Online Signature Verification: Is the Whole Greater Than the Sum of the Parts?

Marianela Parodi(✉) and Juan Carlos Gómez

Laboratory for System Dynamics and Signal Processing, FCEIA, National University of Rosario, CIFASIS, CONICET, Rosario, Argentina
parodi@cifasis-conicet.gov.ar

Abstract. To choose the best features to model the signatures is one of the most challenging problems in online signature verification. In this paper, the idea is to evaluate whether it would be possible to combine different feature sets selected by different criteria in such a way that their main characteristics could be properly exploited and the verification performance could be improved with respect to the case of using each set individually. In particular, the combination of an automatically selected feature set, a feature set inspired by the ones used by Forensic Handwriting Experts (FHEs), and a set of global features is proposed. Two different fusion strategies are used to perform the combination, namely, a decision level fusion scheme and a pre-classification scheme. Experimental results show that the proposed feature combination approaches result not only in improvements regarding the verification error rates but also the simplicity, flexibility and interpretability of the verification system.

Keywords: Online signature verification · Forensic handwriting examination · Information fusion

1 Introduction

Automatic signature verification is an important research area in the field of biometrics [1], being the most popular method for identity verification. Signatures are recognized as a legal means of identity verification by financial and administrative institutions, and people is familiar with their use in everyday life.

Two categories of signature verification systems can be distinguished, namely, offline (only the image of the signature is available), and online (dynamic information acquired during the signing process is available). The interest in the online approach has increased in recent years due to the widespread use of electronic pen-input devices in many daily applications. In addition, it is reasonable to expect that the incorporation of dynamic information would make signatures more difficult to forge. Nevertheless, there are certain applications that demand the use of the offline approach. For example, Forensic Handwriting Experts (FHEs) only have the signature image available in their daily work, although in the future it might occur that FHEs will also have to deal with online signatures.

© Springer International Publishing Switzerland 2015
A. Pardo and J. Kittler (Eds.): CIARP 2015, LNCS 9423, pp. 151–159, 2015.
DOI: 10.1007/978-3-319-25751-8_19

In online systems, the signature is parameterized by discrete time functions, such as pen coordinates, velocity and pressure, among others. Researchers have long argued about their effectiveness for verification purposes, and the conflicting results make the discussion still open [2], [3]. To decide which features extract from the available time functions is also an important design step. Local (computed for each point in the time sequence) and global (computed from the whole signature) features can be considered [4], [5].

In this paper, the idea is to evaluate whether it would be possible to combine different online feature sets selected by different criteria taking advantage of their main characteristics in order to improve the verification performance with respect to the case of using them individually. The combination of three feature sets that have already proved to have interesting qualities, resulting not only in good verification performances, but also providing different advantages to the verification systems, is then proposed. In particular, an automatically selected feature set, a set of features relevant to FHEs, and a global feature set are combined and the discriminative power of the resulting combination is evaluated. The advantages of using each of these feature sets will be highlighted along this paper. Two different strategies are proposed for the combination of the feature sets. One is based on a decision level fusion strategy and the other one on a pre-classification approach. A well known state-of-the-art classifier, namely, Random Forest (RF), is used to perform the verification experiments. The verification performance of the proposed combination approaches is evaluated for two different signature styles of a publicly available database, namely, Western (Dutch) and Chinese.

2 Feature Selection

Typically, the measured data consists of three discrete time functions: pen coordinates x and y, and pressure p. Several extended time functions are usually computed from them [5], [6]. In this paper, the velocity magnitude v_T and direction θ, the total acceleration a_T and the log-radius curvature ρ are computed. The first and second order time derivatives of these functions are also computed. The different features are then extracted from the above mentioned time functions, as described in the following Subsections.

2.1 Global Features

In [7], a set of widely used global features corresponding to the better ranked ones in [5] and [4] is used. These global features (hereafter referred to as GF) will be the ones considered in this paper (for both, Dutch and Chinese signatures), and they are: signature total time duration T, pen down duration T_{pd}, positive x velocity duration T_{vx}, average pressure $\bar{P}$, maximum pressure P_M and the time of maximum pressure T_{P_M}. To include global features to the combination would simplify and improve the interpretability of the system since they are simple, intuitive and easily to compute and compare.

