standardization. We apply z-score to detect signatures having significantly differences in length from the rest of signatures collected from a user. The length variation may appear due to the fact that this type of signatures may be incorrectly acquired in terms of the acceleration signals and graphic form.

Also, the properties of the normal distribution were helpful in assessing and improving out signature data set (Larsen and Marx, 2006).

We calculated z-score in order to understand how various subsets of data signatures contribute to the authentication system performance results. Z-scores were used to find subsets of signatures which could negatively affect the rates in the evaluation process. We computed z-scores by formula (1), where μ represents the mean of the distribution and σ the standard deviation.

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

By using the three-sigma rule or 68-95-99.7 rule, we have considered the interval $\mu \pm 2\sigma$ a 95% confidence interval (2).

$$\Pr(\mu - 2\sigma \le x \le \mu + 2\sigma) \approx 0.09545 \tag{2}$$

For every subject in the database we have computed the z-score for each of his genuine signature. Then, we eliminated from the dataset the signatures with a z-score outside the interval of confidence. For the 95% prediction interval chosen, the corresponding z-score is 1.96 (Larsen and Marx, 2006). The score is computed in terms of the quantile function using the formula (3), where Φ represents the standard normal cumulative distribution function; μ represents the mean of the distribution and σ the standard deviation.

$$\Phi^{-1}_{\mu,\sigma^2}(p) = \mu + \sigma\Phi^{-1}(p) \tag{3}$$

Another procedure applied on the signatures database was to eliminate the signatures given by the subject in the first acquisition session, when he was not fully accustomed to the biometric acquisition device. We observed that by removing the signatures acquired in the first acquisition session we obtain better results in the evaluation phase.

4 EXPERIMENTAL RESULTS

In order to prove the effects of the post-processing procedures presented in previous section we will

compute two performance coefficients: FAR (False Accept Rate) and FRR (False Reject Rate) over the collected signatures database. To compute these performance coefficients we will use an online authentication system. It is a distance-based system containing a set of algorithms which compare two strings of symbols extracted from the two signatures: the input signature from the authentication phase and the specimen signatures from the registration phase. The pen signals are translated into these symbols arrays called "invariants" which have attached a cost and are used in several algorithms like the „Levenshtein" algorithm (Andrei, Rusu, Diaconescu and Dinescu, 2011).

We compute the performance coefficients FAR and FRR for each user registered in the database, using 5 signatures declared as being genuine, and we consider them to be specimens. Then, for each user, we send all the remaining genuine signatures to be authenticated, computing the FRR coefficient. For all subjects in the database, to compute the FAR, we send for authentication all the signatures that were captured as forgeries for a subject.

The performance coefficients obtained when using the collected database before any of the post-processing procedures presented above are: FRR 19.44% and FAR 2.29%. By removing the acquisition errors and the operating errors, the FRR decreased with 2.29 percentages, meaning that there will be more with 2.29 percentages genuine signatures accepted correctly. The FAR also decreased with 0.45 percentages, meaning that more forgery attempts will be rejected correctly. By applying also z-score for data standardization over the post-processed dataset, we obtain the same value for FRR while the value for FAR decreased with another 0.57 percentages. The above results prove that, in order to build a strong data collection of dynamic signatures, you need to make sure that the signatures of a user are consistent; otherwise there will be negative effects in the evaluation process.

We observed that, the additional method applied for eliminating the set of signatures acquired in the first session (Table 1) of the acquisition process, further improves results: FRR decreased with 2.56 percentages. This is due to the fact that the signer was not fully accustomed to the signing pen in the first day of acquisition. After applying all post-processing procedures mentioned above, over the data collection, the performance coefficients improved: FRR decreased with approximately 5 percentages, while FAR decreased with approximately 1 percentage.

5 CONCLUSIONS

In this paper we presented our methods to achieve a major step in evaluating performances of online signature authentication systems: data collection. We described our acquisition process, acquisition device and acquisition application. We presented our approach used to collect genuine signatures and forgeries. We applied post-processing methods on the collected database. Besides, using the raw databases and also the post-processed database we report the performance results of our dynamic signature verification system. The obtained results support the claim that signature data collecting strategies have an impact over the performance coefficients of an authentication system.

In future works, it is interesting to study how the performance results will change if we use forgeries collected from professional forgers which would be motivated to break the system and also signatures which are collected over a long period of time.

REFERENCES

Rusu, S., Dinescu, A., Diaconescu, S., 2011. *Systems and methods for assessing the authenticity of dynamic handwritten signature.* World Intellectual Property Organization WO/2011/112113.

Bernadette, D., Chollet, G., Petrovska-Delacrétaz, D., 2009. *Guide to Biometric Reference Systems and Performance Evaluation*. In London: Springer-Verlag London, pp. 125-166.

Yeung, D.Y., Chang, H., Xiong,Y., George, S., Kashi, R., Matsumoto, T., Rigoll, G., 2004. *SVC2004: First International Signature Verification Competition*. In Springer LNCS, Volume 3072, pp. 16-22.

Houmani, N., Mayoue, A., Garcia-Salicetti, S., Dorizzi, B., Khalil, M. I., Moustafa, M. N., Abbas, H., Muramatsu, D., Yanıkoğlu, Berrin, Kholmatov, Alisher, Martinez-Diaz, M., Fierrez, J., Ortega-Garcia, J., Roure Alcobe, J., Fabregas, J., Faundez-Zanuy, M., Pascual-Gaspar, J. M., Carednoso-Payo, V., Vivaracho-Pascual, C., 2011. *BioSecure signature evaluation campaign (BSEC'2009): evaluating online signature algorithms depending on the quality of signatures.* In: Pattern Recognition Volume 45, Issue 3, pp. 993-1003.

Larsen, R. J., Marx, M. L., 2006. *An Introduction to Mathematical Statistics and Its Applications*. 4th ed. Prentice Hall, pp. 308-869.

Andrei, V., Rusu, M. S., Diaconescu S., Dinescu, A., 2011. *Securing On-line Payment using Dynamic Signature Verification.* LISS (3) 2011, pp. 58-65.