www.ietdl.org

Published in IET Biometrics
Received on 13th February 2013
Revised on 13th May 2013
Accepted on 15th June 2013
doi: 10.1049/iet-bmt.2013.0024

Special Issue: Selected papers from International Conference on Frontiers in Handwriting Recognition (ICFHR 2012)

ISSN 2047-4938

Hybrid writer-independent–writer-dependent offline signature verification system

George S. Eskander, Robert Sabourin, Eric Granger

Laboratoire d'imagerie, de vision et d'intelligence artificielle, Ecole de technologie supérieure, Université du Québec, 1100 rue Notre-Dame Ouest, Room A-3600, Montréal, QC H3C 1K3, Canada
E-mail: geskander@livia.etsmtl.ca

Abstract: Standard signature verification (SV) systems are writer-dependent (WD), where a specific classifier is designed for each individual. It is inconvenient to ask a user to provide enough number of signature samples to design his WD classifier. In practice, very few samples are collected and inaccurate classifiers maybe produced. To overcome this, writer-independent (WI) systems are introduced. A global classifier is designed using a development database, prior to enrolling users to the system. For these systems, signature templates are needed for verification, and the template databases can be compromised. Moreover, state-of-the-art WI and WD systems provide enhanced accuracy through information fusion at either feature, score or decision levels, but they increase computational complexity. In this study, a hybrid WI–WD system is proposed, as a compromise of the two approaches. When a user is enrolled to the system, a WI classifier is used to verify his queries. During operation, user samples are collected and adapt the WI classifier to his signatures. Once adapted, the resulting WD classifier replaces the WI classifier for this user. Simulations on the Brazilian and the GPDS signature databases indicate that the proposed hybrid system provides comparative accuracy as complex WI and WD systems, while decreases the classification complexity.

1 Introduction

Signature verification systems (SV) are employed to authenticate individuals based on their handwritten signatures. There are two modes of operation for SV systems: online and offline. For online systems, users use special devices such as special pens and tablets to acquire signature trajectory dynamics such as velocity, pressure, etc. On the other hand, offline SV systems employ digitised signature images for authentication. Only static information can be acquired from the signature images, producing less informative signals, and hence, a harder pattern recognition task [1, 2].

Standard SV systems are writer-dependent (WD), where an individual classifier is designed for each user using his enrolment samples [3]. During verification, only query signature samples are processed by the classifier. Hence, WD systems are secure as no templates are stored for verification. Accuracy of these systems requires that users provide enough number of samples to train their classifiers. Hence, the WD approach implies a trade-off between accuracy and user convenience.

A more user-convenient approach is to design a writer-independent (WI) SV system. A single global classifier is designed using an independent (development) database prior to enrolling real users to the system. During verification, both query signature samples and at least one signature template are required to produce the classification decision. Hence, users can start using the system with providing a single signature sample. However, such systems are not secure as signature templates are needed for verification. The stored templates can be stolen, deleted or modified. Moreover, these systems do not model the individual signatures, but rather a universal model that should generalise on current and future users. Accordingly, the produced models are complex and ensemble methods are applied for enhanced performance at the expense of significantly increased complexity [4, 5].

This paper proposes a solution to compromise between the pros and cons of the WI and WD systems. A hybrid system is proposed where switching between the two approaches is possible. A universal WI classifier is designed with a development database. This enables starting system operation, even if users provide a single signature sample in the enrolment phase. Through operation, signature samples are collected and stored with the user profile. Once enough samples are collected for a specific user, they are used to adapt the universal classifier to this user. From this time on, the resulting WD classifier is used to verify signatures for the specific user. While the universal classifier compares the query samples to the stored user signature templates, the user-specific classifier only uses the query sample to produce the classification decision. Applying this scenario facilitates starting the system without asking the users to provide high number of enrolling samples. Then, switching to a more secure and less complex operational mode is possible whenever specific number of user samples exist.

To design the WI stage, pairwise dissimilarities are computed between feature representations of intra-personal and inter-personal samples from the development dataset.

Then, boosting feature selection (BFS) [6] is employed in a dissimilarity representation space [5]. To design the WD stage, the resulting global classifier is adapted to each user based on his stored samples. The features embedded in the WI classifier constitute a universal signature representation that can represent all users. Hence, subsets of this representation are discriminant for the different users. Accordingly, we tune the universal representation to a user by selecting the subset of feature representation that discriminates him from the others. To this end, stored user samples are represented in the universal feature space. These representations are used to train a WD classifier, by employing another BFS process that produces a more compact and secure classification system.

The proposed system is previously presented in [7]. In this paper, the robustness of the system is further investigated by conducted simulations using the public GPDS signature databases [8], besides the Brazilian database [9]. The next section provides an overview of state-of-the-art pure WI and WD offline SV systems. Section 3 describes the proposed WI–WD hybrid approach. Section 4 describes the experimental methodology applied in this paper. The experimental results are presented and discussed in Section 5.

2 Pure WD and WI signature verification systems

The design of WD systems relies on modelling user signatures in a feature representation space. Accuracy of the resulting models is limited by the available samples for training. Enhanced recognition rates of WD systems are recently achieved by training multi-classifier systems [10, 11]. On the other hand, WI systems do not produce models for the individual signatures, but rather a universal model that is valid for all users. In practice, it is impossible to locate a feature representation space in which signatures of all current and future users share the same distribution. The dissimilarity concept, where samples that belong to same class are similar, while samples that come from different classes are dissimilar, provides a solution.

The concept of dissimilarity-based classification has been proposed by Pekalska and Duin [12]. For this approach, the proximity between objects is modelled rather than modelling the objects themselves. Objects belong to a specific class have a shared degree of commonality that could be captured by a dissimilarity value. The dissimilarity measures can be derived in many ways, for example, from raw (sensor) measurements, histograms, strings or graphs. However, it can also be build on top of a feature representation [13].

In the SV context, the WI approach is realised using a dissimilarity (distance) measure, to compare samples (query and reference samples) as belonging to either the same or different user. As most of work on SV is feature-based, where many techniques of feature extraction are already proposed [1], the employed dissimilarity representations are built on top of a feature representation.

First implementation of the dissimilarity concept to the author identification domain was presented by Cha [14]. Simultaneous works are done to apply the dissimilarity learning concept on the offline SV problem by Srihari *et al.* [15], and Santos *et al.* [16]. While the first group used the correlation between binary features as a distance measure, the second group employed the Euclidean distance between graphometric feature vectors. In both implementations, the concept of WI-SV system was introduced. Instead of building a single WD classifier for each user using his enrolling signatures, a single global classifier is designed by learning the dissimilarities between signatures of all users. In [16], a neural network is trained to find the optimal boundary that splits the genuine and forgery classes in the dissimilarity representation space. Later, Oliveira *et al.* [17] and Bertolini *et al.* [4] applied the same concept, where they generated different dissimilarity spaces based on different feature representations. A set of SVM classifiers is trained to model the decision boundaries for the different subspaces. Finally, each SVM is used to produce a partial classification decision, while the final decision relies on the fusion of these partial decisions in the receiver operating curve (ROC) space. Kumar *et al.* [18] proposed a WI-SV based on surroundedness features.

More recently, Rivard *et al.* [5] extended the system in [4] to perform multiple feature extraction and selection. In this work, information fusion is also performed at the feature level. Multiple features are extracted based on multiple size grids. Fusion of these features and projecting them in a dissimilarity space results in dissimilarity representation of high dimensionality. This complex representation is then simplified by applying the BFS approach [6]. By applying the multi-feature approach with BFS, it was possible to design WI systems with higher performance than the earlier implementations. Moreover, the complex dissimilarity representation (possibly tens of thousands of features) is condensed to a compact universal representation of few hundreds in dimensionality. This representation can classify samples from unknown users, whose signatures had no share in the training process. The accuracy of this WI system could be enhanced through combining multiple decisions based on multiple templates.

Pure WI are insecure because of the need to store reference signatures for verification; however, they are user-convenient as they do not need user samples for training. On the other hand, pure WD are secure but user-inconvenient. Both techniques can produce SV systems with acceptable accuracy, but they are complex because of the fusion of responses from multiple classifiers. The work proposed in this paper merge the two techniques to overcome the limitations of the pure approaches. To this end, a WI system is designed as in [5], to start system operation with only one enrolling sample. Then, the universal representation embedded in the WI classifier is adapted to each specific user, whenever enough genuine samples are collected. This adaptation step aims to reduce the classification complexity (number of features and number of classifications fused for a decision), and avoiding the need of using reference signatures for verification.