2.2 Time Function Based Features

A wavelet approximation of the time functions is proposed to model them. The Discrete Wavelet Transform (DWT) decomposes the signal at different resolution levels, splitting it in low (*approximation*) and high (*details*) frequency components. The idea is to use the DWT approximation coefficients to represent the time functions. In particular, the widely used `db4` wavelet is employed. Resampling of the time functions, previous to the DWT decomposition, is needed in order to have a fixed-length feature vector.

Automatically Selected Time Function Based Features. In [8], an automatic feature selection based on the variable importance provided by the RF algorithm is performed from the original set of time functions listed at the beginning of Section 2. The automatically selected features are: x, a_T, y, v_T, p, dp, ρ, dx, θ, dy, d^2x, d^2y and dv_T for the Dutch data, and y, x, p, v_T, a_T, dy, dx, d^2y, θ, ρ, dp, d^2x, $d\theta$, d^2p, dv_T, $d\rho$ and $d^2\theta$ for the Chinese data. Here, df and d^2f denote de first and second order time derivatives of the corresponding time function f, respectively. Note that different features are selected for each signature style, then to include these features to the combination will improve its flexibility and capability to adapt to each type of signature. These feature sets will be referred to as ASF.

FHE Based Features. Although FHEs work with the static image of the signature, they can infer some dynamic properties from it. FHEs consider velocity and curvature as distinctive features, since in natural handwriting the stroke velocity is determined by its curvature, while in a forgery process this would not be the case. The pen pressure is not useful for them since it is strongly dependant on external factors such as the writing material and surface, although the pressure fluctuations are highly individualistic to the writer. In this paper, the set of features presented in [8], hereafter referred to as TFFHE, is considered as the one relevant to FHEs: v_T, θ, ρ and dp. To include these features to the combination would make it meaningful for FHEs, then the system could be integrated into toolkits that could be useful for them. This could contribute towards bridging the gap between the FHE and the Pattern Recognition (PR) communities. The TFFHE features have been selected based on FHE criteria for Latin scripts. Since information about FHE criteria for Chinese scripts was not available for the authors, the same TFFHE set is used for both signature styles.

3 Feature Combination Approaches

Two different combination strategies are proposed. One of them is based on a decision level fusion (DLF) approach, while the other is based on a pre-classification (PC) of the signatures so that gross forgeries can be early detected and discarded. They are described in Subsections 3.1 and 3.2, respectively.

3.1 Decision Level Fusion

Traditionally, three main approaches for information fusion can be distinguished, namely, feature, classifier or decision level fusion. In the feature level case, the feature vectors coming from different sources are concatenated to obtain a combined feature vector which is then used in the classification task. In the classifier level approach, a composite classifier is generated by combining the individual classifiers used to process the different signals involved. Finally, in the decision level approach, a final decision is obtained by combining the probability/likelihood scores from the separate classifiers processing the different signals.

In this paper, classifier level fusion is not possible due to the particular classifier being used (RF). Regarding a fusion at feature level, it is clear that since the ASF feature set contains the TFFHE set, feature level fusion of these two sets would not make sense. Two separate experiments fusing GF features with ASF features on one hand, and fusing GF features with TFFHE features on the other, were carried out. The verification results obtained (not shown here) did not improve the ones corresponding to the case of using the ASF and the TFFHE feature sets individually.

Based on the above comments, only DLF is considered in this paper. Three independent RF classifiers are fed by each type of features (GF, ASF and TFFHE) and the final decision is computed as a combination of the likelihood scores associated which each classifier based on the widely used *weighted geometrical combination rule*, that is:

$$P_{fused} = P_{GF}^{\beta} \, P_{ASF}^{\gamma} \, P_{TFFHE}^{(1-\beta-\gamma)}, \tag{1}$$

where P_{fused} is the likelihood score for the combined scheme, P_{GF}, P_{ASF} and P_{TFFHE} are the likelihood scores for the classifiers based on GF, ASF and TFFHE features, respectively, and $0 \leq \beta \leq 1$ and $0 \leq \gamma \leq 1$ are user defined parameters weighting the individual likelihood scores.

