Online Signature Verification: To What Extent Should a Classifier be Trusted in?

Marianela Parodi

Lab. for System Dynamics & Signal Processing
Universidad Nacional de Rosario, CIFASIS, CONICET
Rosario, Argentina
parodi@cifasis-conicet.gov.ar

and

Juan C. Gómez

Lab. for System Dynamics & Signal Processing
Universidad Nacional de Rosario, CIFASIS, CONICET
Rosario, Argentina
gomez@cifasis-conicet.gov.ar

Abstract

To select the best features to model the signatures is one of the major challenges in the field of online signature verification. To combine different feature sets, selected by different criteria, is a useful technique to address this problem. In this line, the analysis of different features and their discriminative power has been researchers' main concern, paying less attention to the way in which the different kind of features are combined. Moreover, the fact that conflicting results may appear when several classifiers are being used, has rarely been taken into account. In this paper, a score level fusion scheme is proposed to combine three different and meaningful feature sets, *viz.*, an automatically selected feature set, a feature set relevant to Forensic Handwriting Experts (FHEs), and a global feature set. The score level fusion is performed within the framework of the Belief Function Theory (BFT), in order to address the problem of the conflicting results appearing when multiple classifiers are being used. Two different models, namely, the Denoeux and the Appriou models, are used to embed the problem within this framework, where the fusion is performed resorting to two well-known combination rules, namely, the Dempster-Shafer (DS) and the Proportional Conflict Redistribution (PCR5) one. In order to analyze the robustness of the proposed score level fusion approach, the combination is performed for the same verification system using two different classification techniques, namely, Random Forests (RF) and Support Vector Machines (SVM). Experimental results, on a publicly available database, show that the proposed score level fusion approach allows the system to have a very good trade-off between verification results and reliability.

Keywords: Online Signature Verification, Feature Combination, Score Level Fusion.

1 Introduction

Biometric systems aim to automatically recognize or verify an identity. Among the numerous available biometric techniques, signature verification is one of the most popular [1–5]. The widespread acceptance of the signatures is due to the fact that they are recognized as a legal means of verifying an individual's identity by financial and administrative institutions, and that people are familiar with the use of signatures for identity verification in their everyday life. In addition, signatures have the advantage of being easy to collect, since no invasive methods are required to collect them.

According to the signatures' acquisition method, signature verification systems can be categorized into

offline and online systems. For offline systems, only the image of the signature is available, while for online systems, dynamic information acquired during the signing process, such as x and y pen coordinates, pen pressure, and even pen angles, is available. Online verification systems are believed to be more reliable than offline ones, since it would be reasonable to expect that the incorporation of dynamic information acquired during the signing process would make signatures more difficult to forge. In addition, although there exists some applications that still demand the use of the offline approach, such as Forensic Handwriting Experts (FHEs) caseworks, the use of different electronic pen-input devices, such as digitizer tablets and PDAs, has significantly increased in recent years, making online systems the current option for signature verification.

Which features should be used to model the signatures has always been a crucial issue in the field of online signature verification. Although the effectiveness of the different features for verification purposes has been largely discussed, the conflicting results in the literature make the discussion still open [6–11]. Despite the lack of conclusions, researchers agree that an interesting strategy is to combine different feature sets, selected by different criteria, in order to take advantage of their individual discriminative capabilities. In line with this idea, different online signature verification systems have been proposed in the literature combining different kind of features. Some of them, propose the combination of local features, which are computed for each point in the time sequences, and global features, which are computed from the whole signatures, and highlight the advantages of incorporating global features to the verification systems [11–14]. Some others, propose the combination of features (independently of being local or global) obtained on the basis of different feature extraction strategies, showing that the combination improves the performance with respect to the case of using each of the feature sets individually [15–18].

Despite most of the works in the literature of signature verification concerning feature combination focus their analysis on which features should be combined, it is also important to address the problem of how to combine them. Information fusion techniques have long been used in the field of biometrics in order to improve the systems' accuracy by considering multiple pieces of evidence [19–22]. The combination of different authentication systems can offer substantial improvements in their accuracy depending upon the information being combined, for instance, the different kind of features, and the fusion methodology adopted. In particular, information fusion can be performed at different levels, *viz.*, data level, feature level, score level, or decision level. The fusion at data or feature level, usually referred to as tightly coupled integrations [23], assumes a strong interaction among the input measurements. On the other hand, the fusion at score or decision level, usually called loosely coupled systems, assumes a very little or no interaction among the inputs, since the integration occurs at the output of relatively autonomous agents, each agent independently assessing the input from its own perspective. Several works in the literature of biometrics have shown that the fusion at score level offers the best trade-off between information content and the ease in fusion [24]. Researchers in the area of signature verification agree with this idea, being score level fusion the trend in this field [13], [15–18], [25–28]. Nevertheless, examples using feature level fusion [18], [29, 30], and decision level fusion [31], can also be found.

The most commonly used technique in the literature of signature verification to perform score level fusion consists in combining the scores of the individual classifiers with simple rules, such as, *max*, *min*, *product*, *simple sum*, or *weighted sum* [20]. In [13], local and global features are combined on the basis of the *sum* and *max* rules, while in [17], such kind of features are combined on the basis of the *sum* and *product* rules. A user dependent weighting fusion is introduced in [25] for combining features in the Discrete Wavelet Transform (DWT) domain, and a *weighted geometrical* fusion rule is used in [18] for combining three different feature sets. To train a fusion algorithm using kernel methods, such as, Support Vector Machines (SVM) [15], [32], or Neural Networks (NN) [33], is also a widely used technique. In [15], a combination of two different histogram features is performed on the basis of a SVM classifier, while in [32], the SVM classifier is used to combine offline and online features. In [33], a two-stage NN is used to perform the combination of global and local features.

When combining scores obtained by multiple classifiers, conflicting results may appear, in the sense that the different classifiers may not achieve similar scores, introducing some degree of uncertainty to the classification problem. Given such a situation, it would be desirable to quantify to what extent each classifier should be trusted in. In fact, to use this information for performing the combination would allow it to be more realistic and reliable. This can be done by computing a confidence level for each classifier and embedding the classification problem in a framework capable of handling the problem's inherent degree of uncertainty. Although being widely used, score level fusion approaches based on simple rules and kernel techniques do not perform well under this condition. Moreover, in such cases, the fusion performance decreases considerably. A suitable framework to address the problem of conflicting information is provided by the Belief Function Theory (BFT) [34, 35]. The BFT is a general framework for reasoning with uncertainty, first introduced by A. P. Dempster in the context of statistical inference [34], and developed later by G. Shafer into a general framework for modeling epistemic uncertainty [35]. The BFT has been widely used for multi-

biometric applications [24], [36–39], where conflicting results are the common scenario. Although combining multiple classifiers is a widely used technique in the field of signature verification, to the best of the authors’ knowledge, the degree of uncertainty that this technique may introduce to the classification problem has not been significantly explored. In fact, there exist few works in the literature taking into account such situation resorting to the BFT [27, 28]. In [27], the authors propose the fusion of online and offline information using SVM classifiers, and in [28], the fusion of SVM classifiers is proposed to combine different feature sets for offline verification.

