Online SignatureVerification: Automatic Feature Selection vs. FHE's Choice

Marianela Parodi [1], Juan C. Gómez [1], Linda Alewijnse [2], Marcus Liwicki [3]

Abstract. *In this paper, the discriminative power of a set of features which seems to be relevant to signature analysis by Forensic Handwriting Experts (FHEs) is analyzed and particularly compared to the discriminative power of automatically selected feature sets. This analysis could help FHEs to further understand the signatures and the writer behaviour. In addition, two information fusion schemes are proposed to combine the discriminative capability of the two types of features being considered. The coefficients in the wavelet decomposition of the different time functions associated with the signing process are used as features to model them. Two different signature styles are considered, namely, Western and Chinese, of one of the most recent publicly available Online Signature Databases. The experimental results are promising, especially for the features that seem to be relevant to FHEs, since the obtained verification error rates are comparable to the ones reported in the state-of-the-art over the same datasets. Further, the results also show that it is possible to combine both types of features to improve the verification performance.*

Reference: Marianela Parodi, Juan C. Gomez, Linda Alewijnse, Marcus Liwicki (2014). Online Signature Verification: Automatic Feature Selection vs. FHS's Choice. J. Forensic Document Examination, Vol. 24, pp. 5-19

Keywords: Online signature verification, FHE based features, Automatic feature selection, Information fusion

1. Introduction

Signature verification is one of the most popular methods for identity verification. It is a non-invasive biometric technique and people are familiar with the use of signatures in their everyday life. Automatic signature verification has long been considered an important research area in the field of biometrics [Plamondon and Lorette, 1989], [Leclerc and Plamondon, 1994], [Plamondon and Srihari, 2000], [Impedovo and Pirlo, 2008].

Two categories of signature verification systems can be distinguished taking into account the acquisition device, namely, offline and online systems. For offline systems, only the image of the signature is available. For online systems, dynamic information acquired during the signing process is available. In this case, the signature is parameterised by several discrete time functions such as x and y pen coordinates, pen pressure and, when available, pen inclination angles. Researchers have long argued about the effectiveness of the different time functions for verification purposes. There are conflicting results regarding their importance [Kholmatov and Yanikoglu, 2005], [Maramatsu and Matsumoto, 2007], [Houmani et al., 2009], and the discussion is still open.

The interest in the online approach has increased in recent years due to the widespread use of electronic pen-input devices. Nevertheless, there are certain applications that demand the use of the offline approach. Forensic Handwriting Experts (FHEs) only have the offline data available in their daily casework. To perform a forensic signature comparison it is then

1. FCEIA, Universidad Nacional de Rosario, and CIFASIS, ARGENTINA
2. Netherlands Forensic Institute,The Hague, THE NETHERLANDS
3. German Research Center for Artificial Intelligence, Kaiserslautern, GERMANY
{parodi,gomez}@cifasis-conicet.gov.ar, liwicki@dfki.uni-kl.de, l.alewijnse@nfi.minvenj.nl

necessary to work with offline data, while online data can be used to perform biometric person verification/identification. Furthermore, in the future it might occur that FHEs will also have to deal with online signatures.

To bridge the gap between the Pattern Recognition (PR) and the FHE communities is an important task, being crucial for the automatic signature verification approach to be useful to FHEs. It would then be interesting to investigate an online feature set containing features which are relevant to FHEs. This could help them to better understand the signatures and the writer behavior. FHEs work with the offline specimens of the signature, so it is not possible for them to look at online features. Nevertheless, they have plenty of experience in the interpretation of some dynamic information that can be inferred from the offline signature [Bird et al., 2011], [Caligiuri and Mohammed, 2012], [Will, 2012]. FHEs try to understand the forgery process from the forger's point of view. For a successful forgery, the forger must imitate all habits of the authentic signer and the qualities of the authentic signature, and must discard all conflicting elements of his own writing. There will be then a trade-off between accuracy and velocity. To produce an accurate copy of the specimen signature, the forger will likely write slowly, resulting in bad line fluency and hesitations that will be visible for the FHEs. On the other hand, if the forger focuses on the writing velocity to make the forgery more fluent he will aim at a variation that fits within the writer's variability. A monotonous pressure over all the signature, the presence of more pressure in unusual places or slightly different curves at specific spots, can also be signs of forgeries. Some results regarding the use of features motivated by FHEs are presented in [Pervouchine and Leedham, 2007] and [Santos et al., 2007].

Another factor that is important in the evaluation of signature verification systems is their versatility to deal with signatures from different cultural origins. In [Pal et al., 2012], an updated survey of non-English and non-Latin signature verification systems can be found. This interest in the analysis of non-English and non-Latin signatures is evidenced in the main conferences of the field, where signature verification competitions presenting non-English/non-Latin databases have been held. For instance, in ICDAR[i] 2009 [Blankers et al., 2009] and ICDAR 2011 [Liwicki et al., 2011] a Chinese database was made available, while in ICDAR 2013 [Malik et al., 2013] a Japanase database was introduced.

The idea in this paper is to show that an automatic signature verification system based only on the use of a small set of FHE based features could provide verification results comparable with the ones in the state-of-the-art, and to show that FHEs could benefit by combining automatically selected features and the usual features they employ. For this purpose, the discriminative power of a set of features relevant to FHEs is analysed and compared to the discriminative power of automatically selected feature sets. To perform the above mentioned combination, two different fusion strategies, namely, feature level fusion and decision level fusion, are employed. The coefficients in a wavelet approximation of the signature time functions are used as features to model them. A Random Forest (RF) classifier is employed to perform the verification [Breiman, 2001]. To analyse the influence of the cultural origin of the signatures in the performance of the proposed signature verification systems, two different signature styles, namely, Western and Chinese, of a recent publicly available database are used in the verification experiments.

The reader should note that a previous version of this paper has been published in [Parodi et al., 2013b]. In that work, the discriminative power of the feature set chosen by FHEs was compared to the one corresponding to the automatically selected features. In the present paper, the proposed signature verification approach is enhanced by combining the different feature sets being analysed on the basis of two different fusion techniques. Furthermore, more related work is reviewed, a clearer description of the proposed method is presented, and finally an improved analysis of the overall approach is performed. It is the authors' belief that the results presented in this paper could be of some benefit for both the Pattern Recognition and FHE communities.

The main contributions of this paper are the following:

- A set of features which seems to be relevant to FHEs is analysed. These features are selected based on those FHEs would look at when doing their daily casework.