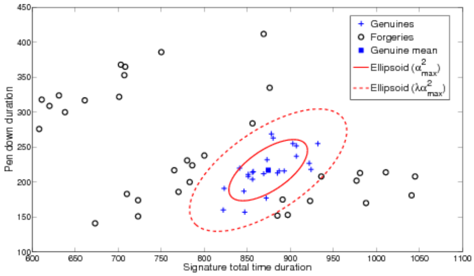
3.2 Pre-classification

It would be reasonable to expect that for gross forgeries some features such as global features and the ones based on the FHE criteria, would present a wide variability. This leads to the idea of using GF and TFFHE features for pre-classification in order to quickly recognize and discard gross forgeries.

In this paper, a multivariate version of the univariate PC approach introduced in [9] is proposed. The decision rule is shown in Fig. 1 (right), where $\mathbf{g}_{test}$ denotes the feature vector corresponding to the test signature, $\bar{\mathbf{g}}_{train}$ and Σ_{train} are the feature vector sample mean and sample covariance over the genuine training set, respectively, and α is a coefficient defining the threshold. The decision rule means that signatures whose feature vectors lie outside the hyperellipsoid define as $(\mathbf{g}_{test}-\bar{\mathbf{g}}_{train})^T \Sigma_{train}^{-1}(\mathbf{g}_{test}-\bar{\mathbf{g}}_{train}) = \alpha^2$, are considered as forgeries. Figure 1 (left) illustrates this, for the case of a two-dimensional feature vector. Coefficient α^2 is computed as:

$$\alpha^2 = \max_{A} \max_{A_i} \left\{ (\mathbf{g}_{test} - \bar{\mathbf{g}}_{train})^T \Sigma_{train}^{-1} (\mathbf{g}_{test} - \bar{\mathbf{g}}_{train}) \right\}, \quad (2)$$

where A is the set of all the authors in the Training Set and A_i denotes the i-th author in the same set.



Decision Rule

If $(\mathbf{g}_{test} - \bar{\mathbf{g}}_{train})^T \Sigma_{train}^{-1} (\mathbf{g}_{test} - \bar{\mathbf{g}}_{train}) > \alpha^2$

then signature=FORGERY

else continue classification

Fig. 1. Left: Distribution of the global feature vectors for the genuine (+) and forged (o) signatures of an author in the database. A bounding ellipsoid ((red) solid line) as defined above, and an enlarged ellipsoid ((red) dashed line) as defined in Section 5, are also represented. In this case, the feature vector is composed by T and T_{PD}. Right: Decision rule.

In this paper, two different experiments employing the PC approach are proposed. One of them, referred to as PC-GF, uses GF features for PC while the subsequent classification stage is performed based on a DLF between two RF classifiers fed by ASF and TFFHE features, respectively. The other one, referred to as PC-FHE, uses TFFHE features for PC while for the subsequent classification stage employs a DLF between two RF classifiers fed by ASF and GF features, respectively.

4 Evaluation Protocol

The SigComp2011 Dataset [10] is used for the experiments. Since it contains Dutch and Chinese signatures, the influence of the cultural origin of the signatures in the verification performance can be evaluated, which is crucial in order for the system to be widely accepted. Each dataset is divided into a Training and a Testing Set. Skilled forgeries (simulated signatures in which forgers are allowed to practice the reference signature for as long as they deem it necessary) are available. The measured data are: pen coordinates x and y, and pressure p.

To evaluate the performance, the Equal Error Rate (EER) and the cost of the log-likelihood ratios ($\hat{C}_{llr}$ and $\hat{C}_{llr}^{min}$) are computed. A smaller value of $\hat{C}_{llr}^{min}$ (minimal possible value of $\hat{C}_{llr}$) indicates a better performance of the system. The use of log-likelihood ratios has been recommended by the experts in the lastest main conferences of the area since they allow FHEs to give an opinion on the strength of the evidence.

The optimization of the tunning parameters of the proposed verification systems is performed over the Training Set, while the Testing Set is used for independent testing purposes. To obtain statistically significant results, a 5-fold cross-validation (5-fold CV) is performed over the Testing Set to estimate the verification errors. Forgeries are not usually available in real applications during the training phase, then only genuine signatures are used for training purposes.