In this paper, the combination of three different and meaningful feature sets, selected by different criteria, is proposed on the basis of a score level fusion approach based on the BFT, which would provide the appropriate framework to quantify the confidence in each of the classifiers, and to handle the conflicting results that may appear. In particular, an automatically selected feature set [10], a set of features relevant to Forensic Handwriting Experts (FHEs) [40], and a set of global features [11], are combined. These feature sets have already shown to have interesting discriminative capabilities, resulting not only in good verification performances, but also providing, from their own perspective, several advantages to the signature verification system. In a previous work [18], these feature sets were combined on the basis of a simpler combination approach, where the possibility of conflicting results among the different classifiers was not taken into account. Here, by actually incorporating this possibility to the analysis (through the BFT approach), it is expected for the proposed combination approach to be more reliable and, in this way, to make the whole signature verification system more secure. The idea is then to train an independent classifier using each of the three mentioned feature sets, and combine the corresponding three output scores on the basis of a score level fusion approach based on the BFT. The first step towards managing conflicting information, is to transform the information (classifiers’ output scores) into evidence, in order to embed the problem within the BFT framework. This constitutes a crucial step in the whole fusion process. In this paper, two different transfer models, namely, the Denoeux [41] and Appriou [42] models, are proposed to perform the transformation. In addition, the proposed models allow to explicitly take into account the confidence level corresponding to each classifier, by incorporating a confidence factor when converting the classifiers’ scores into belief assignments. The second step, is to actually perform the combination in the framework of the BFT, resorting to combination rules specially developed (most of them based on Dempster’s rule [43]) to work within this context. In this paper, two different and widely used rules, namely, the Dempster-Shafer (DS) [44] and the Proportional Conflict Redistribution (PCR5) one [45], are employed. Finally, to evaluate the robustness of the proposed score level fusion approach, two different and well-known classifiers, namely, Random Forests (RF) [46] and SVM [47, 48], are used to build two different signature verification systems where the score level fusion approach is performed. That is, in one case, the three independent classifiers are implemented based on RF, and in the other, they are implemented based on SVM. In addition, the verification performance of the proposed score level fusion approach is analyzed for two different signature styles, namely, Western (Dutch) and Chinese, in a publicly available database [49].

The paper is organized as follows. The basics of the BFT are described in Section 2. Section 3 describes the feature sets that are to be combined. The proposed score level fusion scheme is described in Section 4. In Sections 5 and 6, the database and the evaluation protocol are described, respectively. In Section 7, the experimental results are presented and discussed. Finally, some concluding remarks are given in Section 8.

2 Basics of BFT

The BFT is a general framework for reasoning with uncertain information. It arose from the reinterpretation and development of the work of A. P. Dempster [34] by G. Shafer [35], and it is a generalization of the Bayesian theory of subjective probability. The BFT allows to combine evidence from different sources and arrive at a degree of belief (or confidence, or trust), which is represented by a mathematical object called belief function, that takes into account all the available evidence. To do so, the idea is to obtain degrees of belief for one question from subjective probabilities for a related question, and combine such degrees of belief based on specially developed combination rules, such as, Dempster’s rule or any other rule derived from it.

Generally, signature verification leads to a two-class classification problem, in the sense that an input signature can be either genuine or a forgery. For this reason, the main definitions regarding BFT are presented in the following Subsections for the case of a two-class problem.

2.1 Belief Assignments

To embed the problem within the framework of BFT, it is necessary to convert the scores of the classifiers, which are the ones that are meant to be combined, into belief assignments. In this paper, the Denoeux [41] and Appriou [42] transfer models are used to perform this transformation. Such models allow, in addition, the

possibility of actually measuring the degree of confidence in each classifier by letting to explicitly introduce the confidence factor α associated with each classifier.

Let θ_{gen} and θ_{false} be associated with the genuine and the forged classes, respectively, being $\Theta = \{\theta_{gen}, \theta_{false}\}$ the so-called frame of discernment. The BFT takes into account the possibility of having ambiguity in the results by considering also the case where the evidence is not strong enough to discern between the two classes, assuming the power set of Θ (2^Θ) to be the (extended) set of possible results $\{\theta_{gen}, \theta_{false}, \{\theta_{gen}, \theta_{false}\}\}$. In this context, the Denoeux [41] model is given by

$$\begin{aligned} m_i(gen) &= \alpha_i \times \Psi_i(s_i) \\ m_i(false) &= 1 - \alpha_i \times \Psi_i(s_i) \\ m_i(\Theta) &= 0, \end{aligned} \quad (1)$$

while the Appriou [42] one is given by

$$\begin{aligned} m_i(gen) &= \alpha_i \times \Psi_i(s_i) \\ m_i(false) &= \alpha_i \times (1 - \Psi_i(s_i)) \\ m_i(\Theta) &= 1 - \alpha_i, \end{aligned} \quad (2)$$

where m are the belief assignments, i corresponds to the classifier, $0 < \alpha_i < 1$ is a confidence factor for each classifier, s_i is the score of the classifier i for a given signature, and Ψ_i is a monotonically increasing function, which maps the scores into the interval $[0, 1]$. In the particular signature verification application proposed in this paper, there is no need of using the mapping function, since the scores are already in the interval $[0, 1]$ due to the classification methods being used.

As already mentioned, the proposed transfer models allow the possibility of actually measuring the degree of confidence in each classifier by letting to explicitly introduce the confidence factor α associated with each classifier. This is an important step in the whole proposed score level fusion scheme, since it is a step towards answering the question in the paper's title. In addition, each of the proposed models perform the conversion and incorporates the confidence factor α resorting to different criteria, making it possible to analyze the impact of using one model or another in the whole score level fusion process.

Finally, from Equation (1), the reader can notice that Denoeux model does not actually take into account the case of ambiguity in the classifiers' results since $m_i(\Theta) = 0$. This is not the case for the Appriou model (Equation (2)), where $m_i(\Theta) = 1 - \alpha_i$. In any case, that is, whether the Denoeux or Appriou model is used, the conflicting results are taken into account at the combination stage by using the DS and PCR5 combination rules described in Subsection 2.2, which explicitly handle uncertainty (see Equations (3) and (4)). Then, even in the case of using the Denoeux model, where the conflicting results are not considered in the transformation stage, they are handled in the combination stage.

2.2 Combination Rules

To perform the combination in the framework of the BFT, that is, to combine the previously calculated belief assignments, it is necessary to resort to combination rules specially developed to work in this context. In this paper, the DS [44] and PCR5 rules [45], are used.

Let $m_1(\cdot)$ and $m_2(\cdot)$ be the belief assignments corresponding to two different classifiers (for instance, computed as in Equations (1) or (2)). The DS combination rule [44] is defined as

$$m_{DS}(A) = \frac{\sum_{(X,Y \in 2^\Theta, X \cap Y = A)} m_1(X).m_2(Y)}{1 - \sum_{(X,Y \in 2^\Theta, X \cap Y = \emptyset)} m_1(X).m_2(Y)}, \quad (3)$$

where $A \in 2^\Theta = \{\theta_{gen}, \theta_{false}, \{\theta_{gen}, \theta_{false}\}\}$, and the denominator is a normalizing factor which quantifies the conflict between $m_1(X)$ and $m_2(Y)$. On the other hand, the PCR5 rule [45] is defined as

$$m_{PCR5}(A) = m_{12}(A) + \sum_{(A,B \in 2^\Theta, A \cap B = \emptyset)} \left[\frac{m_1(A)^2.m_2(B)}{m_1(A) + m_2(B)} + \frac{m_2(A)^2.m_1(B)}{m_2(A) + m_1(B)} \right], \quad (4)$$

where $m_{12}(A) = \sum_{(X,Y \in 2^\Theta, X \cap Y = A)} m_1(X).m_2(Y)$ corresponds to the conjunctive consensus on A between the classifiers.

