Online Signature Verification: Improving Performance through Pre-classification Based on Global Features

Marianela Parodi and Juan Carlos Gómez

Lab. for System Dynamics and Signal Processing, FCEIA,
Universidad Nacional de Rosario, and CIFASIS, Argentina
{parodi,gomez}@cifasis-conicet.gov.ar

Abstract. In this paper, a pre-classification stage based on global features is incorporated to an online signature verification system for the purposes of improving its performance. The pre-classifier makes use of the discriminative power of some global features to discard (by declaring them as forgeries) those signatures for which the associated global feature is far away from its respective mean. For the remaining signatures, features based on a wavelet approximation of the time functions associated with the signing process, are extracted, and a Random Forest based classification is performed. The experimental results show that the proposed pre-classification approach, when based on the apppropriate global feature, is capable of getting error rate improvements with respect to the case where no pre-classification is performed. The approach also has the advantages of simplifying and speeding up the verification process.

Keywords: Online Signature Verification, Global Features, Pre-classification.

1 Introduction

Signature verification is the most popular method for identity verification since people are familiar with the use of signatures in their everyday life [1]. Two categories of signature verification systems can be distinguished taking into account the acquisition device, namely, offline (only the signature image is available) and online systems (dynamic information about the signing process is available).

In online systems, the signature is parameterized by several discrete time functions, such as, pen coordinates, pen pressure and, when available, pen altitude and azimuth angles. Researchers have long argued about the effectiveness of these different time functions for verification purposes [2], [3]. To choose the features that could be extracted from them is also an important design step. The different features can be classified into local, calculated for each point in the time sequence, and global, calculated from the whole signature. Many researchers accept that approaches based on local features achieve better performance than those based on global features, but still there are others who favor the use of global features [4], [5]. In fact, it is interesting to use global features since they

have the advantage of being simple features, usually more intuitive than local ones, and can be easily computed and compared. Furthermore, it would be reasonable to expect that local and global features could provide complementary information [5]. To combine both types of features in such a way that their main characteristics could be exploited, is an interesting and still open challenge. In [6], an online signature verification system is designed as a multilevel system which uses three different signature representations, one based on global features and the other two based on local features. In addition, several fusion strategies have been proposed in the literature. In [5], fusion strategies based on the `max` and `sum` rules are compared. Some approaches using a pre-classification stage based on global features for the purposes of early detecting bad forgeries, have been proposed in the mid 1990's in [6] and [7]. Global features, such as signature total time duration and pen down duration, have been used in those papers.

In this paper, global based features are used for pre-classification purposes. The idea is to pre-classify signatures, declaring as forgeries those that are far away from their mean, in terms of the global based feature being considered. It is expected that this could help to quickly recognize and classify gross forgeries, speeding up and simplifying the verification process. The remaining signatures continue with the subsequent classification stage which consists in extracting new features (time function based ones modeled by the coefficients in their wavelet approximations) in order to achieve a more detailed representation, and classifying them on the basis of a Random Forest (RF) classifier. For the verification experiments, two different signature styles are considered, namely, Western and Chinese, of one of the most recent publicly available Signature Databases.

2   Pre-classification Approach

In this paper, features based on global parameters as well as features based on time functions (modeled by their wavelet based approximations), are used. Since these different types of features could provide complementary information, the idea is to exploit their intrinsic characteristics. When using global based features, it would be expectable to get a rough and quick representation of the signature. This could be useful to have an idea of some distinctive characteristics and detect some anomalies of the signature. On the other hand, if a more precise representation is needed, the time function based features could provide more detailed information, at the cost of a more time consuming feature extraction.

Global based features are then used for pre-classification purposes. It is reasonable to expect that some global based features, such as signature total time duration, pen down duration and average pressure, for the genuine samples would be far away from the corresponding ones for the forged samples. This is illustrated in Fig. 1 (left), where the distributions of the global based feature "signature total time duration" for the genuine (left) and forged (right) signatures of an author in the database, are depicted.