4.1 Decision Level Fusion Approach

For each instance of the 5-fold CV, a signature of a particular writer from one of the testing sets in the 5-fold CV is fed to the system. The GF, ASF and TFFHE features are computed. Then, three RF classifiers are trained using GF, ASF and TFFHE features, respectively. Each classifier is trained by a genuine class consisting of the current writer's genuine class in the training set of the 5-fold CV, and a forged class consisting of the genuine signatures of all the remaining writers in the same set. The result of the verification process is then the combination of the outputs of these three RF classifiers computed as in (1).

4.2 Pre-classification Approach

For each instance of the 5-fold CV, a signature of a particular writer from one of the testing sets in the 5-fold CV is fed to the system. The GF (for the PC-GF) or the TFFHE (for the PC-FHE) features are computed to construct $\mathbf{g}_{test}$. Then, the distance between $\mathbf{g}_{test}$ and $\bar{\mathbf{g}}_{train}$ (sample mean computed over the current writer's genuine signatures available in the training set of the 5-fold CV) is computed. If this distance is larger than the threshold (α^2), the signature is declared to be a forgery. If this is not the case, the signature is subjected to the subsequent classification stage, where two RF classifiers are trained, one of them with the ASF features and the other one with the TFFHE (for the PC-GF) or the GF (for the PC-FHE). Each RF classifier is trained as described in Subsection 4.1. A DLF is performed over the two RF classifier outputs, giving the final output of this classification stage. Then, the result of the verification process is either the result of the PC (the input signature is declared to be a forgery), or the one of the DLF of the two RF classifiers. Note that, in case the result is given by the PC, the verification process is simplified and speeded up.

5 Results and Discussion

The tuning parameters are optimized over the corresponding Training Sets. For both approaches, the number of trees and randomly selected splitting variables in the RF classifiers were set to 500 and $\sqrt{P}$ (being P the feature vector dimension), respectively. The time functions were resampled to a normalized length of 256. Regarding the resolution level, a better approximation accuracy is obtained using a lower resolution level, at the cost of increasing the amount of the modeling DWT coefficients. To increase the amount of DWT coefficients to model each time function is not a limitation when using the TFFHE features since the feature vector contains only four features, although it will significantly increase the feature vector dimension in the case of using the ASF features. Then, the DWT resolution level was set to 2 when computing the TFFHE features, and to 3 in the case of the ASF ones.

For the PC approach, parameter α is computed resorting to (2) over the Training Sets. Experiments carried out over these sets showed that such a computation of α leads to several genuine signatures lying outside the defined hyper-ellipsoid and so wrongly classified as forgeries. This is probably due to the fact that α is always computed over a separate subset of genuine signatures used exclusively for training purposes, without taking into account the forgeries which are also available in the Training Sets. The experiments also showed that it is possible to enlarge the hyper-ellipsoid in such a way that less genuine signatures lie outside it. This is illustrated in Fig. 1, where an enlarged ellipsoid containing the original one has been plotted in (red) dashed line. Then, better results can be obtained by redefining the decision threshold multiplying α^2 by a factor $\lambda > 1$. The parameter λ was also optimized over the Training Sets, being set to $\lambda = 5$ for both PC approaches (PC-GF and PC-FHE), and both datasets. For the DLF approach, the parameters β and γ are optimized by minimizing $\hat{C}_{llr}^{min}$ over the Training Sets, being set to $\beta^{Dutch} = 0.2$ and $\gamma^{Dutch} = 0.5$ for Dutch data, and $\beta^{Chinese} = 0.1$ and $\gamma^{Chinese} = 0.8$ for Chinese data.

The verification results obtained when using the PC-GF approach are shown in the first row of Table 1, for the Dutch (left) and Chinese (right) datasets, respectively, while the ones corresponding to the PC-FHE approach are not good and it does not make sense to include them in Table 1. The verification results corresponding to the case of using the DLF approach are presented in the second row of Table 1. The best verification results for the case of using each feature set individually correspond to the case of using the ASF features, and they are shown in the third row of Table 1. In addition, state-of-the-art results corresponding to the best commercial and non-commercial systems in the SigComp2011 Competition reported over the same datasets [10], are included in the last two rows of Table 1 (information about the EER was not given).