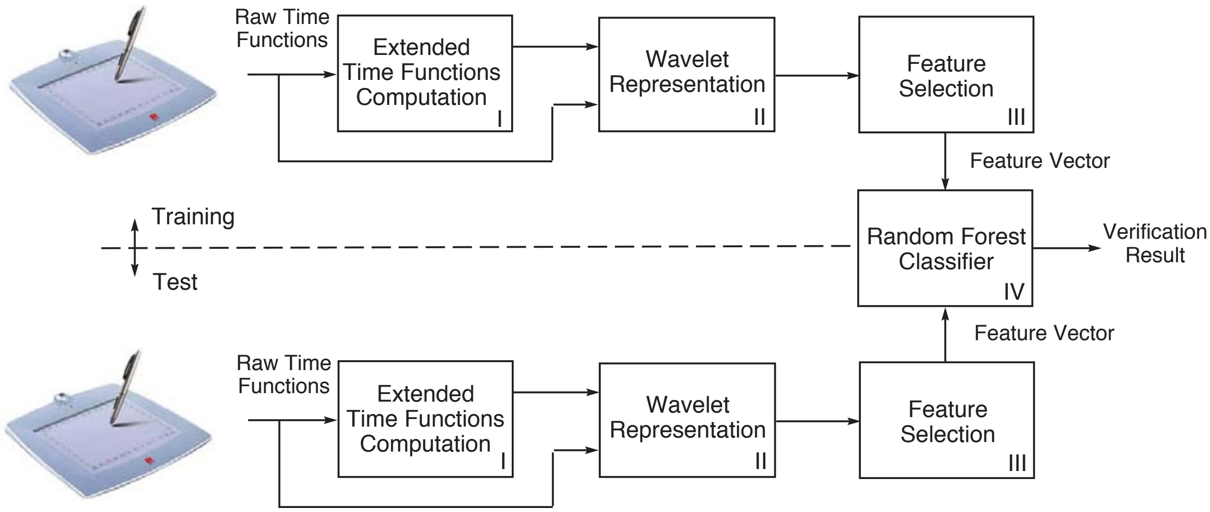

Figure 1: Online signature verification scheme.

- Different sets of automatically selected features are compared to the set of features which are meaningful for FHEs.
- Two different fusion techniques are proposed to combine the discriminative power of both types of features (FHE based and automatically selected ones).
- The experiments are performed on a signature database, containing Western and Chinese signatures, which has been used in one of the latest signature verification competitions.
- To quantify the verification performance, the EER (Equal Error Rate) and the cost of the log-likelihood ratios $\hat{C}_{llr}$ are reported. To compute log-likelihood ratios is important since they allow FHEs to give an opinion on the strength of the evidence.

The paper is organized as follows. In Section 2 a description of the proposed signature verification system is provided. The considered time functions are introduced in Section 3. The wavelet representation used to model the signatures is described in Section 4. Section 5 is devoted to the different feature combinations proposed in this paper. In particular, Subsection 5.1 focuses on the FHE based features, while Subsection 5.2 focuses on the automatically selected features. In Section 6 and Section 7 the database and the evaluation protocol are described, respectively. In Section 8 the experimental results are presented and discussed. In addition, the different fusion techniques are introduced and the verification results obtained with the combined systems are also presented and discussed. Finally, some concluding remarks are given in Section 9.

2. Online Signature Verification System Description

Figure 1 schematically depicts the proposed online signature verification system. The measured data (output of the acquisition device) consists of three discrete time functions: pen coordinates x and y, and pen pressure p. Depending on the acquisition device, the pen altitude and azimuth angles could also be available. In addition, several extended functions are usually computed from the raw data, for instance, velocity and acceleration (Block I). Each of the time functions (measured data and extended time functions) is then modelled based on its wavelet decomposition so that a time function representation is obtained which is composed by the wavelet approximation coefficients (Block II). Finally, the feature vectors are obtained by concatenating the wavelet representations of the different time functions. Different time function combinations will be considered in this paper (Block III). The verification is performed by a Random Forest classifier (Block IV).

3. Time Functions

The measured data consists of the three discrete time functions that are usually available, that is, pen coordinates x and y, and pen pressure p. As already mentioned, several extended functions can be computed from them [Kholmatov and Yanikoglu, 2005], [Richiardi and Drygajlo, 2003], [Fierrez-Aguilar et al., 2007]. In [Kholmatov and Yanikoglu, 2005], the incremental variations of the x and y pen coordinates are proposed. In [Richiardi and Drygajlo, 2003], several time functions, such as, the x and y velocities and accelerations and the log curvature radius, among others, are used as well as their first and second order time derivatives. 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 [Fierrez-Aguilar et al., 2007]. Prior to their computation, the original x and y pen coordinates are normalized regarding scale and translation. Let $n=1,2,...,L_{sign}$ be the discrete time index of the measured functions and L_{sign} the time duration of the signature in sampling units, then the above mentioned extended functions are computed as:

- **Path velocity magnitude:** $v_T(n)=\sqrt{\dot{x}^2(n)+\dot{y}^2(n)}$
- **Path-tangent angle:** $\theta(n)=\arctan(\dot{y}(n)/\dot{x}(n))$
- **Total acceleration:** $a_T(n)=\sqrt{\dot{v}_T{}^2(n)+c^2(n)}$ **, where:** $c(n)=v_T(n).\dot{\theta}(n)$
- **Log curvature radius:** $\rho(n)=\log(v_T(n)/\dot{\theta}(n))$

In all cases, the first order time derivatives are computed as in [Fierrez-Aguilar et al., 2007]:

$$\dot{f}(n)\approx\Delta f(n)=\frac{\sum_{\tau=1}^{2}\tau(f(n+\tau)-f(n-\tau))}{2.\sum_{\tau=1}^{2}\tau^2}$$

.

The initial set of features will then be composed by x, y and p, the computed extended functions (v_T, θ, a_T and ρ), and their first and second order time derivatives (dx, dy, dp, dv_T, $d\theta$, da_T, $d\rho$, and d^2x, d^2y, d^2p, d^2v_T, $d^2\theta$, d^2a_T, $d^2\rho$).

	$\ell=1$	$\ell=2$	$\ell=3$
Best FIT_x [%]	99.15	96.88	84.98
Best FIT_y [%]	98.47	89.64	75.23
Best FIT_p [%]	89.00	78.68	68.21

Table 1: Best FIT between the measured and the approximated time functions using db4 wavelets.

