Adaptation of Writer-Independent Systems for Offline Signature Verification

George S. Eskander, Robert Sabourin and Eric Granger
Laboratoire d'imagerie, de vision et d'intelligence artificielle
École de technologie supérieure, Université du Québec
Montreal, Canada
geskander@livia.etsmtl.ca Robert.Sabourin@etsmtl.ca Eric.Granger@etsmtl.ca

***Abstract*— Although writer-independent offline signature verification (WI-SV) systems may provide a high level of accuracy, they are not secure due to the need to store user templates for authentication. Moreover, state-of-the-art writer-dependent (WD) and writer-independent (WI) systems provide enhanced accuracy through information fusion at either feature, score or decision levels, but they increase computational complexity. In this paper, a method for adapting WI-SV systems to different users is proposed, leading to secure and compact WD-SV systems. Feature representations embedded within WI classifiers are extracted and tuned to each enrolled user while building a user-specific classifier. Simulation results on the Brazilian signature database indicate that the proposed method yields WD classifiers that provide the same level of accuracy as that of the baseline WI classifiers (AER of about 5.38), while reducing complexity by about 99.5%.**

***Keywords*- Offline signature verification, writer-dependent, writer-independent, writer-adaptation, dissimilarity representation, boosting feature selection.**

I. INTRODUCTION

Off-line signature verification (SV) systems use digitized handwritten signature images for authentication. While on-line systems exploit signature trajectory dynamics, offline systems depend on less discriminant static signature representations [13]. One way to increase the discrimination power of signature images is to apply information fusion at the feature, score, and decision levels [17]. In the feature level fusion, different feature representations are fused prior to the classification process. On the other hand, for the decision and score level fusion, responses from several classifiers are combined. These methods, however, add complexity to the classification systems.

In literature, two main approaches for SV are applied for information fusion at its various levels. The first approach is called writer-independent (WI), where a global classifier is designed using a development database [4]. The classifier is then used to authenticate any user, by comparing the query sample to its stored templates. To achieve a high level of accuracy with WI systems, feature level fusion is applied and produces high-dimensional feature representations. Moreover, multiple classification decisions are combined [6]. The second approach is called writer-dependent (WD), where an individual classifier is designed for each user using his samples [14]. This type is more secure, as no templates are stored for verification. However, the shortage of training samples limits the classifier accuracy. State-of-the-art WD systems have applied ensemble methods for enhanced performance at the expense of significantly increased complexity [13], [15].

In this paper, a hybrid WI-WD scheme is proposed to exploit the advantages of both approaches while alleviating their drawbacks. This scheme consists of two steps: (1) a WI classifier is designed using Boosting Feature Selection (BFS) [6], and (2) WD classifiers are designed in the feature space defined by BFS. The features embedded in the WI classifier constitutes a universal signature representation, as it can represent all users. To adapt this representation to a specific system user, samples are collected from enrolled user and represented in this universal feature space. These representations are then used to train a WD classifier using the BFS approach, thereby producing a more compact and secure classification system.

To validate this hypothesis, simulations are conducted using the real-world Brazilian signature database [12] that includes random, simple, and skilled forgeries. The next section provides an overview of state-of-the-art offline SV systems, focusing on the baseline WI-SV system used in this work. Section III describes the proposed WI-WD hybrid approach. Section IV describes the experimental methodology applied in this paper. The experimental results are presented and discussed in Section V.

II. OFFLINE SIGNATURE VERIFICATION SYSTEMS

For WD systems, the class distribution of a specific user is modeled using his signature samples. Performance of these systems is limited by the available samples for training. Enhanced recognition rates of WD systems is recently achieved by training multi-classifier systems [15], [16]. On the other hand, WI systems do not model the individual distributions, but rather a universal distribution 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 dissimilarity concept is first introduced by Pekalska et al. [1], and applied to the author identification domain by Cha and Srihari [2].

In this context, the WI approach is realized using a dissimilarity (distance) measure, to compare samples (query and reference samples) as belonging to either the same or different user. The dissimilarity measure allows to represent samples in a multi-dimensional space called dissimilarity space.

Santos et al. [3] applied this concept by using the Euclidian distance between ordinary feature vectors, extracted from both the questioned and reference signatures, as a dissimilarity representation. The number of resulting distance samples is greater than the original samples, hence facilitate learning in a high dimensional space with larger training set. A neural network is then trained to find the optimal boundary that splits the genuine and forgery classes in the dissimilarity space. Later, Oliveira et al. [4], and Bertolini et al. [5] 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 ROC space. Kumar et al. [18] proposed a WI-SV based on surroundedness features.