Table 1. Verification results for the Dutch (left) and Chinese (right) Datasets

	Dutch Dataset			Chinese Dataset		
	EER	$\hat{C}_{llr}$	$\hat{C}_{llr}^{min}$	EER	$\hat{C}_{llr}$	$\hat{C}_{llr}^{min}$
PC-GF	**3.55**	**0.172**	**0.133**	**5.1**	**0.194**	**0.162**
DLF	6.95	0.261	0.228	7.31	0.268	0.218
ASF	6.58	0.243	0.205	7.455	0.296	0.248
Comm.	–	0.259	0.123	–	0.413	0.218
Non-comm.	–	0.493	0.237	–	0.565	0.351

It can be observed that the PC-GF approach obtains better verification results than the DLF one, for both datasets. In addition, the PC-GF approach outperforms the ASF one (for both datasets), while the DLF approach outperforms the ASF one only for the Chinese data. Note also that the proposed combinations obtain results comparable to the best ones in the state-of-the-art, being even better than the non-commercial systems. Moreover, in the case of the Chinese data, results are better than the ones corresponding to the commercial system, which is particularly promising since Chinese signatures are usually more complex than Western ones.

Based on the above discussion, the best combination strategy is the PC-GF approach, that is to use GF features for PC and to perform DLF with the remaining information (ASF and TFFHE features).

Finally, the obtained verification results (shown in Table 1) allow to answer the question in the title of the paper for the positive.

6 Conclusions

The feasibility of combining different feature sets selected by different criteria so that their main characteristics could be properly exploited was evaluated. The experimental results show that the best combination strategy is to use GF for PC and to perform DLF with the additional information (ASF and TFFHE features). The results obtained in this case outperforms the ones obtained when using the feature sets individually. In addition, they are comparable to the best results reported in the state-of-the-art. In particular, for Chinese signatures, they are even better than the best result in the state-of-the-art. This is a promising result since this data is usually more difficult to deal with and, for this reason, it is considered more challenging. Finally, since the best combination scheme is based on a PC, the resulting verification process is simplified and speeded up.

References

1. Impedovo, D., Pirlo, G.: Automatic signature verification: The state of the art. IEEE T. on Syst., Man, and Cybern.-Part C: Appl. and Rev. **38**(5), 609–635 (2008)
2. Muramatsu, D., Matsumoto, T.: Effectiveness of pen pressure, azimuth, and altitude features for online signature verification. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 503–512. Springer, Heidelberg (2007)
3. Houmani, N., Salicetti, S.G., Dorizzi, B.: On assessing the robustness of pen coordinates, pen pressure and pen inclination to time variability with personal entropy. In: Proc. IEEE Int. Conf. on Biom.: Theory, Appl., and Syst., USA (2009)
4. Fierrez-Aguilar, J., Nanni, L., Lopez-Peñalba, J., Ortega-Garcia, J., Maltoni, D.: An on-line signature verification system based on fusion of local and global information. In: Proc. IAPR Int. Conf. on Audio- and Video-Based Biomet. Person Authent., Rye Brook, NY, USA, pp. 523–532 (2005)
5. Richiardi, J., Ketabdar, H., Drygajlo, A.: Local and global feature selection for on-line signature verification. In: Proc. Int. Conf. on Doc. Anal. and Recognit., Seoul, Korea (2005)
6. Fierrez-Aguilar, J., Ortega-Garcia, J., Ramos-Castro, D., Gonzalez-Rodriguez, J.: HMM-based on-line signature verification: feature extraction and signature modelling. Pattern Recognit. Lett. **28**, 2325–2334 (2007)

7. Parodi, M., Gomez, J.C.: Using Global Features for Pre-classification in Online Signature Verification Systems. In: Emerging Aspects in Handwritten Signature Verification, pp. 39–52. World Sci. (2014)
8. Parodi, M., Gómez, J.C., Liwicki, M., Alewijnse, L.: Orthogonal function representation for online signature verification: which features should be looked at? IET Biomet. **2**(4), 137–150 (2013)
9. Lee, L.L., Berger, T., Aviczer, E.: Reliable on-line human signature verification systems. IEEE T. on Pattern Anal. and Mach. Intell. **18**(6), 643–647 (1996)
10. Liwicki, M., Malik, M.I., den Heuvel, C.E., Chen, X., Berger, C., Stoel, R., Blumenstein, M., Found, B.: Signature verification competition for online and offline skilled forgeries (SigComp2011). In: Proc. 11th Int. Conf. on Doc. Anal. and Recognit., China (2011)