The idea is then to classify as a forgery those signatures for which the global features differ significantly from the corresponding genuine feature mean.
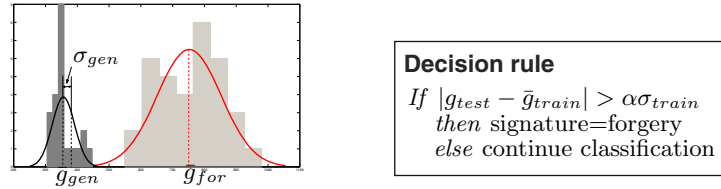
Fig. 1. Distribution of the global feature "signature total time duration" for the genuine and forged signatures of an author in the database (left). Pre-classification decision rule (right).

In particular, the decision rule shown in Fig. 1 (right) is considered, being g_{test} the global based feature corresponding to the test signature, $\bar{g}_{train}$ and σ_{train} the global based feature mean and standard deviation values over the genuine training set, respectively, and α a coefficient defining the threshold.

Coefficient α is computed, for each global based feature, in three different ways for comparison purposes, namely, as: i. the maximum, or ii. the mean, or iii. the minimum, over all the authors in the Training Set of the database, of the maximum, over all the signatures of each author, of the absolute difference between the global feature value of the test signature and the global based feature mean of the genuine training set, normalized by the standard deviation of the set. That is (for case i.):

$$\alpha = \max_{A} \max_{A_i} \left\{ \frac{|g_{test} - \bar{g}_{train}|}{\sigma_{train}} \right\} \tag{1}$$

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.

A different approach is considered in [7], where only the signature total time duration is used for pre-classification purposes. In that paper, the threshold is computed as a fraction of the $\bar{g}_{train}$, and it is heuristically set to 0.2.

3   Feature Extraction

Typically, the measured data consists of three discrete time functions: pen coordinates x and y, and pen pressure p. In addition, several extended functions can be computed from them [4], [8]. Previous to the feature extraction, the original x and y pen coordinates are normalized regarding scale and translation.

3.1   Global Based Features

Several global based features can be extracted from the measured and extended time functions. In [4], a feature selection is performed on a set of 46 global features which seem to be the most commonly used in the literature. In [5], subsets of global features are selected from an initial set of 100 features. In the present paper, the global based features are selected to be discriminative enough in order for the proposed pre-classification to be succesful. In order to

analyze and compare their individual discriminative power, the following global based features, corresponding to the better ranked ones by the feature selection performed in [4] and [5], are used in this paper: signature total time duration T, pen down duration T_{pd}, positive x velocity duration T_{v_x}, average pressure $\bar{P}$, maximum pressure P_M and the time at which the pressure is maximum T_{P_M}.

3.2   Time Function Based Features

Several extended time functions can be computed from the measured ones. In this paper, the path velocity magnitude v_T, the path-tangent angle θ, the total acceleration a_T and the log curvature radius ρ are computed as in [8]. The set of time function based features is then composed by x, y, p, v_T, θ, a_T and ρ, and their first and second order time derivatives.

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 here is to use the DWT approximation coefficients to represent the time functions. Resampling of the time functions, previous to the DWT decomposition, is needed in order to have a fixed-length feature vector. To use a fixed-length feature vector represents an advantage since it makes the comparison between two signatures easier. In [3] and [9], fixed-length representations are proposed based on the Fast Fourier Transform (FFT) and Legendre Polynomials series expansions, respectively. The approximation accuracy is determined by the chosen resolution level, which also determines the length of the resulting feature vector. Since this length has to be kept reasonably small, there will be a trade-off between accuracy and feature vector length. The design parameter is then the length of the feature vector, which determines the resolution level to be used. The widely used db4 wavelets [10] is employed for the representation of the time functions.

4   Experiments

The SigComp2011 Dataset [11] is used for the verification experiments. It has two separate datasets, containing Western (Dutch) and Chinese signatures, respectively. Each dataset, is divided into a Training and a Testing Set. Skilled forgeries (in which forgers are allowed to practice the reference signature for as long as they deem it necessary) are available. The signatures were acquired using a ballpoint pen on paper over a WACOM Tablet, which is the natural writing process. The measured data are the pen coordinates x and y, and pressure p.