More recently, Rivard et al. [6] extended the system in [5] 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 the dissimilarity space results in dissimilarity representation of high dimensionality. This complex representation is then simplified by applying the boosting feature selection approach (BFS) [7]. By applying the multi-feature approach with BFS, it is possible to design WI systems with higher perfromance than the ealier implementations. Moreover, the complex dissimilarity represenation (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.

Current SV techniques with acceptable accuracy are complex due to the fusion of responses from multiple classifiers. Morover, they might be insecure due to the need to store reference signatures for verification. The work proposed in this paper extends on the system in [6] by adapting the universal representation to each specific user, with the aim of reducing the classfication complexity (number of features and number of classifications fused for a decision), and avoiding the need of using reference signatures for verification.

III. A Hybrid WI-WD Approach

Figure 1 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. [6]. To adapt this system to specific users, features embedded in the designed WI-SV are considered as a universal (population-based) represenation. Then, user and forgery samples are translated into this universal space and used to train the WD-SV sub-system.

A. WI Feature Selection

In literature, many types of features could be extracted from offline signature images [13]. 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) [10], and Directional Probability Density Function (DPDF) [11]. 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 spatial information, the DPDF detects the directional information from signature images [6].

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 $\mathbf{M}=(m_1, m_2, \ldots, m_K)$, and K is the dimensionality of the multi-feature representation. To project these samples into the dissimilarity space, dichotomy transformation is applied. For instance, for two samples M_i, M_j, the dissimilarity feature is:

$$DR_{ij} = |(M_i - M_j)| = (\Delta m_1, \Delta m_2, \ldots, \Delta m_K). \qquad (1)$$

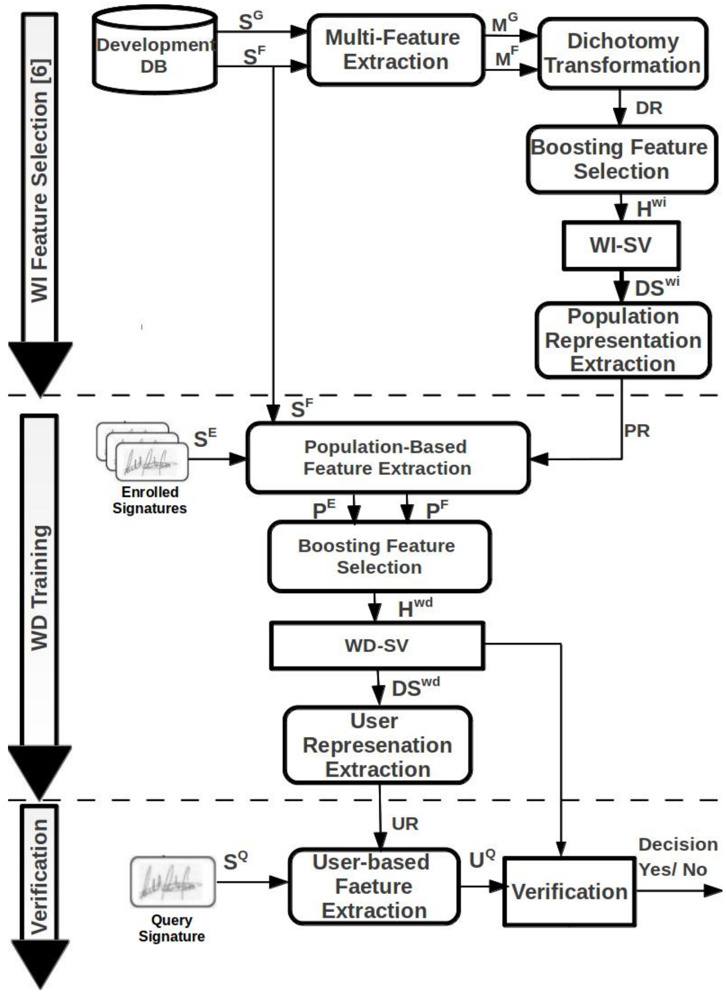
where $\Delta m_k = |(m_{ik} - m_{jk})|$. It is worth noting that both the multi-feature and dissimilarity representations have the same dimensionality K. Also, a sample DR_{ij} is labeled as a within-class or as a between-class instance, when it results from two genuine signatures of the same user, or from two signatures of two different users, respectively.