3 A hybrid WI–WD signature verification system

3.1 Theoretical basis

It is not easy to achieve high generalisation accuracy for the offline SV systems, owing to the high intra-personal variability and inter-personal similarity of signature images. One approach to tackle this problem is to select the most stable and discriminant features from a large pool of feature extractions. To illustrate this approach, see Fig. 1 (left side). In this illustrative example, signatures of three different writers are represented in the feature space. Assume the candidate pool of features is $F = \{f_i\}_{i=1}^{K}$. For simplicity,

IET Biom., 2013, Vol. 2, Iss. 4, pp. 169–181
doi: 10.1049/iet-bmt.2013.0024

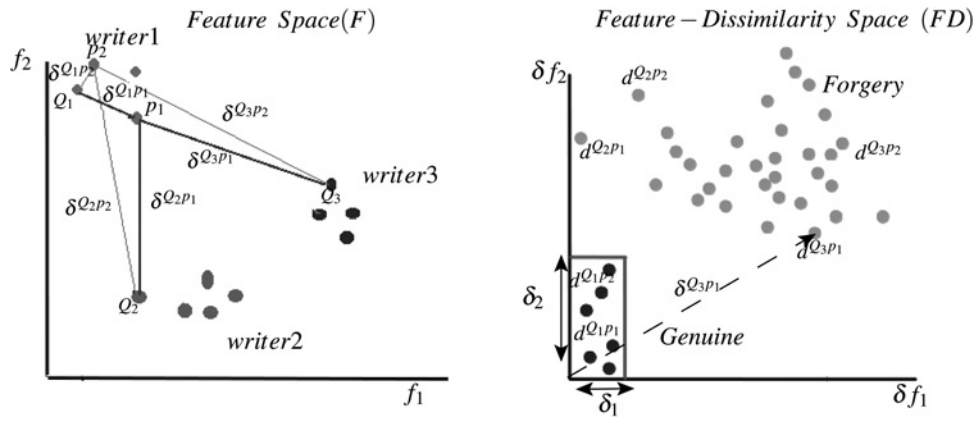

Fig. 1 *Illustration of feature selection in the original feature space (left) and in the feature-dissimilarity space (right)*

only two features f_1 and f_2 are shown in this figure, while the dimensionality of this space is K, which might be a high number for typical representations. For good generalisation performance, within-class (intra-personal) distances should be small, while between-class (inter-personal) distances should be large. For instance, assume that writer 1 has two prototypes (templates) p_1 and p_2. Good generalisation implies that, feature representation of any query Q_1 of this writer should be close to his prototypes, while queries of other writers as Q_2 and Q_3 should be far from them.

The proposed approach relies on selecting a condensed feature representation of dimensionality L, from a very high-dimensional space of dimensionality K, so that distances between intra-personal signature representations are minimised and the inter-personal distances are maximised. Consider the Euclidean distance, so that distance between a signature sample Q_j and a prototype p_r is $\delta^{Q_j p_r}$

$$\delta^{Q_j p_r} = \sqrt{\sum_{i=1}^{L} \left(\delta f_i^{Q_j p_r}\right)^2} \tag{1}$$

where $\delta f_i^{Q_j p_r} = \| f_i^{Q_j} - f_i^{p_r} \|$.

Hence, the overall distance between two signatures is an accumulation of the individual distances between every two corresponding features of the signature representations. To increase the separation between the intra-personal and inter-personal distance ranges, we select features that decrease the intra-personal distances and that increase the inter-personal distances. In this illustrative example, it is obvious that features f_1 and f_2 are discriminative. Distances among intra-personal signatures (like $\delta^{Q_1 p_1}$) are generally smaller than the distances among inter-personal signatures (like $\delta^{Q_2 p_1}$). However, in this space it is not clear which feature is more discriminative. With representations of high dimensionality, high number of users, unknown forgeries and a small number of training samples, it is not feasible to select the best features in the feature space F.

Accordingly, we project this representation on a feature-dissimilarity space FD, as shown in the right side of Fig. 1. In this space, distance between each corresponding features, for each pair of signatures, is computed and used as a new set of features $\{\delta f_i\}_{i=1}^{K}$. So, dimensionality of the F and FD spaces is the same. A distance $\delta^{Q_j p_r}$ between a query Q_j and a prototype p_r is mapped from F to FD as a point $d^{Q_j p_r}$

$$d^{Q_j p_r} = \left\{\delta f_i^{Q_j p_r}\right\}_{i=1}^{K} \tag{2}$$

where, $\delta^{Q_j p_r}$ is represented by the distance from the origin point to $d^{Q_j p_r}$. Here, the impact of every individual feature on the signature dissimilarities is clear. It is obvious that f_2 is more discriminative than f_1. For all genuine query samples such as Q_1, $\delta f_2^{Q_1 p_r} < \delta_2$ and for all forgery query samples such as Q_2 and Q_3, $\delta f_2^{Q_j p_r} > \delta_2$. On the other hand, f_1 is less discriminant. For the forgery query Q_2, $\delta f_1^{Q_2 p_1} < \delta_1$, same as that for the genuine sample Q_1. Accordingly, it is easier to rank and select features in the FD space, as the impact of the individual features on the overall dissimilarity is clear in this space. Moreover, while signatures of users are modelled in the F space, the proximity between user signatures are modelled in the FD space. This property maps the multi-class problem, with few training samples per class, to a two-class problem, with more training samples per class. The constituted classes are: the genuine class and the forgery class. Samples of the genuine class result from comparing two signatures of the same person. The forgery class samples result from comparing two signatures of different persons.

Employing the aforementioned feature selection in the FD space, increases the separation between the genuine and forgery classes, and hence decreases the generalisation error. Moreover, Rivard *et al.* [5] have shown that, classifiers that are designed through feature selection in such dissimilarity representation spaces, can generalise for even users whose signature templates are not used during the design phase. Accordingly, if a classifier is trained with samples from a specific signature database, the same classifier can be used to detect signatures from another database with good accuracy. This observation leads to the concept of WI-SV systems, where an independent (development) database is used to design a classifier, and then signatures of real system users are detected by this global classifier. Good generalisation performance could be achieved, when the development database consists in a large enough number of users. As the resulting classifier can classify samples from unknown users, whose signatures had no share in the training process; so the embedded condensed representation of dimensionality $L < K$ is considered as a population-based representation.

However, such WI-SV systems have some drawbacks. First, in order to model the proximity of samples that

belong to a large population, the dimensionality L of the population-based feature representation is high, and that produces complex classifiers. Second, the dissimilarity feature representation relies on signature prototypes (templates). Storing user templates in a database might cause security vulnerabilities, as stored signatures can be stolen or edited.

In this paper, the drawbacks of a WI–SV system are alleviated through adapting it to specific users. The adaptation approach relies on two main hypothesis:

- For each specific user, the population-based representation of dimensionality L contains a feature subset of dimensionality $N < L$, where this more concise representation discriminates the specific user from the other users. The logic behind this hypothesis is that: although the global representation could represent the specific user, not all of the representation dimensions are mandatory for discriminating the user. Also, the importance of representation dimensions differs for the different users. So, re-ordering the features for each user and selecting the most important subset, produces a more compact and maybe more discriminant representation space.
- Features that are discriminant when represented in the FD space are discriminant when represented in the F space. For instance, if δf_i is discriminant in FD, then this implies that f_i is discriminant in F. This is because that, translation between F and FD spaces can be considered as a direct mapping, where this mapping does not impact the Euclidean distances between the signature representations. This property facilitates the design of the user-specific classifier in the feature space; so no signature templates are needed for verification, and hence a more secure WD–SV could be designed. To this end, the population-based representation obtained in the WI–SV design phase, is translated back to a feature space of the same dimensionality L. Then, user-specific feature selection and classifier design processes are employed in the feature space, to produce a more condensed user-specific space of dimensionality $N < L < K$. As the population-based representation is much condensed than the original feature space, so the few training samples available for real system users, can be used to search for the most discriminant space dimensions.

3.2 System overview

Fig. 2 shows a block diagram of the proposed hybrid WI–WD system in training and verification modes. First, a WI-SV sub-system is designed as proposed by Rivard *et al.* [5]. Query samples of recently enrolled users are verified by this sub-system. After collecting a specific number of reference signatures of a user, they are used to adapt this global sub-system to the specific user. To this end, features embedded in the designed WI–SV are considered as a universal (population-based) representation. The user and forgery samples are translated into this universal space and used to train a WD–SV sub-system. From this time on, signatures of the specific user are verified by his WD–SV sub-system.