In order to compare the discriminative power of the different global based features, experiments using each one of them for pre-classification purposes were carried out. In addition, the three different ways of computing the coefficient α, defining the threshold $\alpha\sigma_{train}$, are also tested.

For each dataset (Dutch and Chinese), the optimization of the meta-parameters of the system is performed over the corresponding Training Set while the

corresponding Testing Set is used for independent testing purposes. The meta-parameters of the system are: α for the pre-classification stage, the normalized length of the resampled time functions, the resolution level for the wavelet approximations, and the number of trees and randomly selected splitting variables for the RF classifier. To obtain statistically significant results, a 5-fold cross-validation (5-fold CV) is performed over the Testing Set to estimate the verification errors. 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. After its pre-processing, one of the global based features (g_{test}) presented in Subsection 3.1 is extracted from it. Then, the pre-classification is performed as follows: the g_{test} value of the input signature is compared with the $\bar{g}_{train}$ (mean value of the same feature) computed over the current writer's genuine signatures available in the corresponding training set of the 5-fold CV. If the difference between these values is larger than the corresponding threshold for that global based feature, the signature is declared to be a forgery. If this is not the case, the signature is subjected to the subsequent classification stage, as follows: the DWT approximation coefficients are computed for the different time funtions presented in Subsection 3.2. Then, a RF [12] classifier is trained by the genuine class consisting of the current writer's genuine signatures available in the corresponding training set of the 5-fold CV, and a forged class which consists of the genuine signatures of all the remaining writers in the dataset available in the same training set. The result of the verification process is then either the result of the pre-classification (the input signature is considered a forgery), or the result of the RF classifier. If the result is given by the pre-classification, the verification process is speeded up.

To evaluate the performance, the EER (Equal Error Rate) is calculated, using the Bosaris toolkit, from the Detection Error TradeOff (DET) Curve as the point in the curve where the FRR (False Rejection Rate) equals the FAR (False Acceptance Rate). The cost of the log-likelihood ratios $\hat{C}_{llr}$ and its minimal possible value $\hat{C}_{llr}^{min}$ [13] are computed using the toolkit as well [14]. A smaller value of $\hat{C}_{llr}^{min}$ indicates a better performance of the system.

5 Results and Discussion

The experiments were performed using 500 trees and $\sqrt{P}$ randomly selected splitting variables (P = feature vector dimension), for the RF classifier. The time functions were resampled to a normalized length of 256. The wavelet resolution level was set to 3, in order to obtain a feature vector of a reasonable length.

The verification results, with pre-classification, for the six global based features considered, and the three different values of α, are shown in Table 1, for the Dutch (left) and Chinese (right) data, respectively. The best results are indicated in **boldfaced** style. In order to analyze the advantages of using the proposed pre-classification scheme, the results obtained without pre-classification are also included in the last row section of Table 1. For the purposes of comparison, the results in [9], based on Legendre polynomials representations (without pre-classification), are also included in that section.