2.3 Pignistic Transformation

To convert the belief functions resulting from the combination of the belief assignments (for instance, computed as in Equations (3) or (4)) into probability measures, the pignistic transformation [50] is usually

employed. It is defined as

$$P(\theta_j) = \sum_{\theta_j \in A \subseteq \Theta} \frac{|A \cap \theta_j|}{|A|} m(A), \quad (5)$$

where $|A|$ denotes the cardinality of A , and $j = \{gen, false\}$. The classification result is then based on probabilities computed as in Equation (5).

3 Feature Sets

Three different feature sets are considered. A set composed by automatically selected features (ASF), a set of features relevant to the FHEs (FHEF), and a set of global features (GF), described in Subsections 3.1.1, 3.1.2 and 3.2, respectively.

Typically, the acquired data during the signing process consists of three discrete time functions: pen coordinates (x and y), and pen pressure p . Several extended functions can be computed from them [12], [51]. In this paper, the path-velocity (magnitude v_T and direction θ), the total acceleration a_T , and the log-radius curvature ρ , are computed from the x and y pen coordinates and the pen pressure p . Before their computation, the original x and y pen coordinates are normalized regarding scale and translation. The initial set of time functions, on which the selection of the three different feature sets will be performed, is composed by the x and y pen coordinates, the pen pressure p , the above mentioned extended functions, viz, v_T , θ , a_T and ρ , and their first and second order time derivatives.

It is important to note that, before performing the selection of the ASF and FHEF features, each of the time functions listed above is represented on the basis of the feature extraction technique described in Subsection 3.1. On the other hand, the GF features are computed directly from the original time functions, without resorting to the technique described in Subsection 3.1.

3.1 Time Function based Features

Each of the time functions listed above is decomposed by a Discrete Wavelet Transform (DWT) at different resolution levels, splitting them into low (approximation) and high (details) frequency components. The idea is to use the corresponding DWT approximation coefficients to represent the time functions.

To actually perform the DWT decomposition, the mother wavelet and the resolution level have to be chosen. In addition, previous to the DWT decomposition, the time functions have to be resampled in order to have a fixed-length feature vector. The approximation accuracy is determined by the chosen resolution level, which also determines the length of the resulting feature vector. This means a trade-off between accuracy and feature vector length, since this length has to be kept reasonably small. Then, the design parameter is, together with the mother wavelet, the length of the feature vector. The described feature extraction approach has been first introduced in [10], where experiments using different mother wavelets and resolution levels have been carried out. The best results have been obtained using the widely known **db4** wavelets [52] and a resolution level of 3, that are also the mother wavelet and resolution level used here.

3.1.1 Automatically Selected Features

The ASF features are selected from the initial set of time functions listed above, based on the variable importance given by the RF algorithm. They were first introduced in [10], achieving very good verification results at the cost of being a large and complex feature set.

The feature selection is performed over a set of signatures exclusively reserved for training purposes, that is, the Training Set (see Table 1 in Section 5), for each dataset in the database, namely, Dutch and Chinese. The corresponding automatically selected feature sets are composed by:

- Dutch dataset: $x, a_T, y, v_T, p, dp, \rho, dx, \theta, dy, d^2x, d^2y, dv_T$,
- Chinese dataset: $y, x, p, v_T, a_T, dy, dx, d^2y, \theta, \rho, dp, d^2x, d\theta, d^2p, dv_T, dp, d^2\theta$,

where df and d^2f denote the first and second order time derivatives of the corresponding time function f , respectively.

Looking at the selected features for the Dutch and Chinese datasets, it can be noticed 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.

3.1.2 Features Relevant to FHEs

The FHEF features, first introduced in [10], are fully interpretable by FHEs, which is an important characteristic, since these features allow automatic signature verification systems to be integrated into toolkits useful for FHEs. This is a fundamental step towards bridging the gap between the Pattern Recognition (PR) and the FHE communities, which is one of the most important current challenges in the field of signature verification. In [14], it was shown that, despite being quite simple, these features allow to build an automatic signature verification system based exclusively on them, obtaining competitive verification results.

In general, FHEs work with the image of the signature, so it is not possible for them to look at online features. However, they can infer some dynamic properties from the signature image. When FHEs analyze a signature, they consider the velocity and the curvature as distinctive features, since they are hard to copy. On the other hand, the acceleration and the pen position (it can be established by striae and inkless starts), are less useful for them. They do not take into account the local pressure for the analysis because it tends to be easily influenced by external factors such as surface and writing material. However, they do take into account pressure fluctuations. The set of features relevant to FHEs is then composed by:

- velocity (v_T (magnitude) and θ (direction)),
- log-radius curvature (ρ),
- first order time derivative of the pressure (dp).

3.2 Global based Features

Despite being less accurate than local ones, global features are simple and intuitive, and have the advantage of being easy to compute and compare. In [11], the discriminative power of this type of features has been exploited by incorporating a pre-classification step to an existing signature verification system, leading to improvements not only in the verification results but also in its simplicity and speed.

Different global features can be extracted from the measured and extended time functions. In [12] and [13], several global features were ranked on the basis of two different feature selection strategies. In [11], in order to successfully pre-classify the signatures, some of the better ranked features in [12] and [13] were selected. The set of global features proposed in [11] and used in this paper, is composed by:

- 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),
- time at which the pressure is maximum (T_{P_M}).

4 Score Level Fusion Scheme

Each of the three feature sets described in Section 3 is used to train an independent classifier. The output of each of these three independent classifiers are the scores that will be combined. Since the idea here is to evaluate the advantages of the proposed score level fusion approach, the independent classifiers are built on the basis of two different classifiers, one based on RF [46], and the other based on SVM [47, 48]. Then, the proposed combination is performed, in one case, by fusing the three outputs corresponding to three RF classifiers trained by each of the three different feature sets, and in the other case, by fusing the three outputs corresponding to each of the three SVM classifiers trained by each of the three different feature sets. In this way, it is expected to show that the proposed score level fusion approach improves the systems' performance independently of the classification method being used.

The proposed combination rules (DS and PCR5), are designed to perform the combination between two sources of information. Then, to perform the combination of the scores corresponding to each of the three classifiers trained by each of the three different feature sets, it is necessary to resort to a two-step cascade scheme. That is, on a first step, a combination between the scores of two independent classifiers trained by two of the proposed feature sets is performed. Then, on a subsequent step, the outcome of this combination is combined with the scores of the third classifier trained by the remaining feature set. The details of the proposed score level fusion scheme can be observed in Figures 1 and 2, for the case of using the RF and

SVM classifiers, respectively. Of course, the performance of the proposed two-step cascade scheme (shown in Figures 1 and 2), will depend on its implementation. That is, the performance will depend on which feature sets are selected to train the two independent classifiers involved in the first step combination, and which feature set is left to be combined at the second step. Then, this constitutes an important tuning decision, and the way in which the two-cascade scheme is implemented has to be optimized. This is done, over the corresponding set of signatures reserved exclusively for training purposes (Training Set described in Table 1, Section 5), together with the optimization of the other tuning parameters of the verification system.

Figure 1: Proposed signature verification system based on a two step cascade score level fusion scheme, when using the RF classifier.

Figure 2: Proposed signature verification system based on a two step cascade score level fusion scheme, when using the SVM classifier.