To build the WI-SV system, the BFS approach is applied [7]. This method applies Gentle AdaBoost algorithm [8] to learn an optimal decision boundary between the within-class and the between-class dissimilarity samples, by boosting Decision Stump (DS) weak learners [9]. At a boosting iteration t, a DS is designed by locating the best dimension d_t in the dissimilarity space that splits the training samples based on a splitting threshold T_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:

$$DS_t = \begin{cases} \rho_t^{left} & if \quad d_t < T_t \\ \rho_t^{right} & otherwise \end{cases}. \qquad (2)$$

where ρ_t^{left}, ρ_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:

Figure 1. Block diagram of the proposed Hybrid WI-WD approach. A WI-SV sub-system is designed as proposed by Rivard et al. [6]. To adapt this system to specific users, features embedded in the designed WI-SV are considered as a universal population-based representation. Then, user and forgery samples are translated into this universal space and used to train the WD-SV sub-system.



$$H^{wi}=\sum_{t}^{T^{wi}} DS_t^{wi}. \qquad (3)$$

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. Refer to [6] for more detailed algorithms of the WI-SV design process.

In this baseline system, a questioned signature S^Q is represented in dissimilarity space as D^Q, and be classified by the WI-SV system, where

$$WI-SV\left(D^Q\right)=sign\left(\sum_{t}^{T^{wi}} DS_t^{wi}\left(D^Q\right)\right). \qquad (4)$$

In the proposed approach, the WI-SV is not used directly for signature verification, but only for dimensionality reduction through WI feature selection. The feature representation embedded in a WI classifier is extracted and stored as a population-based representation *(PR)* 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.

B. WD Training

Although the universal *PR* 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 (to enlarge the training set and hence facilitate learning in a high dimensional space), the WD training phase, on the other hand, could be performed in either the dissimilarity or the original feature space.

Operating the SV verification 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 (*PR*) of dimensionality L is used for feature extraction. For each enrolled user, sample signatures are collected. Both the enrolling samples S^E and some samples S^F are selected from the development DB (to represent the random forgery class), are represented in the *PR* feature space as P^G and P^F respectively. Finally, the same BFS process is applied, by using this WD data to model the decision boundary H^{wd} that splits the genuine and forgery classes, where

$$H^{wd} = \sum_{t}^{T^{wd}} DS_t^{wd}. \quad (5)$$

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.

C. Verification

To authenticate a specific user, 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 (*UR*) of dimensionality $N<L<K$. Then, the query image S^Q is represented in this concise representation space as U^Q, and then fed the classifier for recognition, where

$$WD-SV\left(U^Q\right) = sign\left(\sum_{t}^{T^{wd}} DS_t^{wd}\left(U^Q\right)\right). \quad (6)$$

It is worth noting that, 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.

IV. EXPERIMENTAL METHODOLOGY

The Brazilian signatures DB [12] is used in this experimentation. It contains signatures of 168 users, that were digitized as 8-bit grayscale images over 400×1000 pixels, at resolution of 300 dpi. It is split into two parts. The first part contains signatures of the last 108 users, and is used as a development DB for the WI feature selection phase. The second part contains signatures of the first 60 users, and used for the WD training, and for performance evaluation.

Table I describes the development dataset used for WI feature selection. A total of 93,960 samples are used for training, and 64,800 are used as holdout validation set to avoid overfitting. Multi-feature representations of signature images are produced by different grid scales. The dimensionality of the resulted multi-feature vector K=30,201 [6].

TABLE I. EVELOPMENT DATABASE: (108 USERS X 40 GENUINE SIGNATURES EACH)

Training set (30 signatures/user)		**Validation set** (10 signatures/user)	
Within-Class distances among the 30 signatures / user.	**Between-Class** distances among 29 signature/user and 15 signatures of other users.	**Within-Class** distances among the 10 signatures/ user and the 30 signatures of the training set.	**Between-Class** distances among the 10 signatures/ user and 30 signatures selected randomly from other users.
108x30x29/2= 46,980 samples	108x29x15= 46,980 samples	108x10x30= 32,400 samples	108x10x30= 32,400 samples

TABLE II. WD DATABASE: (60 USERS X 60 SIGNATURES EACH: 40 GENUINE+10 SIMPLE FORGERY+10 SIMULATED FORGERY)

Training set (30 genuine signatures/user)		**Testing set** (40 signatures/user)	
Genuine-class Signature subsets of different sizes.	**Forgery-class** Signatures of the training set of the development DB.	**Genuine-class** 10 genuine signatures/user.	**Forgery-class** 10 simple +10 simulated + 10 random forgeries selected randomly from other users.
5,7,9,11,13,15,30 samples	108x30= 3240 samples	60x10= 600 samples	60x30= 1,800 samples