4. Wavelet Representation

In this paper, a fixed-length representation of the signatures based on the wavelet approximation of the different time functions associated with them is proposed. 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 perform a multilevel decomposition of the time functions associated with the signatures using the DWT and to use the approximation coefficients to represent them. This approach was introduced by the present authors in [Parodi et al., 2013a], where also Legendre polynomials series expansions were proposed to represent the time functions. A method where the detail coefficients together with the approximation ones are used to represent the time functions within the framework of online signature verification, was presented in [Chang et al., 2012].

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 with respect to the case of having different feature vector lengths. Several works in the literature have proposed a fixed-length representation of the signatures, see for instance [Yanikoglu and Kholmatov, 2009], where the Fast Fourier Transform is employed, and [Parodi et al., 2013a], where the present approach using wavelets was introduced. Further, using fixed-length feature vectors can be required for certain biometric applications [Tuyls et al., 2005], [Xu et al., 2008].

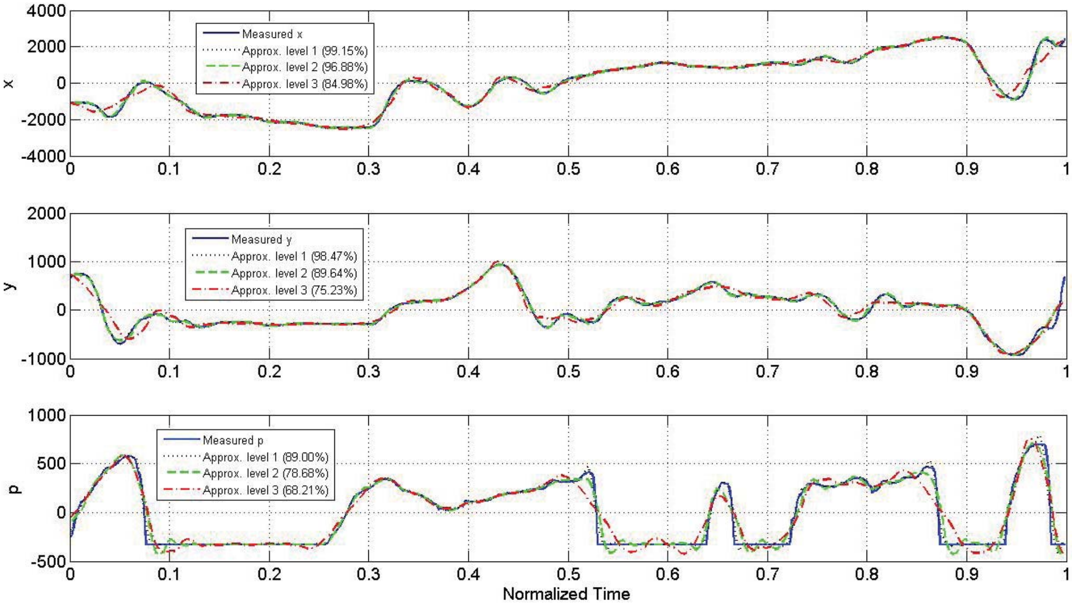

Figure 2: Measured time functions (red solid line): x (top), y (middle) and p (bottom) and their corresponding approximations by the DWT (db4) with levels of resolutions $\ell = 1$ (black dotted line), $\ell = 2$ (green dashed line) and $\ell = 3$ (red dash-dotted line). The best FIT for each case is indicated in brackets.

The approximation accuracy will be determined by the chosen resolution level, which also will determine 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 will then be the ength of the feature vector, which will determine the resolution level to be used. The widely used db4 wavelets [Daubechies, 1992] will be employed for the representation of the time functions. Figure 2 shows the time functions x, y and p associated with a signature and the corresponding first, second and third level approximation using wavelets.

As a measure of the accuracy of the approximation the Best FIT[ii] has been computed. In Table 1 the Best FIT for the different time functions x, y and p, and levels of resolution $\ell = 1, 2, 3$, are shown.

5. Feature Selection

In recent years, much effort has been devoted to try to incorporate the forensic handwriting expertise into the field of automatic signature verification. In particular, some works in the literature [Santos et al., 2007], [Pervouchine and Leedham, 2007] have proposed the use of features motivated by forensic handwriting examination. In this paper, features which are relevant to FHEs are analysed and their discriminative capability is compared with that of a set of automatically selected features. In addition, the possible combinations (at both feature level and decicion level) of FHE based features and automatically selected features is proposed in order to improve the verification performance.

5.1 FHE Based Features

As mentioned in Section 1, FHEs work with the static image of the signature, so it is not possible for them to look at online features, however, they can infer dynamic properties from the signature image, to some extent. They consider velocity and curvature as distinctive features. On the other hand, the acceleration and the pen position (which can be established by striae and inkless starts) are less useful for them. The pen pressure is not useful either since it is strongly dependant on external factors such as the writing material and surface, but pressure fluctuations could be interesting for them, since they are highly individualistic to the writer.

Dutch Dataset			Chinese Dataset		
Training Set			Training Set		
Authors	Genuines	Forgeries	Authors	Genuines	Forgeries
10	240	119	10	230	429
Testing Set			Testing Set		
Authors	Genuines	Forgeries	Authors	Genuines	Forgeries
54	1296	611	10	219	461

Table 2: Online Dutch (left) and Chinese (right) Datasets.

In this paper, the following FHE based online features are considered: velocity (v_T (magnitude) and θ (direction)), curvature (ρ) and first order time derivative of the pressure (dp). They were selected based on FHEs criteria for Latin scripts. It is likely that for FHEs who examine Chinese scripts, the criteria would be different. Since information about FHEs criteria for Chinese scripts was not available for the authors, the same FHE based feature set is used in this paper for both signature styles.

5.2 Automatically Selected Features

An automatic feature selection is proposed based on the variable importance provided by the Random Forest algorithm in order to compare the verification performance obtained when using the FHE based features with the one obtained when using automatically selected ones. As further described in Section 6, the used datasets are divided into the Training and Testing Sets. The feature selection is performed over the Training Sets for both datasets. The optimal automatically selected feature set consists of: x, a_T, y, v_T, p, dp, ρ, dx, θ, dy, d^2x, d^2y and dv_T, for the Dutch data, and of: 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.