The proposed score level fusion is performed as follows. The combination at the first step of the cascade, is performed between the scores corresponding to two independent classifiers (in one case RF, and in the other, SVM) trained by two of the proposed feature sets. A value of α is computed for each of these two classifiers. Then, the scores corresponding to each of these classifiers are converted into belief assignments resorting to Equations (1) or (2). Once the belief assignments are computed, the fusion is performed on the basis of the proposed combination rules resorting to Equations (3) or (4). Then, the belief functions associated to this combination are converted into probabilities resorting to Equation (5). The combination at the second step of the cascade, is performed between the scores corresponding to the output of the combination performed in the previous first step and the ones corresponding to the third independent classifier, trained by the remaining of the proposed feature sets. A new value of α , associated to the output of the combination performed at the first step, is computed, and so it is a value of α associated to the third independent classifier (RF or SVM correspondingly). Then, the scores are converted into belief assignments resorting to Equations (1) or (2), and the fusion is performed on the basis of the proposed combination rules resorting to Equations (3) or (4). The output of this combination, performed at the second step of the cascade, is the final output of the system. This output is a belief function itself, and indicates the degree of belief for the proposed combination. Finally, to make a decision about an input signature, it is necessary to work with probabilities rather than belief functions. Then, to arrive at the desired output of the system, that is, to be able to decide whether the input signature corresponds to the claimed identity (is genuine) or not (is a forgery), it is mandatory to convert the belief functions into probabilities resorting to Equation (5), and as a subsequent stage, apply some decision rule to these probabilities. Then, a simple decision rule defined as

$$Decision = \begin{cases} \text{genuine}, & \frac{P(\theta_{gen})}{P(\theta_{false})} \geq t \\ \text{false}, & \text{otherwise.} \end{cases} \quad (6)$$

can be applied to actually decide whether the input signature is genuine or a forgery. Usually, the threshold t can be defined by the user depending on the particular application and the required level of security. It

is a common practice to choose t as the one that minimizes the error measurement being used to evaluate the performance of the signature verification system. Of course, the optimization of t should be performed over the corresponding set of signatures reserved exclusively for training purposes (Training Set described in Table 1, Section 5).

5 Signature Database

The publicly available SigComp2011 Dataset [49] is used to carry out the verification experiments. It has two separate datasets, containing Western (Dutch) and Chinese signatures, respectively. Each dataset is divided into independent Training and a Testing Sets. Table 1, shows a detailed description of the Dutch (left) and Chinese (right) datasets¹. Skilled forgeries, which are simulated signatures 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 a natural writing process. The measured data are the pen coordinates x and y , and the pen pressure p .

Table 1: Online Dutch (left) and Chinese (right) datasets.

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

6 Evaluation Protocol

The Equal Error Rate (EER) has been widely used to evaluate the automatic signature verification systems' performance. In this paper, the EER is computed, using the Bosaris toolkit, from the Detection Error TradeOff (DET) Curve as the point in the curve where the False Rejection Rate (FRR) equals the False Acceptance Rate (FAR) [54]. In recent years, several efforts have been made in order to bridge the gap between the PR and FHE communities. In line with this idea, to use the cost of the log-likelihood ratios to evaluate the performance of automatic signature verification systems was introduced in AFHA 2011 Workshop² (ICDAR 2011), where the importance of computing such measurements was highlighted, since they allow FHEs to give an opinion on the strength of the evidence [13], although they would not be in the position to make a leap of faith and judge about guilt or no guilt. In this paper, the cost of the log-likelihood ratios $\hat{C}_{llr}$ and its minimal possible value $\hat{C}_{llr}^{min}$ are computed using the above mentioned toolkit. A smaller value of $\hat{C}_{llr}^{min}$ indicates a better performance of the system.

The optimization of the tuning parameters of the proposed verification systems is performed, for each dataset in the SigComp2011 Dataset, over the corresponding set of signatures reserved exclusively for training purposes, that is, the Training Set (Table 1 in Section 5), while the Testing Set (Table 1 in Section 5) 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. Either in the case of using RF or SVM classifiers, the 5-fold CV is carried out as follows. The first step, is to train each of the three independent classifiers, using each of the three feature sets described in Section 3. For each of these three independent classifiers, the training procedure is the same. For each instance of the 5-fold CV, a signature model is trained for each writer, using only genuine signatures. It is important to note that forgeries are not usually available in real applications during the training phase, then, avoiding their use at this stage, makes the developed system more realistic. Then, to train the signature model for a particular writer, the genuine class consists in the genuine signatures of the writer available in the corresponding training set of the 5-fold CV, while the forged class consists in the genuine signatures of all the remaining writers in the dataset available in the same training set. Once the training phase is completed, the second step is to compute the scores corresponding to each of the three independent classifiers. The genuine and forged signatures of the writer under consideration available in the corresponding testing set of the 5-fold CV are used for testing the

¹The amount of genuine and forged signature samples may differ from those in [49] since when making signatures available for the research community some of them were missing [53].

²AFHA 2011: 1st Int. Workshop on Automated Forensic Handwriting Analysis

model corresponding to each of the three independent classifiers that have already been trained. As a result of this testing procedure, the three scores corresponding to the three classifiers' outputs are available for the current testing set of the 5-fold CV. The third step, is then to combine these three independent scores resorting to the score level fusion described in Section 4. In this way, the final output of the system (the combined output) is obtained for the current instance of the 5-fold CV.

7 Results and Discussion

As already mentioned in Section 6, the tuning parameters of the system are optimized over the corresponding set of signatures exclusively reserved to training purposes (Training Set in Table 1, Section 5), for the Dutch and Chinese datasets. In particular, for the proposed score level fusion scheme, it is necessary to optimize the order in which the feature sets are combined, that is, the implementation of the two-step cascade scheme, and the confidence factor α associated with each classifier's output involved in each combination step. In addition, the tuning parameters corresponding to both of the classification techniques that are being used, namely, RF and SVM, have also to be optimized.

The proposed two-step cascade scheme (Figures 1 and 2) can be performed in different ways depending on which feature sets are selected to train the two independent classifiers involved in the first step combination, and which feature set is left to be combined at the second step. Experiments taking into account all the possible alternatives have been carried out, over the corresponding Training Sets, for the Dutch and Chinese datasets. The best option, that is, the one that minimizes the verification errors in these Training Sets, is the one that performs the combination between ASF and FHEF features at the first step, and leaves the GF features to perform the combination at the second step, for both, RF and SVM classifiers, and for the Dutch and Chinese datasets. Then, this is the chosen way for the implementation of the two-step cascade scheme for computing the verification errors over the Testing Set. Note that, although the ASF and the FHEF feature sets may contain features in common (it will depend on the algorithm used to perform the automatic feature selection), they are selected by two explicitly different criteria. In fact, the FHEF features are meaningful for FHEs, and they intend to represent some of the features they look at in their daily work. Then, it makes sense to combine these feature sets because they represent information from two very different sources. That is, the ASF features could be provided by a PR researcher, since automatic feature selection is a widely used technique to choose features within researchers in the PR community, while the FHEF could be provided by a FHE researcher, allowing a fruitful contribution between both communities.

The value of the confidence factor α , plays a crucial role in the whole score level fusion process, and so, it has to be carefully chosen. In this paper, α is selected by performing an exhaustive search in the parameter space. The optimized values of α are shown in Tables 2 and 3, for the Dutch and Chinese datasets, respectively.

Table 2: Confidence factor α for the Dutch dataset.