3.3 WI training

A feature-level fusion is performed, by employing a multi-type multi-scale feature extraction. In literature, many types of features could be extracted from offline signature images [1]. Any combination of these features may be concatenated into a single high-dimensional representation, and used for the proposed framework. However, we focused on using feature extracted using extended-shadow-code (ESC) [19] and directional probability density function (DPDF) [20]. Features are extracted based on different grid scales, hence a range of details are detected in the signature image. These features have shown complementary functionality: while ESC detects the distribution of the signature in the spatial space, DPDF detects the orientation of the signature strokes. For more details on the employed the multi-feature extraction technique, see [5].

A development signature database is used to train the WI–SV classifier. To this end, the multi-feature representations M_G and M_F are extracted from some genuine signature samples S^G and forgery signature samples S^F, respectively, where $M = \{m_k\}_{k=1}^{K}$, and K is the dimensionality of the multi-feature representation. To project these samples into a dissimilarity representation space, dichotomy transformation is applied. For instance, for two samples M_i, M_j, the dissimilarity feature is

$$D_{ij}^{M} = \left\| M_i - M_j \right\| = \{\Delta m_k\}_{k=1}^{K} \tag{3}$$

where $\Delta m_k = \|m_{i_k} - m_{j_k}\|$, and D^M is the dissimilarity representation built on top of the multi-feature representation M.

It is worth noting that both the multi-feature and dissimilarity representations have the same dimensionality K. Also, a sample D_{ij}^{M} is labelled as a within-class or as a between-class instance, when it results from two genuine signatures of the same user, or from two signatures of two different users, respectively.

To build the WI-SV system, the BFS approach is applied [6]. This method applies Gentle AdaBoost algorithm [21] to learn an optimal decision boundary between the within-class and the between-class dissimilarity samples, by boosting Decision Stump (DS) weak learners [22]. At a boosting iteration t, a DS is designed by locating the best dimension d_t in the dissimilarity representation space that splits the training samples based on a splitting threshold δ_t. The DS either has positive or negative polarity, depending on the direction of splitting the classes. At a boosting iteration t, a DS_t is formulated as

$$DS_t(d_t) = \begin{cases} p_t^{\text{left}} & \text{if} \quad (d_t < \delta_t) \\ p_t^{\text{right}} & \text{otherwise} \end{cases} \tag{4}$$

p_t^{left}, p_t^{right} represent the confidence of decisions taken by this DS, when the feature value lies to the left or to the right of the splitting threshold, respectively. Accordingly, each DS shares in the final classification decision based on its expected accuracy. The boosting process runs for T^{wi} boosting iterations, and the final decision boundary is defined by

$$H^{\text{wi}} = \sum_{t=1}^{T^{\text{wi}}} DS_t^{\text{wi}} \tag{5}$$

where DS_t^{wi} is the DS designed at boosting iteration t based on the development data, and T^{wi} is the number of boosting iteration in the WI training process. See [5] for more detailed algorithms of the WI–SV design process.

This WI–SV sub-system is used to verify user signatures, before switching to the WD mode (the WI verification

IET Biom., 2013, Vol. 2, Iss. 4, pp. 169–181
doi: 10.1049/iet-bmt.2013.0024

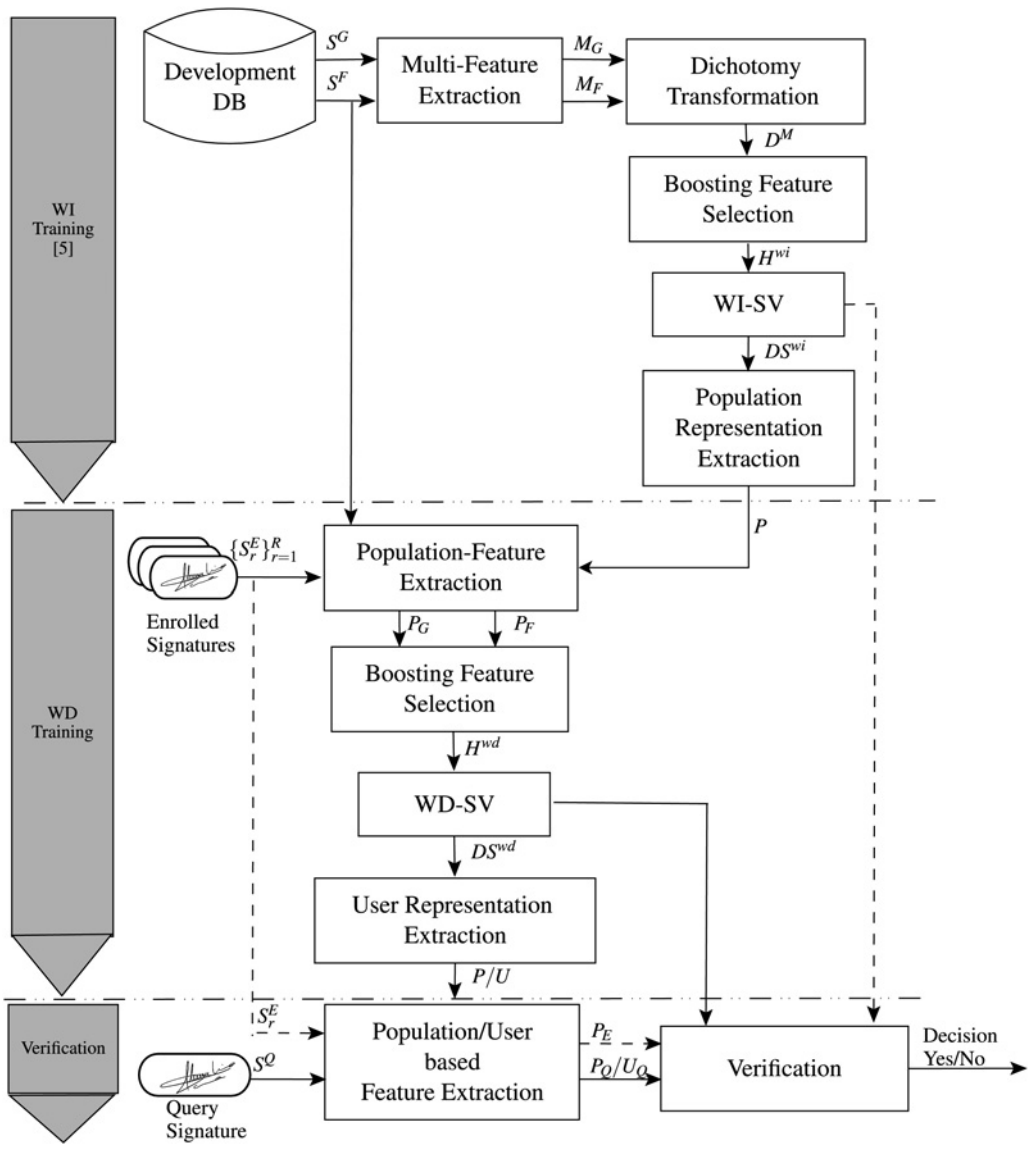

Fig. 2 *Hybrid WI–WD SV system*

mode is represented by dotted arrows in Fig. 2 and illustrated in Section 3.5). However, after collecting enough number of genuine user samples, they are used to adapt the WI–SV sub-system to the user samples. To this end, the WI–SV is used for dimensionality reduction through WI feature selection. The feature representation embedded in a WI classifier is extracted and stored as a population-based representation $P = \{p_l\}_{l=1}^{L}$ of dimensionality $L < K$, by which signatures of all users are represented. This step reduces the representation dimensionality, and allows for the design of compact user-specific (WD) classifiers.

3.4 WD training

Although the universal representation P contains discriminant features for all users, not all dimensions of this space are needed to discriminate specific users from other populations. Moreover, the dissimilarity thresholds selected in the WI system are not optimal for each user. In this design step, selection of discriminant features for each specific user is achieved, while selecting the best splitting threshold in each dimension.

While the WI training phase should be performed in the dissimilarity space (as it is impossible to locate a feature representation space in which signatures of all current and future users share the same distribution), the WD training phase, on the other hand, could be performed in either the dissimilarity or the original feature space. Operating the SV system in the feature space is more secure, as no signature references need to be stored for verification. Accordingly, the WD training phase is implemented in the feature space.

To this end, the population-based representation (P) of dimensionality L is used for feature extraction. For each enrolled user, R signature samples are collected. Both the enrolling samples $S^E = \{S_r^E\}_{r=1}^R$ and some samples S^F are selected from the development database (to represent the random forgery class), are represented in the P feature space as P_G and P_F, respectively. Finally, a similar BFS process is applied, by using this WD data to model the decision boundary H^{wd} that splits the genuine and forgery classes, where

$$H^{wd} = \sum_{t=1}^{T^{wd}} DS_t^{wd} \tag{6}$$

where DS_t^{wd} is the decision stump designed at boosting iteration t based on the WD training data, and T^{wd} is the number of boosting iteration in the WD training process.