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

		Dutch Dataset						Chinese Dataset					
		T	T_{pd}	T_{v_x}	$\bar{P}$	P_M	T_{P_M}	T	T_{pd}	T_{v_x}	$\bar{P}$	P_M	T_{P_M}
α_{max}	EER	5.35	6.22	4.95	4.93	6.58	7.08	5.65	7.64	7.01	7.05	7,84	7.91
	$\hat{C}_{llr}$	0.2374	0.2389	0.2283	0.2123	0.2544	0.2704	0.2244	0.3017	0.2644	0.2631	0.2918	0.2849
	$\hat{C}_{llr}^{min}$	**0.2015**	0.2053	**0.167**	**0.1683**	0.2146	0.2408	**0.193**	0.2532	**0.2188**	**0.2155**	**0.2437**	0.2512
α_{mean}	EER	5.98	6.61	6.64	6.01	8.17	8.91	4.93	7.63	6.88	6.92	8.42	8.85
	$\hat{C}_{llr}$	0.2562	0.2728	0.265	0.2558	0.3019	0.3039	0.2008	0.293	0.2745	0.2719	0.3215	0.3331
	$\hat{C}_{llr}^{min}$	0.2263	0.2418	0.2413	0.2272	0.2763	0.2848	**0.1781**	0.2505	**0.228**	**0.2311**	0.2834	0.2861
α_{min}	EER	14.94	15.91	13.24	10.63	10.63	10.2	11.4	14.8	10.47	10.22	12.67	12.16
	$\hat{C}_{llr}$	0.388	0.4407	0.3839	0.3484	0.3484	0.3382	0.3336	0.4532	0.3383	0.3331	0.427	0.3849
	$\hat{C}_{llr}^{min}$	0.3827	0.4247	0.3656	0.335	0.335	0.3235	0.3183	0.4148	0.3113	0.314	0.3873	0.3589
		Without pre-class			Results in [9]			Without pre-class			Results in [9]		
EER		6.78			5.91			7.9			10.03		
$\hat{C}_{llr}$		0.2491			0.237			0.3126			0.36		
$\hat{C}_{llr}^{min}$		0.2055			0.195			0.2476			0.2969		

The actual values of α belong to the intervals: $[3,4]$, $[2,3]$ and $[1,2]$, for the max, the mean and the min criteria, respectively. When choosing α_{max}, the threshold is conservative, while when α_{min} is chosen, the threshold allows for more signatures to be pre-classified at the cost of larger errors (in the sense of classifying genuine signatures as forgeries). Then, selecting the mean criterion for computing α is a trade-off between these extreme values. This is confirmed by the results in Table 1, where the results using α_{max} are, in most of the cases, better than the results using α_{mean}, while using α_{min} leads to the worst results. In Fig. 2, the percentage of signatures that are pre-classified out of the total amount of signatures, for each of the three α criteria, are shown for the Dutch (right) and Chinese (left) data, respectively.
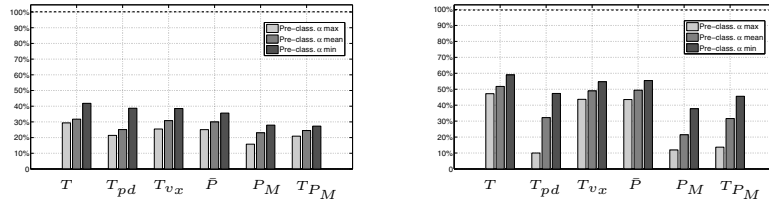

Fig. 2. Percentage of signatures pre-classified as forgeries, using the three different α values to compute the threshold

From Table 1, it can be seen that the proposed pre-classification does improve the error rates with respect to the case of not using it, and the ones in [9] (where no pre-classification takes place). This is not the case for every value of α, since when using α_{min} the results are not good due to the fact that several genuine signatures may be wrongly classified as forgeries. For the Dutch data, the error rates improve only when using the most conservative threshold

(defined by α_{max}). Then, if the trade-off (mean) value of α would to be used, there would be no error improvements. This shows that many genuine signatures are beyond $2\sigma_{gen}$. This is probably due to the fact that the feature genuine distribution is not a Gaussian. For the Chinese data, the error rates are still improved when using the trade-off α. Moreover, the best error rate is achieved in this case. Then, for this data, the threshold can be chosen to be a more robust one.

The pre-classifications based on P_M, T_{P_M} and T_{pd} did not get any error rate improvements. In the cases of P_M and T_{P_M}, this seems to be reasonable. The time at which the pressure is maximum T_{P_M} is probably an unstable feature, since people is not likely to be consistent in the time where the pen pressure reaches a peak. The value of P_M, is likely to be dependant on the writing surface, the pen, etc., making it hard to make pre-classification decisions based only on this feature. On the other hand, in the case of T_{pd}, the result was unexpected, since it is believed that forgers are not able to accurately reproduce the pen down time of the genuine writers. The global based features used for pre-classification, leading to error rate improvements, were: T, T_{v_x} and $\bar{P}$. In the case of T, the results do nothing but confirm the well known fact that this feature is a good discriminator [7]. The average pressure $\bar{P}$, is likely to be more consistent than the other pressure based features analyzed here, since people may not make considerable changes in the average pressure when signing. Finally, T_{v_x} proved to have a high discriminative power. Since T_{v_x} indicates the time in which the writer is writing forward, its discriminative power is probably due to the fact that forgers may go back several times during the writing process.