	Dutch Dataset							
	Denoeux				Appriou			
	DS		PCR5		DS		PCR5	
	RF	SVM	RF	SVM	RF	SVM	RF	SVM
α_{ASF}	0.9	0.9	0.3	0.9	0.9	0.7	0.9	0.9
α_{FHEF}	0.5	0.1	0.2	0.1	0.1	0.4	0.1	0.1
α_{Comb}	0.4	0.9	0.8	0.9	0.9	0.2	0.9	0.2
α_{GF}	0.1	0.1	0.2	0.3	0.1	0.1	0.1	0.1

Table 3: Confidence factor α for the Chinese dataset.

	Chinese Dataset							
	Denoeux				Appriou			
	DS		PCR5		DS		PCR5	
	RF	SVM	RF	SVM	RF	SVM	RF	SVM
α_{ASF}	0.9	0.9	0.9	0.9	0.9	0.5	0.8	0.5
α_{FHEF}	0.1	0.1	0.1	0.1	0.3	0.1	0.4	0.1
α_{Comb}	0.9	0.9	0.9	0.9	0.8	0.5	0.9	0.5
α_{GF}	0.1	0.1	0.1	0.1	0.3	0.1	0.1	0.1

Regarding the tuning parameters of the classifiers, for the RF classifiers, the number of trees to grow and the number of randomly selected splitting variables to be considered at each node, have to be chosen. It is well known that, in general, the sensitivity to those parameters is not meaningful [55], and the default values are a good choice. Then, the number of trees and the number of randomly selected splitting variables are set to 500 and $\sqrt{P}$ (being P the feature vector dimension), respectively, for all the verification experiments. For the SVM classifiers, it is necessary to choose the kernel and its tuning parameters. The Radial Basis Function (RBF) kernel is used, being the scale σ^2 and the regularization parameter $C > 0$, which provides a trade-off between model complexity and the training error in the SVM cost function, the corresponding tuning parameters. These values, optimized over the corresponding Training Sets, are $\sigma^2 = 10^{-7}$ and $C = 10$, for all the verification experiments.

The verification results are shown, in terms of the EER and the $\hat{C}_{llr}^{min}$, in Tables 4 and 5 for the Dutch and Chinese datasets, respectively. These results are obtained resorting to a score level fusion scheme performed as the two-step cascade shown in Figures 1 and 2 implemented as described above, the optimized values of α shown in Tables 2 and 3, and the above mentioned optimized parameters for the RF and SVM classifiers, respectively. Tables 4 and 5 include results obtained in the case of using the Denoux as well as the Appriou transfer models, and using the DS combination rule as well as the PCR5 one, for the case of using the RF and the SVM classifiers.

Table 4: Verification results for the Dutch dataset.

	Dutch Dataset							
	Denoux Model				Appriou Model			
	DS		PCR5		DS		PCR5	
	RF	SVM	RF	SVM	RF	SVM	RF	SVM
EER	6.14	4.72	6.24	6.48	7.21	6.19	6.92	6.60
$\hat{C}_{llr}^{min}$	0.1981	0.1430	0.1700	0.1967	0.1914	0.1800	0.1897	0.1875

Table 5: Verification results for the Chinese dataset.

	Chinese Dataset							
	Denoux Model				Appriou Model			
	DS		PCR5		DS		PCR5	
	RF	SVM	RF	SVM	RF	SVM	RF	SVM
EER	7.00	11.05	6.26	11.50	6.51	15.78	6.48	15.78
$\hat{C}_{llr}^{min}$	0.1777	0.3236	0.1637	0.3446	0.1682	0.4385	0.1623	0.4376

When a combination of different feature sets is proposed, such as it is here, the main objective is that the verification performance of the combined system outperforms the corresponding to the ones obtained when using the feature sets individually. In order to compare, Table 6 summarizes the best combination results from Tables 4 and 5 (highlighted in **boldface** in Tables 4 and 5) obtained for each classification method (RF and SVM), together with the results obtained in the case of using each feature set individually, for the Dutch (left) and Chinese (right) datasets, respectively. From Table 6, it can be observed that the best verification results obtained by the proposed score level fusion scheme outperform the ones obtained in the case of using each feature set individually, for both, RF and SVM classifiers, and for the Dutch and Chinese datasets. This shows that the proposed feature combination is meaningful, since achieves the very first goal of any feature combination, that is to improve the verification performance with respect to the case of using each feature set individually. Note that, although only the best results obtained with the proposed combinations are shown in Table 6, all the results in Tables 4 and 5 (except for the ones corresponding to the case of using the Appriou model together with the SVM classifier for Chinese signatures that do not outperform the results obtained when using the ASF features individually) are better than the ones obtained when using the different feature sets individually.

Two different classification methods, namely, RF and SVM, were used in order to evaluate the robustness of the proposed score level fusion approach against the use of different classification methods to implement the classifiers that are meant to be combined. For the proposed score level fusion scheme to be useful, it is reasonable to expect it to be robust against the classification method being used to implement each independent classifier. The verification results in Table 6, show that the proposed score level fusion scheme improves the verification performance independently of using RF or SVM classifiers. This is an important fact, since the idea here is to combine different meaningful feature sets, selected by different criteria, and to show

that combining them properly can enhance their individual discriminative power, making the combination to achieve better verification results, without being influenced by the chosen classification method. Then, although the classifier makes its own contribution to the overall performance of the verification system, the feature combination has to be useful by itself, that is, for any classification method being used.

Table 6: Best combination result vs. individual results for Dutch (left) and Chinese (right) datasets.

	Dutch Dataset				Chinese Dataset			
	RF		SVM		RF		SVM	
	EER	C_{llr}^{min}	EER	C_{llr}^{min}	EER	C_{llr}^{min}	EER	C_{llr}^{min}
Best ASF-FHEF-GF	6.24	0.1700	4.72	0.1430	6.48	0.1623	11.05	0.3236
ASF	6.58	0.205	6.14	0.2227	7.45	0.248	11.90	0.3775
FHEF	8.88	0.2885	8.19	0.2631	8.37	0.2602	16.22	0.4509
GF	12.31	0.4168	9.70	0.3215	14.40	0.4137	32.47	0.7986

As discussed above, the proposed score level fusion scheme does improve the verification performance with respect to the ones corresponding to the case of using the feature sets individually. Nevertheless, this is not the only goal for the proposed score level fusion approach. In fact, the idea here is to make the score level fusion process more robust by taking into account the possibility of conflicting results appearing when multiple classifiers are being used. To do so, the score level fusion is performed within the framework of the BFT. From Tables 4 and 5, an analysis of the use of the BFT to perform the proposed score level fusion can be done. When using the Appriou model, the conversion of the classifiers' scores into belief assignments considers the case of ambiguity in the classifiers' results (see Equation (2)). In this case, very similar results are obtained with both combination rules (DS and PCR5). This stands for both datasets (Dutch and Chinese) and, even more interesting, for both classification techniques, that is, for RF and SVM classifiers. On the other hand, when using the Denoeux model, the case of ambiguity in the classifiers' results is not considered when converting the classifiers' scores into belief assignments (see Equation (1)). In this case, there is an appreciable difference between the results achieved when using each combination rule (DS or PCR5). Moreover, whether using the DS or the PCR5 combination rule yields better results, seems to be influenced by the classification method (RF or SVM) being used. In fact, from Tables 4 and 5 it can be observed that, in the case of using the Denoeux model, the best results are obtained when using the DS combination rule for Dutch signatures, and when using the PCR5 combination rule for Chinese signatures. It would then be reasonable to infer that, despite the fact that, as already mentioned above, the proposed score level fusion approach does improve the verification performance independently of the classification method being used, the system becomes more influenced by the chosen classification method, and thereby, weaker, when the possibility of ambiguity in the classifiers' results is not considered by the transfer model. In other words, it seems that to take into account the possibility of ambiguity in the classifiers' results at the first stage of the process, that is, when embedding the problem within the BFT framework, makes the whole score level fusion scheme more robust. Regarding the BFT approach, the best results are obtained when using the DS combination rule and the Denoeux model for the Dutch data, and the PCR5 combination rule and the Appriou model for the Chinese data. This shows that, to use the Appriou model has not only the advantage of making the system more robust, but also of yielding improvements on the verification results in the case of the Chinese data. Unfortunately, to improve the robustness of the system has a relatively bad impact on the verification results for the Dutch data.