3.5 Signature verification

The WI–SV can be used whenever no user samples are available to train a WD classifier. Switching between WI and WD approaches may depend on the availability of sufficient user samples for training.

3.5.1 WI–SV mode: This operational mode is illustrated by the dotted arrows in Fig. 2. A questioned signature S^Q and a single enrolment sample S_r^E are represented in population representation space (P) of dimensionality $L < K$ as $P_Q = \{p_{Q_l}\}_{l=1}^{L}$ and $P_E = \{p_{E_l}\}_{l=1}^{L}$, respectively. Then, it is classified by the WI–SV system, where

$$WI - SV\left(D_{QE}^P\right) = \mathrm{sign}\left(\sum_{t=1}^{T^{\mathrm{wi}}} DS_t^{\mathrm{wi}}\left(D_{QE}^P\right)\right) \tag{7}$$

where D_{QE}^P is the dissimilarity representation of the query sample S^Q built on top of the population-based feature representation P

$$D_{QE}^P = \|P_Q - P_E\| = \{\Delta p_l\}_{l=1}^{L} \tag{8}$$

where $\Delta p_l = \|p_{Q_l} - p_{E_l}\|$.

3.5.2 WD–SV mode: To authenticate a specific user in this operational mode, the corresponding WD–SV classifier is used. First, the feature representation embedded in the WD–SV is extracted and considered as a user-based representation (U) of dimensionality $N < L < K$. Then, the query image S^Q is represented in this concise representation space as $U_Q = \{u_{Q_n}\}_{n=1}^{N}$, and then fed the classifier for recognition, where

$$WD - SV\left(U_Q\right) = \mathrm{sign}\left(\sum_{t=1}^{T^{\mathrm{wd}}} DS_t^{\mathrm{wd}}\left(U_Q\right)\right) \tag{9}$$

4 Experimental methodology

Performance of the proposed hybrid WI-WD SV system is investigated by considering its two modes of operation:

- WI-SV mode – in this mode, the query signature samples are verified by applying (7). The objective of investigating this operational mode is to measure the minimum accuracy of the system. In this case, it is assumed that only single signature sample is obtained in the enrolment phase and used for the verification task.
- WD–SV mode – in this mode, the query signature samples are verified by applying (9). The objective of testing this operational mode is to determine a reasonable number of user samples that produce reliable user-specific classifiers. To this end, performance of the WD classifiers designed with different number of samples is investigated. Suitable switching point between the WI and WD modes is identified by the number of training samples that produce WD classifiers with higher accuracy than the global WI classifier.

4.1 Signature databases

Two different off-line signature databases are used for proof-of-concept simulations: the Brazilian SV database [9] and the GPDS database [8]. While the Brazilian SV database is composed of random, simple and simulated forgeries, the GPDS database is composed of random and simulated forgeries. Random forgeries occur when the query signature presented to the system is mislabelled to another user. Also, forgers produce random forgeries when they know neither the signer's name nor the signature morphology. For simple forgeries, the forger knows the writer's name but not the signature morphology. He can only produce a simple forgery using his style of writing. Finally, simulated forgeries imitate the signatures as they have access to a genuine signatures sample.

4.1.1 Brazilian database: The Brazilian signatures database contains signatures of 168 users, that were digitised as 8-bit greyscale images over 400 × 1000 pixels, at resolution of 300 dpi. It is split into two parts. The first part contains signatures of the first 60 users. For each user, there are 40 genuine samples, ten simple and ten simulated forgeries. A subset of this part is used for WD training, so it is referenced in this paper as B^{wd}. The remaining of this part is used for performance evaluation (see Table 3). The second part contains signatures of the last 108 users. For each user, there are only 40 genuine signatures. This part is used for WI training, so it is referenced in this paper as B^{wi} (see Table 1).

4.1.2 GPDS database: The GPDS database contains signatures of 300 users that were digitised as 8-bit greyscale at resolution of 300 dpi. This database contains images of different sizes (that vary from 51 × 82 pixels to 402 × 649 pixels). All users have 24 genuine signatures and 30 simulated forgeries. It is split into two parts. The first part contains signatures of the first 160 users. A subset of this part is used for the WD training, so it is referenced in this paper as G^{wd}. The remaining of this part is used for performance evaluation (see Table 4). The second part contains signatures of the last 140 users. This part is used for the WI training, so it is referenced in this paper as G^{wi} (see Table 2).

Table 1 The Brazilian development database (B^{wi}): 108 users × 40 genuine signatures each

Training set (30 signatures/user)		Validation set (10 signatures/users)	
Within-class	Between-class	Within-class	Between-class
distances among the 30 signatures/ user	distances among 29 signatures/ user and 15 signatures of other users	distances among the 10 signatures/ user and the 30 signatures of the training set	distances among the 10 signatures/ user and 30 signatures selected randomly from other users
108 × 30 × 29/ 2 = 46 980 samples	108 × 29 × 15 = 46 980 samples	108 × 10 × 30 = 32 400 samples	108 × 10 × 30 = 32 400 samples

IET Biom., 2013, Vol. 2, Iss. 4, pp. 169–181
doi: 10.1049/iet-bmt.2013.0024

Table 2 The GPDS development database (G^{wi}): 140 users × 24 genuine signatures each

Training set (14 signatures/user)		Validation set (10 signatures/users)	
Within-class	Between-class	Within-class	Between-class
within-class distances among the 14 signatures/user	between-class distances among 13 signatures/user and 7 signatures of other users	within-class distances among the 10 signatures/user and the 14 signatures of the training set	between-class distances among the 10 signatures/user and 14 signatures selected randomly from other users
140 × 14 × 13/2 = 12 740 samples	140 × 13 × 7 = 12 740 samples	140 × 10 × 14 = 19 600 samples	140 × 10 × 14 = 19 600 samples

4.2 Feature extraction

A set of 30 grid scales is used for both of the ESC and DPDF feature types, producing 60 different single-scale feature representations. These representations are then fused to produce a multi-feature representation M of huge dimensionality ($K = 30\,201$) [5].

4.3 WI training

Tables 1 and 2 describe the development dataset used in the WI training stage, for the Brazilian and the GPDS databases, respectively. For the Brazilian database, a total of 93 960 samples are used for training, and 64 800 are used as holdout validation set to avoid overfitting. For the GPDS database, a total of 25 480 samples are used for training, and 39 200 are used for holdout validation.

Multi-feature representations of signature images of both the training and validation sets are produced. Then, these representations are fed to the BFS process. The BFS algorithm is set for 1000 max boosting iterations and 100 early stopping criteria. For the Brazilian database, the constituted WI-SV classifier contained 679 decision stumps, with them only 555 distinct features are used. (i.e. $T^{wi} = 679$, $L = 555$). For the GPDS database, the WI-SV classifier contained 998 decision stumps, with them only 697 distinct features are used. (i.e. $T^{wi} = 998$, $L = 697$).

4.4 WD training

Tables 3 and 4 describe the data sets used to build the WD classifiers and for performance evaluation, for the Brazilian and the GPDS, respectively. To investigate the impact of training samples quantity on the recognition performance (and hence determining a reasonable switching point between the WI and WD modes), different number of samples are used to train the WD classifier. The forgery class is represented by genuine signatures from the development database. Genuine and forgery samples are represented in the population (P) space and used for training (dimensionality of P is $L = 555$ in case of the Brazilian database, and $L = 697$ in case of the GPDS database). Then, these representations are fed to the BFS process. For the WD training, fixed number of boosting iteration T^{wd} is used for early stopping. For both databases, we observed saturation in performance about 100 boosting iterations; so the boosting iterations were set to a fixed number (here, the performance is reported for two cases where, $T^{wd} = 20$ and 100).

4.5 Performance measures

The testing sets are illustrated in Tables 3 and 4. For the Brazilian database, 40 test samples per user are employed. Of them, ten genuine, ten random, ten simple and ten simulated forgeries, for a total of 2400 questioned signatures are employed for system evaluation. For the GPDS database, 50 test samples per user are employed. Of them, ten genuine, ten random and 30 simulated forgeries, for a total of 8000 questioned signatures.