The best error rates were achieved by a pre-classification based on α_{max} and T_{v_x}, and on α_{mean} and T, for the Dutch and Chinese data, respectively. For the Chinese data, the result is not surprising since T is a highly discriminative feature. For the Dutch data, the result could be explained based on the fact that, in most of the cases, horizontal traces are more significant than vertical ones. Then, differences in the time in which the writer is writing forward would indicate that a forged signing process is taking place.

In addition to the error rate improvements, the pre-classification helps to simplify and speed up the verification process. From Fig. 2, it can be seen that, still in the most conservative cases (using α_{max}, where the amount of pre-classified signatures is minimum), an important part of the whole set of signatures is discarded (more than 40% and 25%, for the best cases of the Chinese and Dutch data, respectively), making the system to further process less signatures.

6   Conclusions

A pre-classification approach on the basis of global based features, was proposed to be included in an online signature verification system. The proposed pre-classification approach proved to be capable of exploiting the discriminative power of the global based features to improve the overall performance with respect to the case where no pre-classification is carried out. In addition, the incorporation of the pre-classification stage proved to have the advantages of simplifying and speeding up the signature verification process.

Finally, the proposed pre-classification approach has the advantage of being very simple, since it is based only on one global based feature, but proved to be powerful, allowing improvements regarding the verification errors, the process speed and the simplicity of the whole signature verification system.

References

1. Impedovo, D., Pirlo, G.: Automatic signature verification: The state of the art. IEEE Trans. on Syst., Man, and Cybern. - Part C: Appl. and Reviews 38(5), 609–635 (2008)
2. Maramatsu, D., Matsumoto, T.: Effectiveness of pen pressure, azimuth, and altitude features for online signature verification. In: Proc. of Int. Conf. on Biomet., pp. 503–512 (2007)
3. Yanikoglu, B., Kholmatov, A.: Online signature verification using fourier descriptors. EURASIP J. on Advances in Signal Process., 230–275 (2009)
4. Richiardi, J., Ketabdar, H., Drygajlo, A.: Local and global feature selection for online signature verification. In: Proc. of 8th Int. Conf. on Doc. Anal. and Recognit., Seoul, Korea (2005)
5. 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 Authentic., New York, NY, USA, pp. 523–532 (2005)
6. Plamondon, R.: The design of an on-line signature verification system: from theory to practice. Int. J. on Pattern Recognit. and Artif. Intell. 8(3), 795–811 (1994)
7. Lee, L.L., Berger, T., Aviczer, E.: Reliable on-line human signature verification systems. IEEE Trans. on Pattern Anal. and Machine Intell. 18(6), 643–647 (1996)
8. 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)
9. Parodi, M., Gómez, J.C., Liwicki, M.: Online signature verification based on legendre series representation. Robustness assessment of different feature combinations. In: Proc. of 13th Int. Conf. on Frontiers in Handwriting Recognit, Bari, Italy, pp. 377–382 (September 2012)
10. Daubechies, I.: Ten Lectures on Wavelets. SIAM, Pennsylvania (1992)
11. 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: Proceedings of 11th Int. Conf. on Doc. Anal. and Recognit., Beijing, China (September 2011)
12. Breiman, L.: Random forests. Technical report, Stat. Dep., Univ. of California, Berkeley (2001)
13. Gonzalez-Rodriguez, J., Fierrez-Aguilar, J., Ramos-Castro, D., Ortega-Garcia, J.: Bayesian analysis of fingerprint, face and signature evidences with automatic biometric systems. Forensic Sci. Int. 155, 126–140 (2005)
14. Brümmer, N., du Preez, J.: Application-independent evaluation of speaker detection. Comput. Speech & Language 20, 230–275 (2006)