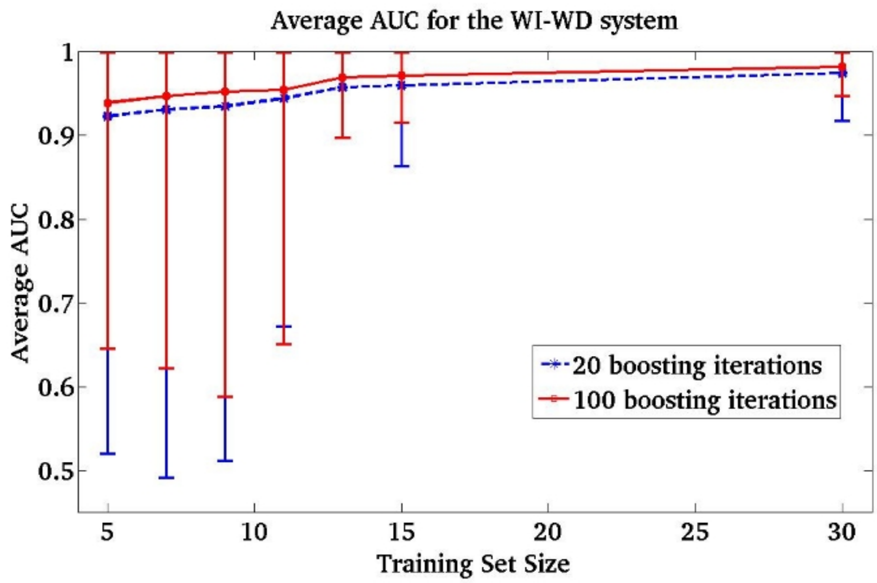
The BFS algorithm is set for 1000 max boosting iterations, and 100 early stopping criterion. The resulted WI-SV classifier contained 679 decision stumps, with them only 555 distinct features are used. (i.e., T^{wi} =679, L= 555). Then, the WI-SV is used to extract the population-based representation (*PR*) of dimension L=555. The WD training is executed in the *PR* space. To this end, a WD-DB consisting of 60 users is used to generate the WD training set, and the testing set.

Table II describes the data sets used to build the WD classifiers, and for performance evaluation. For the WD training, fixed number of boosting iteration T^{wd} is used for early stopping. To investigate the impact of training samples quantity on the recognition performance, different number of samples are used to train the WD classifier. The forgery class is represented by genuine signatures from the development DB. In each WD training run, genuine and forgery samples are represented in the *PR* space and used for training. We observed saturation in performance around 100 boosting iterations, so the boosting iterations was set to a fixed number (here, the performance is reported for two cases where, T^{wd}=20, and 100).

For performance evaluation, 40 test samples per user are employed. Of them, 10 genuine, 10 random, 10 simple, and 10 simulated forgeries, for a total of 2400 questioned signatures are employed for system evaluation.

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 40 questioned signatures S^Q per user in the test set are processed by a classifier. Its outputs are then sorted, and used as a set

Figure 2. Average AUC of ROC curves for WD-SV classifiers designed with different training set sizes and with different number of boosting iteration. 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. The classifier performance increases when increasing both training set size and boosting iterations.



of classifier thresholds. Then, the GAR (Genuine Accept Rate) and FAR (False Accept Rate) 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 of classifiers are averaged over 60 users. The Average Error Rate (AER) is computed as follows:

$$AER=\left(FRR+FAR_{random}+FAR_{simple}+FAR_{simulated}\right)/4 \qquad (7)$$

where FRR is the False Rejection Rate, and FAR_{random}, FAR_{simple}, and $FAR_{simulated}$ are the False Accept Rates when verifying random, simple, and simulated forgeries, respectively.

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 [19], where

$$TFV=\sum_{i=1}^{n} m_i x_i \qquad (8)$$

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.

V. Simulation Results

Figure 2. shows the a AUC of the WD-SV system when designed with different number of boosting iterations, and with different training set sizes. 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. It is shown that with only 5 training samples, and only 20 boosting iterations, the average AUC 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). Table III compares the proposed WI-WD system to some pure WI and WD systems in literature. All systems are investigated using the same data set and testing protocol, and results are reported in terms of AER. The first 3 systems are WI systems, while the last 3 are WD systems. 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).

TABLE III. Overall error rates (%) provided by systems designed with the Brazilian database.

System	Type	FRR	FAR			AER
			Random	Simple	Simulated	
1. Santos [3]	WI	10.33	4.41	1.67	15.67	8.02
2. Bertolini [5]	WI	11.32	4.32	3.00	6.48	6.28
3. Rivard [6]	WI	9.77	0.02	0.32	10.65	5.19
4. Justino [14]	WD	2.17	1.23	3.17	36.57	7.87
5. Batista [16]	WD	9.83	0.00	1.00	20.33	7.79
6. Batista [15]	WD	7.50	0.33	0.50	13.50	5.46
7.proposed	**WI-WD**	**7.83**	**0.016**	**0.17**	**13.50**	**5.38**

The proposed hybrid WI-WD 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 WI-WD system showed similar accuracy with a single classification operation.