The area under ROC curves (AUC) and the average error rate (AER) are used to evaluate the accuracy of classifiers in this paper. For AUC computations, the questioned signatures S^Q of the test set are processed by a classifier. Its outputs are then sorted, and used as a set of classifier thresholds. Then, the genuine accept rate (GAR) and false accept rate (FAR) are computed for each specific threshold. Finally, the ROC curve is plotted using the generated GAR and FAR values, and the AUC is computed [AUC values are used here only as quality indicators for the constituted classifiers. During the WD training phase, we do not take any design selections (like early stopping, decision threshold, etc.), based on the ROC curves and their AUC values. Such decisions are only taken based on the

Table 3 The Brazilian WD database (B^{wd}): 60 users × 60 signatures each: 40 genuine + 10 simple forgery + 10 simulated forgery

Training set (30 signatures/user)		Testing set (30 signatures + 10 random forgeries/user)	
genuine-class signature subsets of different sizes	forgery-class signatures of the training set of the B^{wi} dataset	genuine-class remaining 10 genuine signatures/user	forgery-class 10 simple + 10 simulated + 10 random forgery selected randomly from other users in B^{wd}
5,7,9,11,13,15,30 samples	108 × 30 = 3240 samples	60 × 10 = 600 samples	60 × 30 = 1800 samples

Table 4 The GPDS WD database (G^{wd}): 160 users × 54 signatures each: 24 genuine + 30 simulated forgery

Training set (14 signatures/user)		Testing set (40 signatures + 10 random forgeries/user)	
genuine-class signature subsets of different/user sizes	forgery-class signatures of the training set of the G^{wi} dataset	genuine-class remaining 10 genuine signatures/user	forgery-class 30 simulated + 10 random forgery selected randomly from other users in G^{wd}
4,8,12,14 samples	140 × 14 = 1960 samples	160 × 10 = 1600 samples	160 × 40 = 6400 samples

development dataset in the WI design phase.]. AUC classifiers are averaged over all users of the testing dataset. The AER is computed as follows

$$\mathrm{AER} = \left(\mathrm{FRR} + \mathrm{FAR}_{\mathrm{random}} + \mathrm{FAR}_{\mathrm{simple}} + \mathrm{FAR}_{\mathrm{simulated}}\right)/4 \tag{10}$$

where FRR is the false rejection rate, and $\mathrm{FAR}_{\mathrm{random}}$, $\mathrm{FAR}_{\mathrm{simple}}$ and $\mathrm{FAR}_{\mathrm{simulated}}$ are the false accept rates when verifying random, simple and simulated forgeries, respectively (for the GPDS only random and simulated forgeries are considered).

Computational complexity of the designed classifiers is evaluated by the total number of feature values (TFV) that are extracted and processed to produce the final classification decision [23],

$$\mathrm{TFV} = \sum_{i=1}^{n} m_i x_i \tag{11}$$

where n is the number of partial classification decisions that cooperate to produce the final decision, m_i is the number of features per sample processed by a classifier i, and x_i is the number of signature samples processed by a classifier i.

5 Simulation results

Simulations reported in this section address two main objectives:

- Feasibility of using the proposed system in its both verification modes: practical switching point between the two modes is identified. Also, robustness and computational complexity of the system are investigated.
- Comparing the proposed hybrid system with other pure WI and WD systems in the literature.

5.1 Performance of the WI and WD verification modes

The AUC and AER are computed for both the WI-SV and some WD-SV classifiers designed with different number of training samples and boosting iterations. Based on these measures, we observe a suitable point to switch between the two modes. This point is globally determined for all system users, and identified by the number of training samples that produce a WD classifier outperforms the baseline WI-SV classifier. Robustness of the system is investigated by comparing the verification performance for both the Brazilian and the GPDS experimental databases. Finally, computational complexity and the computer processing times are compared for the two verification modes.

5.1.1 Brazilian database: Fig. 3 shows the AUC of the WI-SV classifier and some WD-SV classifiers, designed with different number of boosting iterations and with different training set sizes. It is shown that with only five training samples, and only 20 boosting iterations, the average AUC of the WD classifier is 0.923. The classifier performance increases when increasing both training set size and boosting iterations (the average and minimum values of AUC are increasing). A WD classifier with 13 training samples has same AUC as the WI classifier [For the WI-SV system, the ROC curve is computed for a single WI classifier tested with a single signature template ($R = 1$). Here, no techniques for decision fusion are applied to generate enhanced ROC curves.]. WD classifiers trained with more samples outperform the WI classifier.

Fig. 4 shows the AER for the different classifiers. Same trend is noticed as that of Fig. 3. However, a higher number of training samples (20) is needed to produce a WD–SV with smaller AER than that of the baseline WI-SV. The reason of this difference in determining the optimal switching point is because we do not tune the decision thresholds of the classifiers. So, some accurate classifiers may produce high AER when the decision threshold is not optimised. For 100 boosting iterations, the secure WD classifier trained with 20 training samples has same performance (AER = 7.24%) as that of insecure WI classifier (tested with a single template ($R = 1$)). Also, a WD classifier trained with 30 training samples has same performance (AER = 5.38%) as that of insecure WI classifier (tested with 15 templates ($R = 15$)).

Fig. 5 shows the FRR and FAR for different level of forgeries, when WD–SV classification mode is employed. It is clear that, although the FRR decreases with using more training samples and boosting iterations, the $\mathrm{FAR}_{\mathrm{simulated}}$ increases. However, the $\mathrm{FAR}_{\mathrm{simple}}$ and $\mathrm{FAR}_{\mathrm{random}}$ are neglected when compared with the other error rates.

5.1.2 GPDS database: Fig. 6 shows the AUC of the WI–SV classifier and WD–SV classifiers for the GPDS

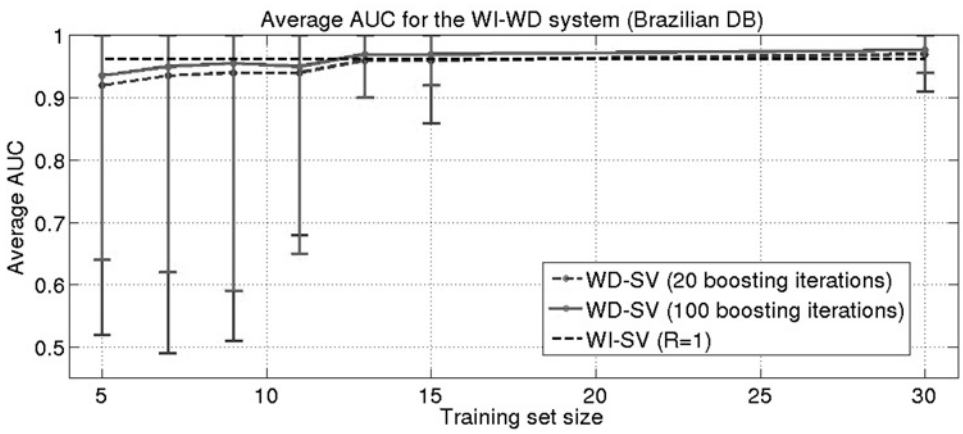

Fig. 3 *Average AUC of ROC curves for the WI and WD classifiers for the Brazilian database*

The points represent the average AUC over the 60 users, and the vertical bars represent the range between the maximum and minimum AUCs for the 60 users

IET Biom., 2013, Vol. 2, Iss. 4, pp. 169–181
doi: 10.1049/iet-bmt.2013.0024

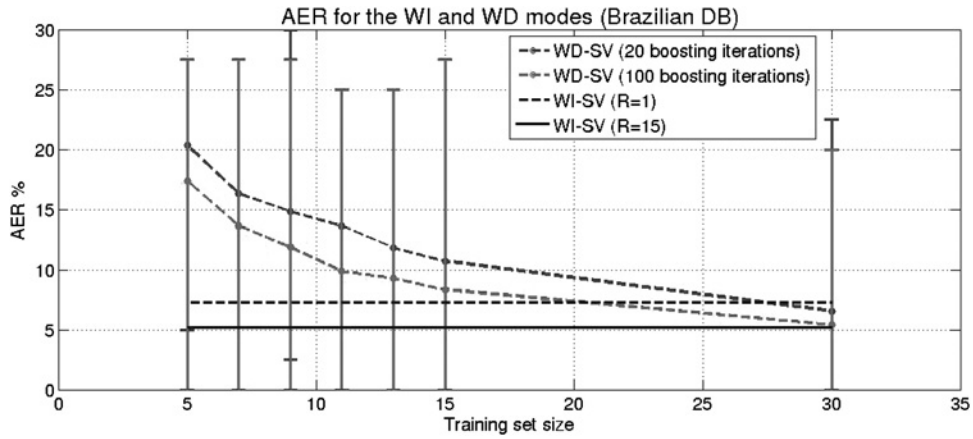

Fig. 4 *AER for the WI and WD classifiers for the Brazilian database*

For 100 boosting iterations, the secure WD classifier trained with 20 training samples has same performance (AER = 7.24%) as that of insecure WI classifier (tested with a single template ($R = 1$)). WD classifiers trained with more samples outperforms the WI classifier

database. Robustness of the proposed system is clear as similar performance trend is shown as that of the Brazilian database. With both databases, classifier performance increases when increasing both training set size and boosting iterations (the average and minimum values of AUC are increasing). However, for the GPDS database, the system has shown lower performance.