6. Signature Database

The publicly available SigComp2011 Dataset [Liwicki et al., 2011] presented within ICDAR 2011 is used. This database consists of two separate datasets, one containing genuine and forged Western signatures (Dutch ones) and the other containing genuine and forged Chinese signatures. The available forgeries are skilled forgeries, which are simulated signatures in which forgers (different writers than the reference one) are allowed to practice the reference signature for as long as they deem it necessary. The signatures were acquired using a ballpoint pen on paper (WACOM Intuos3 A3 Wide USB Pen Tablet). This acquisition setup allows a normal writing behavior.

Each of the datasets in the SigComp2011 Database is divided into two sets, namely, the Training and Testing Sets. The Dutch (left) and the Chinese (right) datasets are described in Table 2.

The measured data in this database consists of the three discrete time functions: pen coordinates x and y, and pen pressure p. As already mentioned, the extended functions are computed as described in Section 3 from this measured data.

7. Evaluation Protocol

A RF classifier is used for the verification experiments. For each dataset, the optimization of the metaparameters of the system is performed over the corresponding Training Set while the corresponding Testing Set is used for independent testing purposes.

For the representation based on the proposed DWT approximations, the user has to choose the mother wavelet, the length of the resampled functions and the resolution level for the approximation. The length of

	Dutch Dataset			Chinese Dataset		
	EER	$\hat{C}_{llr}$	$\hat{C}_{llr}^{\min}$	EER	$\hat{C}_{llr}$	$\hat{C}_{llr}^{\min}$
FHE based feat.	9.59	0.3408	0.2966	10.27	0.3454	0.2760
Auto. Selec. Feat.	6.58	0.2426	0.2049	7.455	0.2962	0.2483
Auto. Selec. Feat. (only 4 feat.)	9.67	0.3365	0.2948	9.91	0.3948	0.3265
System	Acc.	$\hat{C}_{llr}$	$\hat{C}_{llr}^{\min}$	Acc.	$\hat{C}_{llr}$	$\hat{C}_{llr}^{\min}$
commercial	96.27	0.2589	0.1226	93.17	0.4134	0.2179
1st non-comm.	93.49	0.4928	0.2375	84.81	0.5651	0.3511

Table 3: Verification results for the considered feature sets, for the Dutch (left) and Chinese (right) Datasets.

the resulting feature vector is determined by the length of the resampled functions and the resolution level. Regarding the RF classifier, the parameters to adjust are the number of trees to grow and the number of randomly selected splitting variables to be considered at each node.

To obtain statistically significant results, a 5-fold cross-validation (5-fold CV) is performed over the Testing Set to estimate the errors. For each instance of the 5-fold CV, a signature model is trained for each writer, using only genuine signatures. To train the signature model for a particular writer, the genuine class consists of the genuine signatures of the writer available in the corresponding training set of the 5-fold CV, while the forged class consists of the genuine signatures of all the remaining writers in the dataset available in the same training set. The genuine signatures and the skilled forgeries of the writer under consideration available in the corresponding testing set of the 5-fold CV are used for testing.

To evaluate the performance, the EER is computed, 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) [Brümmer and du Preez, 2006]. The cost of the log-likelihood ratios $\hat{C}_{llr}$ and its minimal possible value $\hat{C}_{llr}^{\min}$ are computed using the toolkit as well. A smaller value of $\hat{C}_{llr}^{\min}$ indicates a better performance of the system. Using these measurements to evaluate the performance of a signature verification system was proposed in AFHA 2011 Workshop[iii], where the importance of computing the likelihood ratios was highlighted since they allow FHEs to give an opinion on the strength of the evidence [Gonzalez-Rodriguez et al., 2005], although they would not be in the position to make a leap of faith and judge about guilt or no guilt.

8. Results and Discussion

The verification performance is quantified by the EER, $\hat{C}_{llr}$ and $\hat{C}_{llr}^{\min}$ over the Dutch and Chinese Testing Sets. For the RF classifier, the number of trees was set to 500 and the number of randomly selected splitting variables to $\sqrt{P}$ (P being the feature vector dimension). The time functions were resampled resulting in a normalized length of 256. The resolution level was set to 3, in order to obtain a feature vector of a reasonable length (in this case 38 for each time function). Further increasing the resolution level would deteriorate the approximation accuracy, since only the approximation coefficients of the highest level are kept for the representation. A smaller resolution level would result in an extremely large feature vector. The verification results for the FHE based features and the automatically selected ones are shown in Table 3 for the Dutch (left) and Chinese (right) data, respectively. For the purposes of comparison, the verification results for the best commercial and non-commercial systems in the SigComp2011 competition are also included in the last two rows of Table 3.

From Table 3, it can be observed that the best verification results are obtained when using the automatically selected features, for both datasets. This means that the feature selection done by the RF algorithm is a meaningful one. Note that a different number of features are selected for each dataset, namely, 13 and 17 features for the Dutch and Chinese datasets, respectively. In the case of using the features based on the FHEs criterion, the verification performance is not as good as the one corresponding to the automatically selected features, but it is still a very good performance for both datasets. Note that in this case the number of features is limited to 4. This result is promising since these features have a meaningful interpretation by FHEs. This would suggest that, in case the verification system has to be limited to take into account only FHE based features, its performance would not be substantially deteriorated. In fact, if all the features that FHEs look at could be implemented (which is hard to do since some features used by FHEs, such as line quality and ink intensity variations, are not appropriately defined to be computed automatically), the performance might even be better. Moreover, taking into account other results in the state-of-the-art reported over the same datasets (last two rows of Table 3), it can be concluded that the performance of systems using only the FHE based features will still be comparable to the ones reported in the state-of-the-art.

The fact that the results obtained when using the automatically selected features outperform the ones obtained when using the FHEs based features is probably due to the fact that the automatic feature selection is not limited with respect to the size of the selected set. Nevertheless, the automatic selection is always keeping the FHE based features among the selected ones, that is, the FHE based features are considered important by the automatic feature selection algorithm. These features have been thoroughly investigated by FHEs and are generally accepted by the FHE community.

To perform a fairer comparison between FHE based features and automatically selected ones, the number of automatically selected features was reduced to equal the number of FHE based features (that is 4). The results using only the four most important features in the automatically selected feature set are included in the third row of Table 3. The subset of selected features are x, a_T, y and v_T for the Dutch and y, x, p and v_T for the Chinese datasets, respectively. It can then be observed that, when having a limitation in the set size, the automatically selected features do not coincide with the FHE based features (only v_T is kept) and the discriminative power of the set is deteriorated. This is probably due to the fact that the first features in the ranking are not necessarily the best features by themselves but they are good features when combined with other ones. This is not the case when using the FHE based features, since they are highly discriminative by themselves. Then, when having a limitation in the feature set size, the FHE based features would be more reliable and a better design option. In fact, they outperform (or equal) the results obtained by the reduced set of automatically selected features.