Regarding to the computational complexity, the baseline WI-SV fused 15 partial classification decisions. Each decision is based on processing of 555 features extracted from a query sample and a template. Hence, the *TFV* is *8325*. Adaptation of this WI system to different users produced WD classifiers, that take the classification decision 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, the proposed approach reduces the computational complexity by about 99.5%.

VI. CONCLUSIONS

The adaptation of the writer-independent (WI) classifiers to each specific user is proposed in this paper. The original WI classifier is used to produce a universal signature representation through BFS that is valid for all users. Signature samples collected during enrolment to adapt this universal representation to this specific user, while training his writer-dependent (WD) classifier.

Simulation results confirm the feasibility of the proposed approach, since it decreased the classifier computational complexity by about 99.5%. Only a single compact classifier produced similar level of accuracy (AER of about 5.38) as complex WI and WD systems in literature.

In addition, the proposed WI-WD system is more secure than the baseline WI classifier, eliminating the need to store user templates for verification. Future work will investigate the ability to recognize simulated forgeries by using simulated forgery samples, and by employing other features during the WI training.

ACKNOWLEDGMENT

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

REFERENCES

[1] Pekalska and R.Duin, “Dissimilarity representations allow for building good classifiers,” PR 23(2002): 943–956.

[2] S. Cha, S. Srihari. “On measuring the distance between histograms,” PR 35 (2002): 355–1370.

[3] C. Santos, E. Justino, F. Bortolozzi, R. Sabourin, “An off-line signature verification method based on document questioned experts approach and a neural network classifier,” 9^{Th} IWFHR, pp. 498–502, 2004.

[4] L. Oliveira, E. Justino, R. Sabourin, “Off-line signature using writer-independent approach,” IJCNN, 2007, pp. 2539–2544.

[5] D. Bertolini, L. Oliveira, E. Justino, R.Sabourin. “Reducing forgeries in writer-independent off-line signature verification through ensemble of classifiers,” PR 43(1): 387–396 (2010).

[6] D. Rivard, E.Granger, R.Sabourin, "Multi-feature extraction and selection in writer-independent off-line signature verification," IJDAR(25):1-21, 2012, doi:10.1007/s10032-011-0180-6.

[7] K. Tieu, and P. Viola, “Boosting image retrieval,” Int. J. Comput. Vis. 56(1): 17–36, 2004.

[8] R. Schapire, “The Boosting Approach to Machine Learning: An Overview,” chap. 8. Springer, 2003.

[9] W. Iba, P. Langley, “Induction of one-level decision trees,” Proceedings of the Ninth IMLC. 1992.

[10] R. Sabourin, and G. Genest, “An extended-shadow-code based approach for off-line signature verification,” Proceedings of the third ICDAR(1): 197–201, 1995.

[11] J. Drouhard, R. Sabourin, M. Godbout, “A neural network approach to off-line signature verification using directional pdf,” PR 29(3): 415–424, 1996.

[12] C. Freitas, M. Morita, L. Oliveira, E. Justino, et al, “Bases de dados de cheques bancarios brasileiros,” XXVI Conferencia Latinoamericana de Informatica, 2000.

[13] D. Impedovo, and G. Pirlo, “Automatic signature verification: the state of the art,” IEEE Trans. SMCA 38(5): 609–635, 2008.

[14] E. Justino, F. Bortolozzi, R. Sabourin, “Off-line signature verification using HMM for random, simple and skilled forgeries,” ICDAR, 105–110, 2001.

[15] L. Batista, E. Granger, R. Sabourin, “Dynamic Selection of Generative-Discriminative Ensembles for Off-Line Signature Verification," PR(45) 4:1326-1340,2012.

[16] L. Batista, E. Granger, R. Sabourin, “Improving performance of HMM-based off-line signature verification systems through a multi-hypothesis approach,” IJDAR(13):33–47, 2010.

[17] A. Jain, A. Ross, S. Pankanti, “Biometrics: A Tool for Information Security,” IEEE TIFS.1(2):125-143, 2006.

[18] R. Kumar, J. Sharma, B. Chanda, “Writer-independent off-line signature verification using surroundedness feature,” PR Letters (2012): 301-308.

[19] M. Last, H. Bunke, and A. Kandel, “ A feature-based serial approach to classifier combination,“ Pattern Analysis and Applications. (5): 385–398, 2002.