Although a global training size can be determined, and it results in a better accuracy than that of the original WI-SV system, the best training size differs for the different users. For instance, Fig. 7 shows the ROC curves for a specific user, where eight training samples produce WD-SV systems with higher performance than that for the WI-SV system. Fig. 8 shows the ROC curves for another user where only

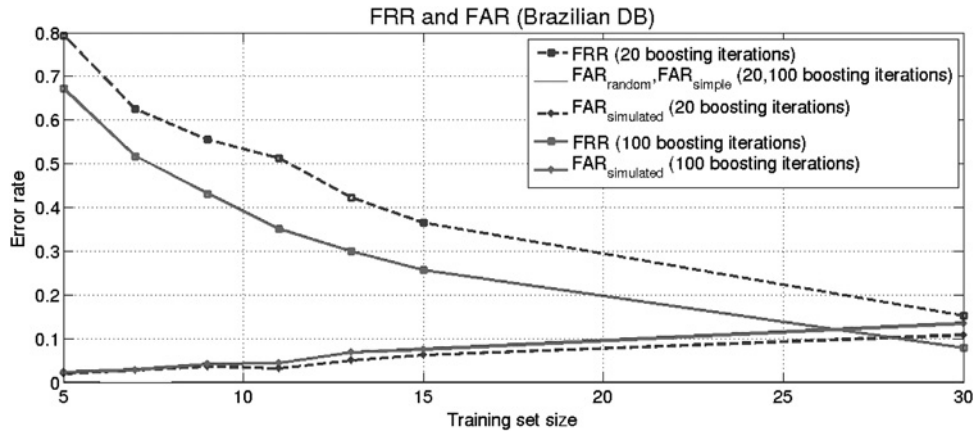

Fig. 5 *FRR and FAR for the WD-SV mode for the Brazilian database*

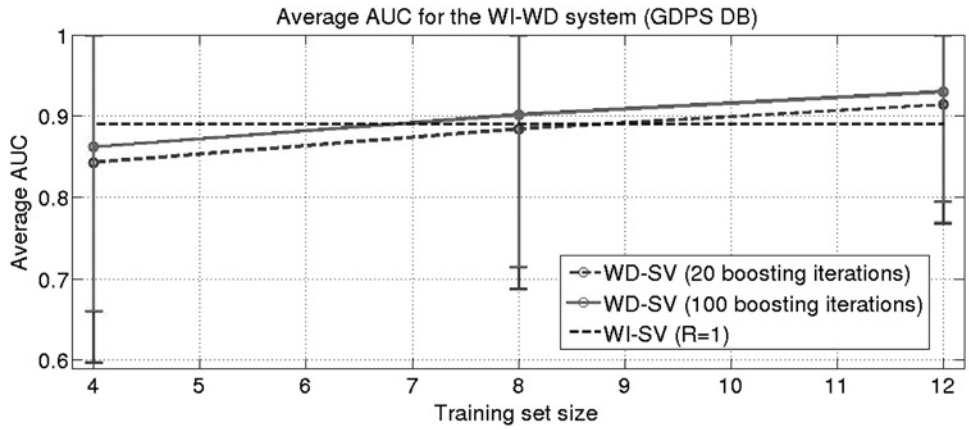

Fig. 6 *Average AUC of ROC curves for the WI and WD classifiers for the GPDS database*

Classifiers performance increase when increasing both training set size and boosting iterations. The WD classifier with only 8 training samples has same AUC as the WI classifier. WD classifiers trained with more samples outperforms the WI classifier

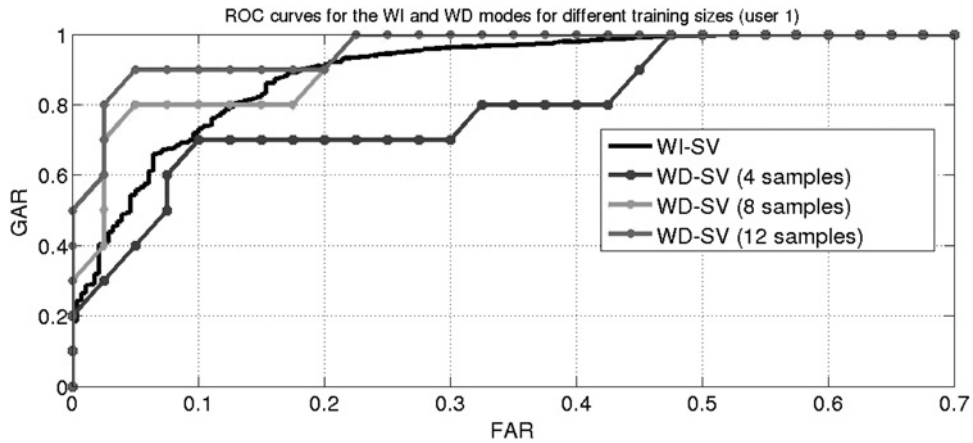

Fig. 7 *ROC curves for the WI and WD modes for different training sizes (user 1)*

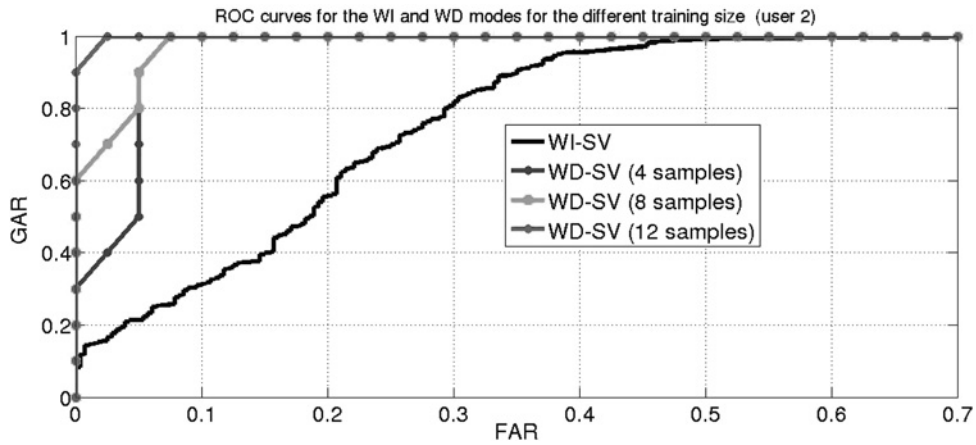

Fig. 8 *ROC curves for the WI and WD modes for different training sizes (user 2)*

four training samples could produce WD-SV systems with higher performance than that for the WI-SV system, as the performance of the WI classifier is very weak for this specific user. Future work is needed to investigate possibility of employing user-specific training size.

Moreover, the classification decisions of both the WI and WD classifiers can be fused in the ROC space, and might produce better performance. For instance, a recently fusion method called IBC, proposed by Khreich *et al.* [24], could be employed to fuse decisions from multiple ROC curves. Such fusion of both modes might be beneficial for the starting operational period (before the switching point to the WD is reached). So, future work will investigate fusing the two classifiers during this operational period.