There are certain signatures that are wrongly classified when using the FHE based features, while they are correctly classified when using the automatically selected ones. Figure 3 shows the FHE based features, and the corresponding wavelet approximations, for a sample (shown in the bottom row) of a Dutch (left) and a Chinese (right) wrongly classified signatures, respectively. Note that the wavelet approximations are not so good due to the fact that the time functions are not smooth. Experiments (not included here) showed that incorporating the detail coefficients improves the approximation accuracy, at the cost of increasing the feature vector length. This would not be a limitation in the case of using the set of FHE based features since this set contains only four features. In any case, to pick this type of signature out of a database and to analyse their stability could be an interesting issue for future work. As will be shown later in this Section, the combination of FHE based features and automatically selected ones will improve the overall error rates.

Based on the results in Table 3, the question arises whether the discriminative capability of both FHE based features and automatically selected ones could be combined to improve the performance of the system. Traditionally, three main approaches for information fusion can be distinguished, namely, early or feature level fusion, intermediate or classifier level fusion, and late 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, which is typically encountered in applications where Hidden Markov Models (HMM) and Dynamic Bayesian Networks (DBN) are used to model the different signals involved, a composite classifier is generated by combining the individual classifiers used to process the different signals. Finally, in the late or decision level fusion approach, a final decision is obtained by combining the probability/likelihood scores from the separate classifiers processing the different signals. A good overview on information fusion techniques for the case of audiovisual signals, in applications of Human-Computer Interfaces can be found in [Shivappa et al., 2010]. References abound on the use of information fusion techniques in the field of biometrics, see for instance [Zhou et al., 2012] and [Lakshmanan, 2013] on ear, and [Kumar and Passi, 2010] on iris biometrics, respectively. Fusion information has also been used in the field of signature verification, see for instance [Rico-Juan and I˜nesta, 2012] and [Fierrez-Aguilar et al., 2005] for offline and online approaches, respectively. Figure 4 schematically depicts the three main approaches for information fusion. In this paper, feature level and decision level approaches will be considered to combine the discriminative capabilities of FHE based features and automatically selected features. In this case, classifier level fusion is not possible due to the particular classifier being used.

It is important to note that the different sets of automatically selected features considered in this paper contain different features for the Dutch and Chinese datasets. Then, it would be reasonable to expect that a verification system based on the fusion of the FHE based features and the automatically selected ones (at feature level or decision level), not only would combine the discriminative capabilities of both feature sets, but also could adapt better to the different signature styles (Dutch and Chinese).

Regarding a combination/fusion at feature level, it is clear that since the automatically selected feature set includes the FHE based features and also the reduced set of automatically selected features, the only combination that would make sense is the one between the FHE based features and the reduced set of automatically selected features. A feature vector consisting of the concatenation of the four FHE based features and the first three ranked features in the automatically selected feature set is considered in this case. Note that only the first three automatically selected features are being considered since the fourth feature is v_T, which is one of the four FHE based features. For the Dutch dataset these three features are: x, a_T and y, while for the Chinese dataset they are: y, x and p. The verification results for this case are shown in the third row of Table 4. It can be observed that for the Dutch dataset an improvement in the verification errors with respect to the ones obtained with each set of features separately (rows 1 and 2 of Table 4) is achieved with the combined feature vector. On the other hand, for the Chinese dataset only the results obtained with the reduced set of automatically selected features are improved with the combined feature vector.

Regarding a combination/fusion at decision level, independent classifiers are used for each of the feature sets to be combined and the final decision is computed as a combination of the likelihood scores associated with each classifier. In particular, the following combination/fusion rule is considered in this paper:

$$P_{fused} = P_{FHE}^{(1-\alpha)} . P_{ASF}^{\alpha} ,$$

where P_{fused} is the likelihood score for the combined scheme, P_{FHE} and P_{ASF} are the likelihood scores for the FHE based features and the automatically selected feature set, respectively, and $0 \leq \alpha \leq 1$ is a user defined parameter weighting the individual likelihood scores. In order to compare the results of this fusion approach with the ones obtained with the fusion at feature level, here also the combination is performed between the FHE based features and the reduced set of automatically selected features. Note that, also for this fusion approach, only the first three automatically selected features are used. In this way, v_T is only taken into account once. The value of α is optimized over the Training Set for each dataset. Figure 5 shows the $\hat{C}_{llr}^{\min}$ error as a function of α for the Dutch (left) and Chinese (right) datasets. From Fig. 5 the optimal values are $\alpha^{Dutch} = 0.7$ and $\alpha^{Chinese} = 0.44$ for the