Finally, the best verification results for the proposed score level fusion approach are summarized in Table 7, for the Dutch (left) and Chinese (right) datasets. These results correspond to the ones obtained when using the Denoeux model together with the DS combination rule and the SVM classifier, in the case of the Dutch data, and to the ones obtained when using the Appriou model together with the PCR5 combination rule and the RF classifier, in the case of the Chinese data. In addition of the above analysis, this explicitly shows the difference between the two signatures styles (Dutch and Chinese), since the score level fusion scheme that obtains the best verification results, although the same in structure, is completely different from the implementation point of view in each case.

The idea of the paper is to combine three different and meaningful feature sets, paying special attention to the way in which the combination is performed. A particular score level fusion approach is proposed, where confidence factors are introduced to quantify the confidence in each classifier, and the conflicting results among the classifiers are handled resorting to the BFT. At this point, it is important to evaluate whether the performance of the proposed combination approach fulfils the expectations. In order to compare, results in previous works, where other combinations among ASF, FHEF and GF features have been proposed, but the conflicting results among the classifiers have not been taken into account, are included in Table 7. In

particular, the results in [11], where a pre-classification scheme is used to combine GF and ASF features, the results in [14], where FHEF features are combined with a subset of the GF features relevant to the FHEs, and the results in [18], where a plain score level fusion of ASF, FHEF and GF features is performed, are included.

Table 7: Best results for Dutch (left) and Chinese (right) dataset.

	Dutch Dataset		Chinese Dataset	
	EER	C_{lr}^{min}	EER	C_{lr}^{min}
Best ASF-FHEF-GF	4.72	0.1430	6.48	0.1623
GF-ASF [11]	3.09	0.112	4.99	0.17
GF-FFHE [14]	4.42	0.178	5.98	0.211
ASF-FHEF-GF [18]	6.95	0.228	7.31	0.218
1st comm.	-	0.123	-	0.237
1st non-comm.	-	0.237	-	0.351

From Table 7, it can be seen that, the verification results obtained with the proposed score level fusion approach outperform the ones obtained in [14] and [18], where other combinations between FHEF and GF, and among ASF, FHEF and GF features, respectively, are proposed. This stands for the Dutch and Chinese data. In the case of the Chinese data, the verification results obtained with the proposed score level fusion approach also outperforms the ones corresponding to the combination of ASF and GF features proposed in [11]. Unfortunately, this is not the case for the Dutch data, since the verification results obtained with the proposed score level fusion approach are not as good as in [11]. As already mentioned, the combinations proposed in [11], [14] and [18] do not take into account the possibility of conflicting results appearing when different classifiers are being combined. To actually taking this possibility into account, yields more conservative results. This could be the reason for the results corresponding to the Dutch data being not as good as the ones in [11]. Nevertheless, even for the case of the Dutch data, the verification results obtained with the proposed score level fusion approach are good results taking into account the trade-off between accuracy and reliability.

Finally, in order to compare the obtained verification results not only with the ones obtained when combining similar feature sets, but also with the best ones reported in the state-of-the-art over the same database, the best results in [49] are also included in the last two rows of Table 7. From Table 7, it can be observed that the verification results obtained with the proposed score level fusion approach are among the best ones in the state-of-the-art. This is particularly notorious in the case of the Chinese data. In fact, for this data, the obtained results are specially worth highlighting, outperforming not only the ones in [11], [14] and [18], but also the ones corresponding to the first ranked commercial and non-commercial systems in [49]. These are very promising results, since Chinese data is one of the most challenging data in the literature of online signature verification [49]. In the case of the Dutch data, the verification results obtained with the proposed combination approach are not as good as in [11], but they are still competitive within the ones in the state-of-the-art, outperforming not only the ones in [14] and [18], but also the first ranked non-commercial system in [49].

8 Conclusions

In this paper, a feature combination for online signature verification is proposed on the basis of a score level fusion scheme. In particular, automatically selected features, features relevant to FHEs and global features, are combined. The idea of the paper is not only to combine three different and meaningful feature sets, but also to pay special attention to the way in which the combination is performed. It is widely known that, when combining scores from different classifiers, conflicting results may appear. To address this problem, the proposed score level fusion is performed within the framework of the BFT. To work within this context, allows to handle the uncertainty in the classifiers' results, and allows the possibility of actually measuring the degree of confidence in each classifier, by letting to explicitly associate a confidence factor with each classifier.

To actually perform the proposed score level fusion within the BFT framework, two different transfer models, namely, the Denoeux and Appriou models, are used to embed the problem within the framework, while two different combination rules, namely, the DS and PCR5 ones, are employed to combine the evidence in such a context.

The major challenge of the paper is to perform a score level fusion that actually takes into account the conflicting results appearing when several classifiers are being used, while keeping a good verification

performance. The experimental results show that this objective has successfully been achieved. The proposed score level fusion approach not only has the advantage of being more reliable and secure by taking into account the possibility of conflicting results (specially in the case of using the Appriou model), but also achieves competitive verification performances (which are robust, in the sense that they do not depend on the chosen classification method), outperforming results obtained when using other proposed combination schemes, and being among the best ones in the state-of-the-art.

Finally, in the case of the Chinese data, the obtained verification results are not only comparable but even better than the best ones in the state-of-the-art. In fact, special attention should be paid for the results obtained for this data, since they are particularly promising.