Figs. 9 and 10 show the AER and FRR/FAR for the different classifiers, respectively. Different than the

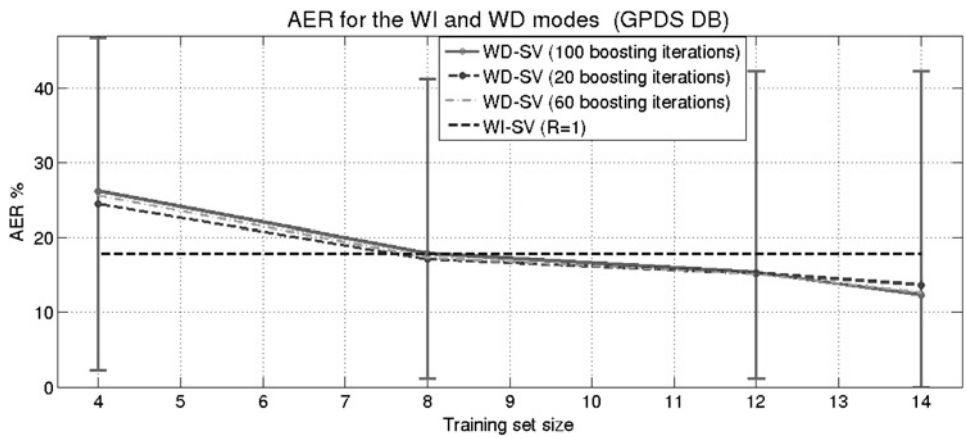

Fig. 9 *AER for WD-SV verification mode, for the GPDS database*

The WD classifier with 8 training samples has same AER as the WI classifier (tested with a single template ($R = 1$)). WD classifiers trained with more samples outperforms the WI classifier

IET Biom., 2013, Vol. 2, Iss. 4, pp. 169–181
doi: 10.1049/iet-bmt.2013.0024

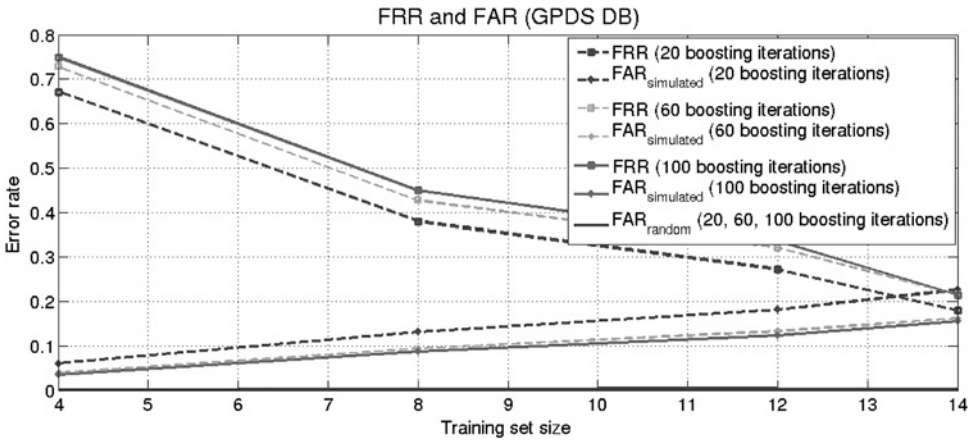

Fig. 10 *FRR and FAR for the WD-SV mode for the GPDS database*

Brazilian database, signatures of the GDPS seem to be less stable as overfitting occurs with only 20 boosting iterations. In Fig. 10, it is clear that the FRR increases after 20 boosting iterations, while the $FAR_{simulated}$ decreases. Also, although the average AER is acceptable (about 12.5% with 14 training samples and 100 boosting iterations), some users have shown inaccurate performance (maximum AER is about 42%).

The explanation of the lower performance of the system, when tested on the GPDS database, than that of the Brazilian database is: while the signature images of the Brazilian database have fixed size, the GPDS database includes images with various sizes. Hence, the population-based feature representation (P) that is designed based on the development dataset may not generalise to some users in the WD dataset, whose signature sizes differ significantly. Accordingly, the proposed system is expected to produce high classification accuracy when employed in real-world SV applications, where signature samples have fixed size as they come from a same type of document, that is, checks from a specific bank.

5.1.3 Computational complexity: To compare the computational complexity of the two modes of verification, we investigated the complexity of a WI-SV and a WD-SV with similar accuracy. For instance, for the Brazilian database, the WI-SV when tested with fifteen template ($R = 15$) is compared with the WD-SV when trained with 30 samples, as both have similar accuracy. For the WI-SV system, the classifier produces the classification decision based on processing of 555 features extracted from a query sample and a template. Hence, when 15 templates are used (1 query + 15 templates = 16 signature images are processed), the TFV is 8880 (see (11)). On the other hand, the WD-SV system frequently produces decisions based on a single classification operation. Only a query sample is used for feature extraction, where about 40 features are processed by the classifier. Hence, the TFV is about 40. Accordingly, adaptation of this WI system to different users reduces the computational complexity by about 99.5%. While similar accuracy is achieved by employing either the WI-SV or the WD-SV systems, the later is lighter and more secure.

We observed consistent computer simulations outcomes. The total verification processing time is dominated by the representation extraction process. While the classifiers compute the classification decision in about 10^{-5} s, the extraction time for population-based representation P and the user-based representation U, for a single image, are 0.25 and 0.02 s, respectively. Hence, the representation extraction time for WI-SV verification mode is about 4 s (for 16 images), and it is about 0.02 s for the WD-SV mode. Accordingly, adaptation of the WI-SV to specific users reduced the verification time by about 99.5%.

5.2 Comparisons with systems in the literature

The proposed systems is compared with pure WI systems on the Brazilian database [4, 5, 16] and on the GPDS database [18]. Also it is compared with pure WD systems on the Brazilian database [3, 10, 11] and on the GPDS database [11, 25–27].

5.2.1 Brazilian database: Table 5 compares the proposed WI-WD system to some pure WI and WD systems that are tested on the Brazilian database. All systems are investigated using the same data set and similar testing protocol, and the results are reported in terms of AER. The

Table 5 Overall error rates (%) provided by systems designed by the Brazilian database

System	Type	No. of templates		FRR	FAR			AER
		Training	Verification		Random	Simple	Simulated	
1. Santos *et al.* [16]	WI	—	5	10.33	4.41	1.67	15.67	8.02
2. Bertolini *et al.* [4]	WI	—	15	11.32	4.32	3.00	6.48	6.28
3. Rivard *et al.* [5]	WI	—	15	9.77	0.02	0.32	10.65	5.19
4. Justino *et al.* [3]	WD	30	—	2.17	1.23	3.17	36.57	7.87
5. Batista *et al.* [10]	WD	30	—	9.83	0.00	1.00	20.33	7.79
6. Batista *et al.* [11]	WD	20	—	7.50	0.33	0.50	13.50	5.46
7. proposed	WI mode	—	1	14.36	0.02	0.35	14.24	7.24
	WD mode	30	—	7.83	0.016	0.17	13.50	5.38

Table 6 Overall error rates (%) provided by systems designed by the GPDS database (*g* means genuine and *f* means forgery)

System	Type	No. of templates		FRR	FAR			AER
		Training	Verification		Random	Simulated	Average	
1. Kumar *et al.* [18]	WI	*f*	1	13.76	—	13.76	—	13.76
2. Ferrer *et al.* [25]	WD	12*g* + 12*f*	—	14.10	—	12.60	—	13.35
3. Vargas *et al.* [26]	WD	12*g* + 12*f*	—	10.01	—	—	14.66	12.33
4. Solar *et al.* [27]	WD	12*g* + 12*f*	—	16.40	—	—	14.20	15.30
5. Batista *et al.* [11]	WD	12*g*	—	19.19	9.81	47.25	—	25.42
		12*g*	—	16.81	–	16.88	—	16.84
6. proposed	WI mode	—	1	26.42	0.0056	27.04	—	17.82
	WD mode	12*g*	—	27.25	0.0031	18.17	—	15.24
		14*g*	—	18.06	0.0031	22.71	—	13.96

first three systems are WI systems, while the last three are WD systems. The WI systems do not use user signature templates for training; however, an independent (development) signature database is used. It is clear that system 2 outperforms system 1 as it applied information fusion on the decision level, instead of the single classifier in system 1. Also, system 3, that applied information fusion on both the feature and decision levels, outperforms system 2 (both systems applied majority vote of decisions based on 15 templates). The proposed system, when employed in the WI mode and with using a single template for verification showed comparable performance of the pure WI-SV systems.

When employed in the more secure WD mode, our system showed similar performance as system 3 (the baseline system of our work), while only single classification decision is executed, instead of fusing 15 classification decisions in the baseline system. Comparing with the WD systems, system 6 has best performance among the other WD systems. Although this system executes a complex dynamic selection of classifiers, the proposed system showed similar accuracy with a single classification operation. For the Brazilian database, the actual accuracy of our system ranges between 5.38 and 7.24%, based on the point of switching between the WI and WD operational modes.