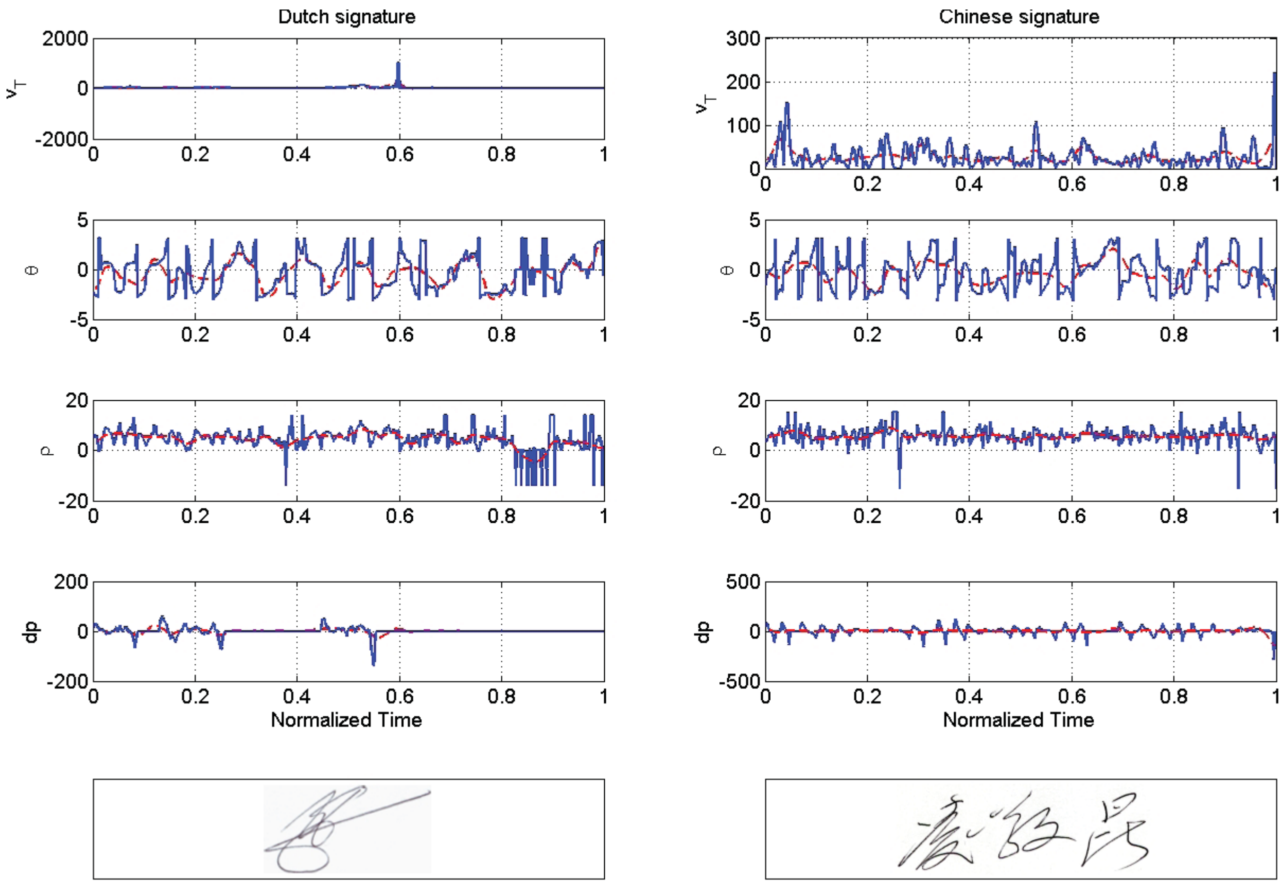

Figure 3: Wrongly classified signatures with FHE based features. Original v_T (first row), θ (second row), ρ (third row) and dp (fourth row) (blue solid line) and their corresponding approximations (red dashed line), for the Dutch (left) and Chinese (right) data. Bottom row: Associated signature images.

Ducth and the Chinese datasets, respectively. From Table 4 it can be observed that an improvement in the verification errors with respect to the ones obtained with each set of features separately (rows 1 and 2 of Table 4) is achieved with the fusion at decision level, for both datasets.

It has already been pointed out that for the Chinese data, the verification performance of the individual systems is only improved when fusion at the decision level is performed. The verification results obtained with the FHE based features are much better than the ones obtained with the reduced set of automatically selected features (see Table 4). Due to this notorious difference in the individual discriminative power, it is likely that the simple combination of these features in a unique feature vector (feature level fusion) would not have enough discriminative power to outperform the results obtained when using only the FHE based features. On the other hand, the fusion at the decision level does improve the verification results obtained with each set of features separately. Then, the use of weights to fuse the individual likelihood scores contributes to combine the individual discriminative power in a more efficient way. Note that the optimal value of α is $\alpha^{Chinese} = 0.44$, meaning that the likelihood scores obtained with the FHE based features (which are more discriminative than the reduced set of automatically selected features) have a higher weight in the combination. For the Dutch data, it has already been shown that the fusion of both feature sets reaches better verification results than the ones obtained with the individual systems, for both fusion approaches. The fact that both fusion approaches yield verification result improvements, is probably due to the fact that the individual verification performances are quite similar.

Note that the fusion at decision level (row 4 in Table 4) yields the best verification results for both

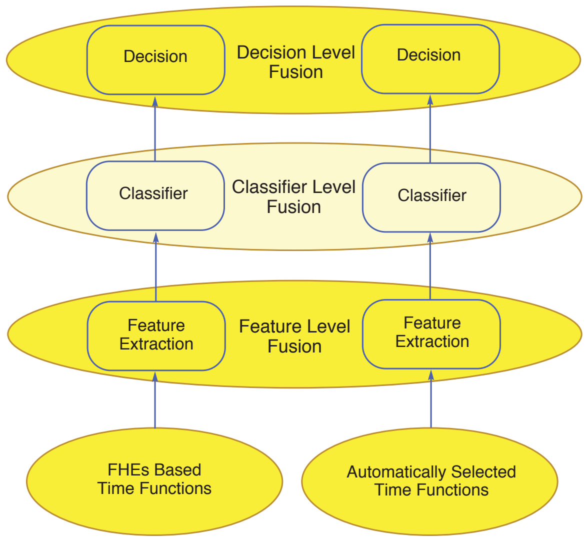

Figure 4: The three main approaches for information fusion.

datasets, and therefore this would be the fusion approach of choice for the considered features and datasets. In addition, the verification results obtained in this case are comparable to the ones in the state of the art (last two rows in Table 3). It is important to highlight that these very good verification results are obtained using only features that are meaningful to FHEs and some other features (the three most important from the automatically selected feature set) that are very simple and easily interpretable. This is promising since it shows that using only few features and, even more important, simple and easy to interpret ones (especially those relevant to the FHEs) good results can be achieved avoiding the use of too many features or more complex ones, which would mean more computational load and less interpretability of the whole system.

For the sake of completeness, the combination at decision level between the FHE based features and the whole set of automatically selected features has been also carried out. The results for this fusion are shown in the last row of Table 5. It can be observed that these results outperform the ones obtained with the FHE based features (row 1 in Table 5), but do not outperfom the ones obtained with the whole set of automatically selected features (row 2 in Table 5). Note that, for the Chinese data, the fusion results are almost equal to the ones corresponding to the whole set of automatically selected features, while for the Dutch data, the fusion results are worse than the ones corresponding to the whole set of automatically selected features. It is likely that the fusion results do not outperform the ones obtained using the whole set of automatically selected features because the FHE based features are contained in the whole set of automatically selected features. Based on these comments it can be concluded that this combination is not worth taking into account, since the results of the combination do not improve the best results obtained individually, and the resulting number of features is large and many of them are difficult to interpret.