References

- [1] R. Plamondon and G. Lorette, "Automatic Signature Verification and Writer Identification: The State of the Art," *Pattern Recognition*, vol. 22, no. 2, pp. 107–131, 1989. [Online]. Available: [https://doi.org/10.1016/0031-3203\(89\)90059-9](https://doi.org/10.1016/0031-3203(89)90059-9)
- [2] F. Leclerc and R. Plamondon, "Automatic Signature Verification: The State of the Art-1989-1993," *Int. J. on Pattern Recognition and Artificial Intelligence*, vol. 8, no. 3, pp. 643–660, June 1994.
- [3] R. Plamondon and S. Srihari, "Online and Offline Handwriting Recognition: A Comprehensive Survey," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 63–84, Jan. 2000. [Online]. Available: <http://dx.doi.org/10.1109/34.824821>
- [4] D. Impedovo and G. Pirlo, "Automatic Signature Verification: The State of the Art," *IEEE Trans. on Systems, Man, and Cybernetics. Part C: Applications and Reviews*, vol. 38, no. 5, pp. 609–635, Sept. 2008. [Online]. Available: <http://dx.doi.org/10.1109/TSMCC.2008.923866>
- [5] D. Impedovo, G. Pirlo, and R. Plamondon, "Handwritten Signature Verification: New Advancements and Open Issues," in *Proc. of 13th Int. Conf. on Frontiers in Handwriting Recognition*, Bari, Italy, Sept. 2012, pp. 367–372. [Online]. Available: <http://dx.doi.org/10.1109/ICFHR.2012.211>
- [6] D. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, and G. Rigoll, "SVC2004: First International Signature Verification Competition," in *Int. Conf. on Biometric Authentication*, ser. LNCS, D. Zhang and A. Jain, Eds., vol. 3072. Springer, Berlin, Heidelberg, 2004, pp. 16–22. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-25948-0_3
- [7] A. Kholmatov and B. Yanikoglu, "Identity Authentication Using Improved Online Signature Verification Method," *Pattern Recognition Letters*, vol. 26, no. 15, pp. 2400–2408, 2005. [Online]. Available: <https://doi.org/10.1016/j.patrec.2005.04.017>
- [8] D. Maramatsu and T. Matsumoto, "Effectiveness of Pen Pressure, Azimuth, and Altitude Features for Online Signature Verification," in *Advances in Biometrics (ICB 2007)*, ser. LNCS, vol. 4642. Springer, Berlin, Heidelberg, 2007, pp. 503–512. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-74549-5_53
- [9] N. Houmani, S. Garcia-Salicetti, and B. Dorizzi, "On Assessing the Robustness of Pen Coordinates, Pen Pressure and Pen Inclination to Time Variability with Personal Entropy," in *IEEE 3rd Int. Conf. on Biometrics: Theory, Applications, and Systems*, Sept. 2009. [Online]. Available: <http://dx.doi.org/10.1109/BTAS.2009.5339074>
- [10] M. Parodi, J. Gómez, M. Liwicki, and L. Alewijnse, "Orthogonal Function Representation for Online Signature Verification: Which Features Should be Looked at?" *IET Biometrics*, vol. 2, no. 4, pp. 137–150, Dec. 2013. [Online]. Available: <http://dx.doi.org/10.1049/iet-bmt.2013.0025>
- [11] M. Parodi and J. Gómez, *Emerging Aspects in Handwritten Signature Verification*. World Scientific, 2014, ch. Using Global Features for Pre-classification in Online Signature Verification Systems, pp. 39–52.
- [12] J. Richiardi, H. Ketabdar, and A. Drygajlo, "Local and Global Feature Selection for On-line Signature Verification," in *Proc. of 8th Int. Conf. on Document Analysis and Recognition*, Seoul, South Korea, Sept. 2005, pp. 625–629. [Online]. Available: <http://dx.doi.org/10.1109/ICDAR.2005.152>

- [13] J. Fierrez-Aguilar, L. Nanni, J. Lopez-Peñalba, J. Ortega-Garcia, and D. Maltoni, "An On-Line Signature Verification System Based on Fusion of Local and Global Information," in *Proc. of IAPR Int. Conf. on Audio- and Video-Based Biometrics Person Authentication (AVBPA 2005)*, ser. LNCS, T. Kanade, A. Jain, and N. Ratha, Eds., vol. 3546. Springer, 2005, pp. 523–532. [Online]. Available: http://dx.doi.org/10.1007/11527923_54
- [14] M. Parodi, J. Gómez, and L. Alewijnse, "Automatic Online Signature Verification Based Only on FHE Features: An Oxymoron?" in *Proc. of 14th Int. Conf. on Frontiers in Handwriting Recognition*, Crete Island, Greece, Sept. 2014, pp. 73–78. [Online]. Available: <http://dx.doi.org/10.1109/ICFHR.2014.20>
- [15] M. Berkay Yilmaz, B. Yanikoglu, C. Tirkaz, and A. Kholmatov, "Offline Signature Verification Using Classifier Combination of HOG and LBP Features," in *Proc. of 2011 Int. Joint Conf. on Biometrics (IJCB)*, Washington DC, USA, 2011, pp. 1–7.
- [16] D. Yang, Y. Qin, Z. Huang, and Y. Lu, *Signal Processing, Image Processing and Pattern Recognition*, ser. Communications in Computer and Information Science. Springer, Berlin, Heidelberg, 2011, vol. 260, ch. Off-line Signature Verification Based on Combination of Modified Direction and Microstructure Features, pp. 270–279. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-27183-0_29
- [17] J. Rico-Juan and J. Iñesta, "Confidence Voting Method Ensemble Applied to Off-line Signature Verification," *Pattern Analysis and Applications*, vol. 15, no. 2, pp. 113–120, April 2012. [Online]. Available: <http://dx.doi.org/10.1007/s10044-012-0270-1>
- [18] M. Parodi and J. Gómez, "Online Signature Verification: Is the Whole Greater Than the Sum of the Parts?" in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, ser. LNCS, A. Pardo and J. Kittler, Eds., vol. 9423. Springer, Oct. 2015, pp. 1–9. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-25751-8_19
- [19] L. Hong, A. Jain, and S. Pankanti, "Can Multibiometrics Improve Performance?" in *Proc. of AutoID99*, ser. LNCS, T. Kanade, A. Jain, and N. Ratha, Eds., vol. 3546. Springer, Berlin, 1999, pp. 59–64.
- [20] A. Ross, K. Nandakumar, and A. Jain, *Handbook of Multibiometrics*, ser. Biometrics. Springer, US, 2006, vol. 6, no. 1. [Online]. Available: <http://dx.doi.org/10.1007/0-387-33123-9>
- [21] M. Gudavalli, S. Raju, A. Babu, and D. Kumar, "Multimodal Biometrics - Sources, Architecture and Fusion Techniques: An Overview," in *Proc. of Int. Symposium on Biometrics and Security Technologies (ISBAST)*, Taipei, Taiwan, March 2012, pp. 27–34. [Online]. Available: <http://dx.doi.org/10.1109/ISBAST.2012.24>
- [22] S. Soruba Sree and N. Radha, "A Survey on Fusion Techniques for Multimodal Biometric Identification," *Int. J. of Innovative Research in Computer and Communication Engineering*, vol. 2, no. 12.
- [23] J. Clark and A. Yuille, *Data Fusion for Sensory Information Processing Systems*, ser. Engineering and Computer Science. Springer, 1990, vol. 105. [Online]. Available: <http://dx.doi.org/10.1007/978-1-4757-2076-1>
- [24] L. Mezai and F. Hachouf, "Score-Level Fusion of Face and Voice Using Particle Swarm Optimization and Belief Functions," *IEEE Trans. on Human-Machine Systems*, vol. 45, no. 6, pp. 761–772, June 2015. [Online]. Available: <http://dx.doi.org/10.1109/THMS.2015.2438005>
- [25] I. Nakanishi, H. Sakamoto, Y. Itoh, and Y. Fukui, "Optimal User Weighting Fusion in DWT Domain On-Line Signature Verification," in *AVBPA 2005*, ser. LNCS, vol. 3546. Springer, Berlin, 2005, pp. 758–766. [Online]. Available: http://dx.doi.org/10.1007/11527923_79
- [26] S. Koike, I. Nakanishi, and Y. Itoh, "Score Level Fusion in DWT Domain On-line Signature Verification," in *Proc. of 23th Int. Technical Conf. on Circuits Systems, Computers and Communication*, 2008, pp. 669–672.
- [27] N. Abbas and Y. Chibani, "Combination of Off-Line and On-Line Signature Verification Systems Based on SVM and DST," in *Proc. of Int. Conf. on Intelligent Systems Design and Applications (ISDA)*, Cordoba, Spain, Nov. 2011, pp. 22–24. [Online]. Available: <http://dx.doi.org/10.1109/ISDA.2011.6121764>
- [28] —, "SVM-DSmT Combination for Off-Line Signature Verification," in *Proc. of Int. Conf. on Computer, Information and Telecommunication Systems (CITS)*, Amman, Jordan, May 2012, pp. 1–5. [Online]. Available: <http://dx.doi.org/10.1109/CITS.2012.6220365>