5.2.2 GPDS database: Table 6 compares the proposed system to some pure WI and WD systems that are tested on the GPDS database. Only the first system is a WI system, whereas the other systems are WD. Although the WI system presented in [18] does not use signature templates of real system users, it used forgery signatures (from the development dataset) in training. For our system, we did not use forgery signatures for training. For systems 2, 3 and 4, genuine and forgery samples (of the real system users) are used for training and/or for selecting optimal decision thresholds. This scenario may bias the reported system accuracy, since forgeries are not available during the design of a real-world SV system. System 5 applied similar experimental protocol to ours, where no forgeries are considered available for training. This system showed comparable performance as that of the aforementioned WD system, despite that it did not use forgeries for training. So, system 5 outperforms the earlier systems; however, it applies complex generative-discriminative system with dynamic selection of classifiers. Also, for systems 1, 2 and 5 (see second row), the equal error rate (ERR) is reported. In this case the decision threshold is selected to produce equal values for FRR and $FAR_{simulated}$. Our system showed AER comparable with that of system 5, where the threshold selection is employed based on an independent (development) database. For the GPDS database, the actual accuracy of our system ranges between 13.96 and 17.82%, based on the point of switching between the WI and WD operational modes.

6 Conclusions and future work

A solution to compromise between pure WD and WI SV systems is proposed. A universal WI classifier is designed with a development database, to enable starting system operation with few signature templates. Switching to a more secure, less complex, and more accurate WD operational mode is possible whenever enough samples are collected for a specific user. Adaptation of the WI classifiers to specific users is achieved through tuning the universal signature representation to each user, while training his WD classifier.

Simulation results on two real-world offline signature databases confirm the feasibility and robustness of the proposed approach. The initial universal (WI) verification mode showed comparable performance to that of state-of-the-art offline SV systems. The final user-specific WD verification mode showed enhanced accuracy with decreased computational complexity. Only a single compact classifier produced similar level of accuracy (AER of about 5.38 and 13.96% for the Brazilian and the GPDS databases, respectively) as complex WI and WD systems in literature. In addition, the produced WD classifiers are more secure than the baseline WI classifiers, eliminating the need to store user templates for verification.

Future work will investigate the ability to enhance the system accuracy by employing other features, and learn from independent forgeries, during the WI training. Also, user adaptation of classifier parameters (such as decision thresholds, image size normalisation and training size), will be investigated. Fusion of both WI and WD system modes will be investigated, in order to enhance the performance during the initial operational period of the system. In this paper, the user samples are assumed to be collected offline. A more practical scenario, where authenticated samples are used to train the WD classifiers online, will be investigated.

7 Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada and BancTec Inc.

8 References

1 Impedovo, D., Pirlo, G.: 'Automatic signature verification: the state of the art', *IEEE Trans. Syst. Man Cybern. C, Appl. Rev.*, 2008, **38**, (5), pp. 609–635

doi: 10.1049/iet-bmt.2013.0024

2 Batista, L., Rivard, D., Sabourin, R., Granger, E., Maupin, P.: 'State of the art in off-line signature verification', in: Verma, B., Blumenstein, M. (eds.): 'Pattern recognition technologies and applications: recent advances' (IGI Global, Hershey, 2007, 1st edn.) pp. 39–62

3 Justino, E., Bortolozzi, F., Sabourin, R.: 'Off-line signature verification using HMM for random, simple and skilled forgeries'. Sixth Int. Conf. on Document Analysis and Recognition (ICDAR'2001), Seattle (Washington, USA), 11–13 September 2001, pp. 1031–1034

4 Bertolini, D., Oliveira, L., Justino, E., Sabourin, R.: 'Reducing forgeries in writer-independent off-line signature verification through ensemble of classifiers', *Pattern Recognit.*, 2010, **43**, (1), pp. 387–396

5 Rivard, D., Granger, E., Sabourin, R.: 'Multi-feature extraction and selection in writer-independent offline signature verification', *Int. J. Doc. Anal. Recognit.*, 2013, **16**, (1), pp. 83–103

6 Tieu, K., Viola, P.: 'Boosting image retrieval', *Int. J. Comput. Vis.*, 2004, **56**, (1), pp. 17–36

7 Eskander, G.S., Sabourin, R., Granger, E.: 'Adaptation of writer-independent systems for offline signature verification'. The 13th Int. Conf. Frontiers in Handwriting Recognition (ICFHR-2012), Bari, Italy, 18–20 September, pp. 432–437

8 Vargas, J., Ferrer, M., Travieso, C., Alonso, J.: 'Off-line handwritten signature GPDS-960 corpus'. Int. Conf. Document Analysis and Recognition, 2007, pp. 764–768

9 Freitas, C., Morita, M., Oliveira, L., *et al*.: 'Bases de dados de cheques bancarios brasileiros'. XXVI Conf. on Latinoamericana de Informatica (CLEI'2000), Atizapan de Zaragosa (Mexico), 18–22 September 2000, CD 1. (in portuguese)

10 Batista, L., Granger, E., Sabourin, R.: 'Improving performance of HMM-based off-line signature verification systems through a multihypothesis approach', *Int. J. Doc. Anal. Recognit.*, 2010, **13**, pp. 33–47

11 Batista, L., Granger, E., Sabourin, R.: 'Dynamic selection of generative-discriminative ensembles for off-line signature verification', *Pattern Recognit.*, 2012, **45**, (4), pp. 1326–1340

12 Pekalska, E., Duin, R.P.W.: 'Dissimilarity representations allow for building good classifiers', *Pattern Recognit. Lett.*, 2002, **23**, (8), pp. 161–166

13 Duin, R.P.W., Loog, M., Pekalska, E., Tax, D.M.J.: 'Feature-based dissimilarity space classification'. Proc. 20th Int. Conf. on Recognizing Patterns in Signals, Speech, Images, and Videos (ICPR'10), Istanbul, Turkey, 23–26 August, 2010, pp. 46–55

14 Cha, S.: 'Use of distance measures in handwriting analysis'. PhD thesis, State University of New York at Buffalo, USA, 2001

15 Srihari, S.N., Xu, A., Kalera, M.K.: 'Learning strategies and classification methods for off-line signature verification'. Proc. Ninth Int. Workshop on Frontiers in Handwriting Recognition (IWFHR'04), Tokyo, Japan, 26–29 October 2004, pp. 161–166

16 Santos, C., Justino, E., Bortolozzi, F., Sabourin, R.: 'An off-line signature verification method based on document questioned experts approach and a neural network classifier'. Proc. Ninth IWFHR Int. Workshop on Frontiers in Handwriting Recognition (IWFHR'04), Tokyo, Japan, 26–29 October 2004, pp. 498–502

17 Oliveira, L., Justino, E., Sabourin, R.: 'Off-line signature using writer-independent approach'. Int. Joint Conf. on Neural Networks (IJCNN'07), Orlando, FL, USA, 12–17 August, pp. 2539–2544

18 Kumar, R., Sharma, J., Chanda, B.: 'Writer-independent off-line signature verification using surroundedness feature', *Pattern Recognit. Lett.*, 2012, **0**, pp. 301–308

19 Sabourin, R., Genest, G.: 'An extended-shadow-code based approach for off-line signature verification'. Proc. 12th Int. Conf. on Pattern Recognition, Jerusalem, 1994, vol. 2, pp. 450–453

20 Drouhard, J., Sabourin, R., Godbout, M.: 'A neural network approach to off-line signature verification using directional PDF', *Pattern Recognit.*, 1996, **29**, (3), pp. 415–424

21 Schapire, R.: 'The boosting approach to machine learning: an overview'. Proc. MSRI Workshop on Nonlinear Estimation and Classification, Berkeley, California, 19–29 March, 2001

22 Iba, W., Langley, P.: 'Induction of one-level decision trees'. Proc. Ninth Int. Machine Learning Conf., Aberdeen, Scotland, 1992, pp. 233–240

23 Last, M., Bunke, H., Kandel, A.: 'A feature-based serial approach to classifier combination', *Pattern Anal. Appl.*, 2002, **5**, pp. 385–398

24 Khreich, W., Granger, E., Miri, A., Sabourin, R.: 'Iterative Boolean combination of classifiers in the ROC space: an application to anomaly detection with HMMs', *Pattern Recognit.*, 2010, **31**, pp. 2732–2752

25 Ferrer, M.A., Alonso, J.B., Travieso, C.M.: 'Offline geometric parameters for automatic signature verification using fixed-point arithmetic', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005, **27**, (6), pp. 993–997

26 Vargas, J.F., Ferrer, M.A., Travieso, C.M., Alsono, J.B.: 'Offline signature verification based on high pressure polar distribution'. Proc. 11th Int. Conf. Frontiers in Handwriting Recognition (ICFHR'2008), 19–21 August 2008, pp. 373–378

27 Solar, J.R., Devia, C., Loncomilla, P., Concha, F.: 'Off-line signature verification using local interest points and descriptors', *Lect. Notes Computer Sci.*, 2008, **5197**, pp. 22–29