The attentive reader has probably already noticed that none of the fusion strategies considered yield results outperforming the ones obtained using the whole set of automatically selected features. However, the results obtained with the decision level fusion of the FHE based features and the reduced set of

	Dutch Dataset			Chinese Dataset		
	EER	$\hat{C}_{llr}$	$\hat{C}_{llr}^{min}$	EER	$\hat{C}_{llr}$	$\hat{C}_{llr}^{min}$
FHE based feat.	9.59	0.3408	0.2966	10.27	0.3454	0.2760
Auto. Selec. Feat. (only 4 feat.)	9.67	0.3365	0.2948	9.91	0.3948	0.3265
Feature level fusion	7.93	0.316	0.2606	8.19	0.3349	0.2905
Decision level fusion	7.3	0.2649	0.2333	7.69	0.3146	0.2668

Table 4: Verification results for the combined systems (feature level and decision level fusion), for the Dutch (left) and Chinese (right) Datasets.

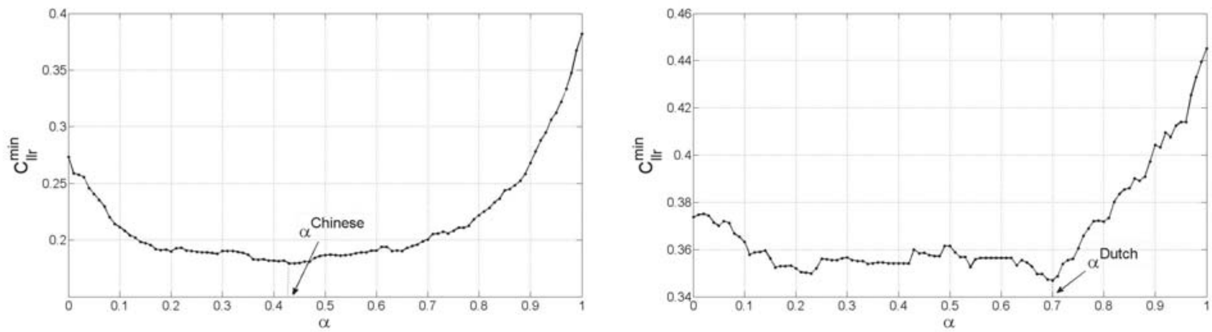

Figure 5: $\widehat{C}_{llr}^{\min}$ error as a function of α over the Training Set for the Dutch (left) and Chinese (right) datasets.

	Dutch Dataset			Chinese Dataset		
	EER	$\widehat{C}_{llr}$	$\widehat{C}_{llr}^{\min}$	EER	$\widehat{C}_{llr}$	$\widehat{C}_{llr}^{\min}$
FHE based feat.	9.59	0.3408	0.2966	10.27	0.3454	0.2760
Auto. Selec. Feat.	6.58	0.2426	0.2049	7.455	0.2962	0.2483
Decision level fusion	6.8	0.2452	0.2125	7.6	0.2926	0.2444

Table 5: Verification results for the decision level fusion between the FHE based features and the whole set of automatically selected features, for the Dutch (left) and Chinese (right) Datasets.

automatically selected features are not that far from the best results, so that by trading-off accuracy with system complexity, this fusion approach would be the best strategy.

9. Conclusions

An online signature verification system based on wavelet feature representation and a Random Forest classifier is proposed in this paper. The discriminative power of a set of features which are relevant to FHEs is analysed in this context. The performance of the proposed signature verification system using these FHE based features is evaluated and compared to the case of using automatically selected features. To take advantage of the discriminative power of both types of features two different fusion techniques, namely, feature level fusion and decision level fusion, are proposed to combine them. In addition, two different signature styles, namely, Western and Chinese, are considered to evaluate the verification performance.

The experimental results using both types of features (FHE based and automatically selected ones) are comparable with those of the state-of-the-art. The results using features which are relevant to FHEs are promising in the sense that the proposed approach could become an automatic verification tool which is useful and reliable for FHEs.

The experiments also show that the proposed decision level fusion approach is capable of improving the verification performance for both datasets (Western and Chinese), without necessarily increasing the system complexity. In addition, the proposed fusion has the advantage of resorting to features that are relevant for FHEs.

References

Bird et al., 2011 Bird, C., Found, B., Ballantyne, K., and Rogers, D. (2011). Dynamic features of naturally written, disguised and forged handwritten text. In *Grassi, E. and Contreras-Vidal, J., editors, Proceedings of 15th Conference of the International Graphonomics Society*, pages 181–184, Cancun, Mexico. College Park, MD: University of Maryland.

Blankers et al., 2009 Blankers, V., van den Heuvel, C., Franke, K., and Vuurpijl, L. (2009). The ICDAR 2009 signature verification competition. In *Proceedings of 10th International Conference on Document Analysis and Recognition*, pages 1403–1407, Barcelona, Spain.

Breiman, 2001 Breiman, L. (2001). *Random Forests*. Technical report, Statistical Department, University of California, Berkeley, USA.

Brümmer and du Preez, 2006 Brümmer, N. and du Preez, J. (2006). Application-independent evaluation of speaker detection. *Computer Speech & Language*, 20(2–3):230–275.

Caligiuri and Mohammed, 2012 Caligiuri, M. and Mohammed, L. (2012). *The Neuroscience of Handwriting*. CRC Press, Florida, USA.

Chang et al., 2012 Chang, H., Dai, D., Wang, P., Xu, Y., Si, F., and Huang, S. (2012). Online signature verification using Wavelet transform of feature function. *Journal of Informatics & Computater Science*, 9(11):3135–3142.

Daubechies, 1992 Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM: Society for Industrial and Applied Mathematics, Pennsylvania.

Fierrez-Aguilar et al., 2005 Fierrez-Aguilar, J., Nanni, L., Lopez-Peñalba, J., Ortega-Garcia, J., and Maltoni, D. (2005). An on-line signature verification system based on fusion of local and global information. In *Kanade, T., Jain, A., and Ratha, N., editors, Proceedings of the 5th International Conference on Audio- and Video-Based Biometric Person Authentication*, volume 3546 of LNCS, pages 523–532, Rye Brook, NY, USA.

Fierrez-Aguilar et al., 2007 Fierrez-Aguilar, J., Ortega-Garcia, J., Ramos-Castro, D., and Gonzalez-Rodriguez, J. (2007). HMM-based on-line signature verification: Feature extraction and signature modelling. *Pattern Recognition Letters*, 28(16):2325–2334.

Gonzalez-Rodriguez et al., 2005 Gonzalez-Rodriguez, J., Fierrez-Aguilar, J., Ramos-Castro, D., and Ortega-Garcia, J. (2005). Bayesian analysis of fingerprint, face and signature evidences with automatic biometric systems. *Forensic Science International*, 155(2–3):126–140.