- [29] D. Shashi Kumar, R. Ravi Kumar, K. Raja, R. Chhotaray, and S. Pattanaik, *Information and Communication Technologies*, ser. Communications in Computer and Information Science. Springer, Berlin, Heidelberg, 2010, vol. 101, ch. Combined Off-Line Signature Verification Using Neural Networks, pp. 580–583. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-15766-0_99
- [30] K. Tselios, E. Zois, A. Nassiopoulos, and G. Economou, “Fusion of Directional Transitional Features for Offline Signature Verification,” in *Proc. of Int. Joint Conf. on Biometrics (IJCB)*, Oct. 2011. [Online]. Available: <http://dx.doi.org/10.1109/IJCB.2011.6117515>
- [31] C. Allgrove and M. Fairhurst, “Majority Voting for Improved Signature Verification,” in *Proc. of IEE Colloquium on Visual Biometrics*, London, UK, March 2000, pp. 1–4. [Online]. Available: <http://dx.doi.org/10.1049/ic:20000469>
- [32] V. Mottl, M. Lange, V. Sulimova, and A. Yermakov, “Signature Verification Based on Fusion of On-line and Off-line Kernels,” in *Proc. of 19th Int. conference on Pattern Recognition (ICPR 2008)*, Tampa, FL, USA, Dec. 2008. [Online]. Available: <http://dx.doi.org/10.1109/ICPR.2008.4761208>
- [33] H. Baltzakisa and N. Papamarkosb, “A New Signature Verification Technique Based on a Two-stage Neural Network Classifier,” *Engineering Applications of Artificial Intelligence*, vol. 14, no. 1, pp. 95–103, Feb. 2001. [Online]. Available: [https://doi.org/10.1016/S0952-1976\(00\)00064-6](https://doi.org/10.1016/S0952-1976(00)00064-6)
- [34] A. Dempster, “Upper and Lower Probabilities Induced by a Multivalued Mapping,” *The Annals of Mathematical Statistics*, vol. 38, no. 2, pp. 325–339, April 1967.
- [35] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, Sept. 1976.
- [36] D. Kisku, J. Sing, and P. Gupta, “Multibiometrics Belief Fusion,” in *Proc. of 2nd Int. Conf. on Machine Vision*, 2009, pp. 37–40. [Online]. Available: <http://doi.ieeecomputersociety.org/10.1109/ICMV.2009.63>
- [37] B. Xiao, “Multi-biometric Feature Fusion Identification Based on D-S Evident Theory of Image Quality with Different Weights,” *J. of Computer Applications*, vol. 32, no. 1, pp. 264–268, 2012.
- [38] F. Wang, *Advanced Materials Research*, 2012, vol. 459, ch. Fusion Fingerprint and Face Multi-Biometrics Recognition Based on D-S Evidence Theory, pp. 580–583. [Online]. Available: 10.4028/www.scientific.net/AMR.459.644
- [39] K. Nguyen, S. Denman, S. Sridharan, and C. Fookes, “Score-level Multibiometric Fusion Based on Dempster-Shafer Theory Incorporating Uncertainty Factors,” *IEEE Trans. Human-Machine Systems*, vol. 45, no. 1, pp. 132–140, Feb. 2015. [Online]. Available: <http://dx.doi.org/10.1109/THMS.2014.2361437>
- [40] M. Parodi, J. Gómez, L. Alewijnse, and M. Liwicki, “Online Signature Verification: Automatic Feature Selection vs. FHE’s Choice,” *J. of Forensic Document Examination*, vol. 24, no. 1.
- [41] T. Denoeux and M. Masson, “Dempster-Shafer Reasoning in Large Partially Ordered Sets: Applications in Machine Learning,” in *Integrated Uncertainty Management and Applications*, ser. Advances in Intelligent and Soft Computing, V. Huynh, Y. Nakamori, J. Lawry, and M. Inuiguchi, Eds., vol. 68. Springer, Berlin, Heidelberg, 2010, pp. 39–54. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-11960-6_5
- [42] L. Fouque, A. Appriou, and W. Pieczynski, “An Evidential Markovian Model for Data Fusion and Unsupervised Image Classification,” in *Proc. of 3rd Int. Conf. on Information Fusion*, Paris, France, July 2000, pp. 25–32. [Online]. Available: <http://dx.doi.org/10.1109/IFIC.2000.862671>
- [43] A. Dempster, “A Generalization of Bayesian Inference,” *J. of the Royal Statistical Society*, vol. 30, no. 2, pp. 205–247, 1968. [Online]. Available: <http://www.jstor.org/stable/2984504>
- [44] N. Wilson, *Handbook of Defeasible Reasoning and Uncertainty Management Systems*. Springer, Dordrecht, 2000, vol. 5, ch. Algorithms for Dempster-Shafer Theory, pp. 421–475. [Online]. Available: http://dx.doi.org/10.1007/978-94-017-1737-3_10
- [45] F. Smarandache and J. Dezert, “Information Fusion Based on New Proportional Conflict Redistribution Rules,” in *Proc. of 7th Int. Conf. on Information Fusion*, Philadelphia, PA, USA, July 2005, pp. 25–29. [Online]. Available: <http://dx.doi.org/10.1109/ICIF.2005.1591955>
- [46] L. Breiman, “Random Forests,” Stat. Dep., U. of California, Berkeley, Tech. Rep., 2001.

- [47] V. Vapnik, *The Nature of Statistical Learning Theory*. NY: Springer, 1995.
- [48] —, *Statistical Learning Theory*. Wiley, 1998.
- [49] M. Liwicki, M. Malik, C. den Heuvel, X. Chen, C. Berger, R. Stoel, M. Blumenstein, and B. Found, “Signature Verification Competition for Online and Offline Skilled Forgeries (SigComp2011),” in *Proc. of 11th Int. Conf. on Doc. Anal. and Recognit.*, Beijing, China, Sept. 2011. [Online]. Available: <http://dx.doi.org/10.1109/ICDAR.2011.294>
- [50] J. Dezert, F. Smarandache, and M. Daniel, “The Generalized Pignistic Transformation,” in *Proc. of 7th Int. Conf. on Information Fusion*, Stockholm, Sweden, June 2004, pp. 384–391.
- [51] J. Fierrez-Aguilar, J. Ortega-Garcia, D. Ramos-Castro, and J. Gonzalez-Rodriguez, “HMM-based On-line Signature Verification: Feature Extraction and Signature Modelling,” *Pattern Recognition Letters*, vol. 28, no. 16, pp. 2325–2334, Dec. 2007. [Online]. Available: <https://doi.org/10.1016/j.patrec.2007.07.012>
- [52] I. Daubechies, *Ten Lectures on Wavelets*. Pennsylvania: SIAM: Society for Industrial and Applied Mathematics, 1992.
- [53] SigComp2011, “ICDAR 2011 Signature Verification Competition,” http://www.iapr-tc11.org/mediawiki/index.php/Datasets_List, 2011.
- [54] N. Brummer and J. du Preez, “Application-Independent Evaluation of Speaker Detection,” *Computer Speech & Language*, vol. 20, no. 2-3, pp. 230–275, April-July 2006.
- [55] A. Liaw and M. Wiener, “Classification and Regression by randomForest,” *R News*, vol. 2, no. 3.