Houmani et al., 2009 Houmani, N., Garcia-Salicetti, S., and Dorizzi, B. (2009). On assessing the robustness of pen coordinates, pen pressure and pen inclination to time variability with personal entropy. In *Proceedings of IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, pages 159–164, Washington DC, USA. IEEE Press Piscataway, NJ, USA.

Impedovo and Pirlo, 2008 Impedovo, D. and Pirlo, G. (2008). Automatic signature verification: The state of the art. *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, 38(5):609–635.

Kholmatov and Yanikoglu, 2005 Kholmatov, A. and Yanikoglu, B. (2005). Identity authentication using improved online signature verification method. *Pattern Recognition Letters*, 26(15):2400–2408.

Kumar and Passi, 2010 Kumar, A. and Passi, A. (2010). Comparison and combination of iris matchers for reliable personal authentication. *Pattern Recognition*, 43:1016–1026.

Lakshmanan, 2013 Lakshmanan, L. (2013). Efficient person authentication based on multi-level fusion of ear scores. *IET Biometrics*, 2(3):97–106.

Leclerc and Plamondon, 1994 Leclerc, F. and Plamondon, R. (1994). Automatic signature verification: The state of the art–1989-1993. *International Journal on Pattern Recognition and Artificial Intelligence*, 8(3):643–660.

Liwicki et al., 2011 Liwicki, M., Malik, M., den Heuvel, C., Chen, X., Berger, C., Stoel, R., Blumenstein, M., and Found, B. (2011). Signature verification competition for online and offline skilled forgeries (SigComp2011). In *Proceedings of 11th International Conference on Document Analysis and Recognition*, pages 1480–1484, Beijing, China.

Malik et al., 2013 Malik, M., Liwicki, M., Alewijnse, L., Ohyama, W., Blumenstein, M., and Found, B. (2013). ICDAR 2013 competitions on signature verification and writer identification for on- and offline skilled forgeries (SigWiComp 2013). In *Proceedings of 12th International Conference on Document Analysis and Recognition*, pages 1477–1483, Washington, DC, USA. IEEE Computer Society, Los Alamitos, California, USA.

Maramatsu and Matsumoto, 2007] Maramatsu,

D. and Matsumoto, T. (2007). Effectiveness of pen pressure, azimuth, and altitude features for online signature verification. In *Lee, S. and Li, S., editors, Proceedings of International Conference on Biometrics (ICB 2007)*, volume 4642 of LNCS, pages 503–512. Springer.

Pal et al., 2012] Pal, S., Blumenstein, M., and Pal, U. (2012). Hindi off-line signature verification. In *Proceedings of 13th International Conference on Frontiers in Handwriting Recognition*, pages 373–378, Bari, Italy. IEEE Computer Society, Washington, DC, USA.

Parodi et al., 2013a] Parodi, M., G´omez, J., Liwicki, M., and Alewijnse, L. (2013a). Orthogonal function representation for online signature verification: which features should be looked at? *IET Biometrics*, 2(4):137–150.

Parodi et al., 2013b] Parodi, M., G´omez, J., Liwicki, M., and Alewijnse, L. (2013b). Wavelet based representation for online signature verification. automatic feature selection vs. fhe's choice. In *Proceedings of 16th Conference of the International Graphonomics Society (IGS2013)*, pages 1–4, Nara, Japan.

Pervouchine and Leedham, 2007] Pervouchine, V. and Leedham, G. (2007). Extraction and analysis of forensic document examiner features used for writer identification. *Pattern Recognition*, 40(3):1004–1013.

Plamondon and Lorette, 1989] Plamondon, R. and Lorette, G. (1989). Automatic signature verification and writer identification the state of the art. *Pattern Recognition*, 22(2):107–131.

Plamondon and Srihari, 2000] Plamondon, R. and Srihari, S. (2000). On-line and off-line handwriting recognition: A comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):63–84.

Richiardi and Drygajlo, 2003] Richiardi, J. and Drygajlo, A. (2003). Gaussian mixture models for online signature verification. In *Proceedings of Workshop on Biometric Methods and Applications*, pages 115–122, Berkley, California, USA. ACM.

Rico-Juan and I˜nesta, 2012 Rico-Juan, J. and I˜nesta, J. (2012). Confidence voting method ensemble applied to off-line signature verification. *Pattern Analysis and Applications*, 15(2):113–120.

Santos et al., 2007 Santos, C., Bortolozzi, F., Oliveira, L., and Justino, E. (2007). Off-line signature verification based on forensic questioned document examination approach. In *Proceedings of Symposium on Applied Computing*, pages 637–638, Seoul, Korea. ACM, New York, NY, USA.

Shivappa et al., 2010 Shivappa, S., Trivedi, M., and Rao, B. (2010). Audiovisual information fusion in humancomputer interfaces and intelligent environments: A survey. *Proceedings of the IEEE*, 98(10):1692–1715.

Tuyls et al., 2005 Tuyls, P., Akkermans, A., Kevenaar, T., Schrijen, G., Bazen, A., and Veldhuis, R. (2005). Practical biometric authentication with template protection. In *Kanade, T., Jain, A., and Ratha, N., editors, Proceedings of the 5th International Conference on Audio- and Video-Based Biometric Person Authentication*, volume 3546 of LNCS, pages 436–446, Rye Brook, NY, USA.

Will, 2012 Will, E. (2012). Inferring relative speed of handwriting from the static trace. *Journal of Forensic Document Examination*, 22.

Xu et al., 2008 Xu, H., Veldhuis, R., Kevenaar, T., Akkermans, A., and Bazen, A. (2008). Spectral minutiae: A fixed-length representation of a minutiae set. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–6, Anchorage, AK.

B. Yanikoglu and A. Kholmatov, 2009 B. Yanikoglu and A. Kholmatov, Online Signature Verification Using Fourier Descriptors. *EURASIP Journal on Advances in Signal Processing 2009*, 2009:260516.

Zhou et al., 2012 Zhou, J., Cadavid, S., and Abdel-Mottaleb, M. (2012). An efficient 3-D ear recognition system employing local and holistic features. *IEEE Transactions on Information Forensics and Security*, 7(3):978–991.

(endnotes)

i ICDAR: International Conference on Document Analysis and Recogniton.

ii The Best FIT is defined as:

$$\text{Best FIT} = 100.\left(1-\frac{\left\|x-x_{aprox}\right\|}{\left\|x-x_{mean}\right\|}\right).$$

iii First International Workshop on Automated Forensic Handwriting Analysis (AFHA 2011) held within ICDAR 2011, Beijing, China