Dissimilarity Representation for Handwritten Signature Verification

George S. Eskander, Robert Sabourin, and Eric Granger
Laboratoire d'imagerie, de vision et d'intelligence artificielle
Ecole de technologie supérieure, Université du Québec
Montréal, QC, Canada
Email: geskander@livia.etsmtl.ca, robert.sabourin@etsmtl.ca, eric.granger@etsmtl.ca

***Abstract*—Signature verification (SV) systems authenticate individuals, based on their handwritten signatures. The standard approach for such systems employ feature representations (FR), where features are extracted from the signature signals and classifiers are designed in the feature space. Performance of FR-based systems is limited by the quality of employed feature representations and the quantity of training data. The dissimilarity representation (DR) approach is recently introduced to pattern recognition community, where proximity among patterns constitute the classification space. Similar concept has been applied by forensic Questioned Document Examination (QDE) experts, where proximity between questioned signatures and a set of templates lead to the authentication decision. Recently, few automatic SV systems are proposed to simulate the QDE approach, by employing DR-based pattern recognition methods. In this paper, we explore different scenarios for employing the DR approach for replacing and/or enhancing the standard SV systems. A general framework for designing FR/DR based systems is proposed, that might guide the signature processing research direction to new areas.**

I. Introduction

Signature Verification (SV) systems verify that a signature sample belongs to a specific writer. Signature signals can be acquired either online or offline. For online systems, signature dynamics such as velocity, pressure, stroke order, etc., are acquired during the signing process. Special pens and tablets are employed for the online acquisition task. On the other hand, for offline systems, signature images are scanned, after the signing process. Only static information are extracted from the signature images, producing a harder pattern recognition problem [1].

Standard SV systems employ feature-based pattern recognition approaches. Discriminative features are extracted from the signature signals, so that each signature is represented as a vector in the Feature Representation (FR) space. The classifiers are then designed in the feature space. Simply, accuracy of such systems relies on to which extend the employed feature representation is discriminative and stable. Signature representations of different users may have high similarities, when features are not discriminative enough. Also, representations of the same writer may differ significantly, when features are not stable. Besides quality of features, enough training data is required to design reliable classifiers in the feature space. The training samples should represent a wide range of genuine signatures and possible forgeries, for all system users. For real world applications, e.g., banking systems, the number of users could be very high and there is a high risk of forgery. The enrolling signature samples, available for designing such systems, are mostly few and no samples of forgeries are available. With these limitations, it is a challenge to extract informative feature representations and to design feature-based classifiers, that absorb the intra-personal variabilities while detecting both the forgeries and the inter-personal similarities.

The Dissimilarity Representation (DR) approach for pattern recognition is recently introduced, by Elzbieta Pekalska and Robert P.W. Duin., [2]. The rational behind this concept is that modeling the proximity between objects may be more tractable than modeling the objects themselves. To this end, dissimilarity measures are computed and considered as features for classification. The dissimilarity measures can be derived in many ways, e.g. from raw (sensor) measurements, histograms, strings or graphs. However, it can also be build on top of a feature representation [3].

In the field of forensic science, similar concept has been applied by the Questioned Document Examination (QDE) experts. A questioned handwritten sample is associated to a specific writer, if it is similar to a set of reference templates of his handwritings. Degree of similarity is determined by comparing a set of graphonomic features, extracted from both the questioned and template samples.

Recently, some automatic SV systems are proposed to simulate the QDE approach [4]-[7]. Distances between intra-personal training samples are computed and used as intra-class samples. Similarly, distances between inter-personal training samples are computed and used as inter-class samples. The produced distance samples are used to train a single two-class classifier, that distinguishes between intra-class and inter-class distances.

The DR approach, besides it enabled automating the forensic expert manual tasks, it alleviates some of the limitations of the FR-based design approach. First, a distance sample is generated for every pair of the original training samples, so it results in a much higher number of samples. This property alleviates the shortage of training data required to model the signatures. Second, dissimilarities between signature signals maybe more discriminative and stable than the feature representations. This is why the QDE experts build their decisions on the dissimilarity between questioned and template samples, and not on the absolute measurements of the questioned sample. Finally, the DR-approach could be applied to develop global classifiers, that are valid for all current and future users. This concept is known as Writer-Independent (WI) systems, developed by Siteargur N. Srihari et al., [8], and Santos and Sabourin et al., [4]. Instead of building a single writer-dependent (WD) classifier for each user using his enrolling signatures, a single global classifier is designed by learning the dissimilarities between signatures of all users. The rational behind the WI approach is: while it is impossible to model a feature-based class distribution that is valid for current and future users, the statistical models for inter-sample distances are generic and can be generalized for users whose signature samples are not used for training.

In this paper, we argue that the DR approach can be applied in different scenarios, in order to design more robust classifiers. It can enable the design of new family of classification systems, such as global and hybrid global/user-specific classifiers. Also, the DR approach can be employed, as an intermediate design tool, for enhanced performance of standard feature-based systems.

In the next section, the DR approach is illustrated, and a general framework for designing FR/DR based systems is proposed. Section III surveys the existing implementations of the DR approach to the offline signature classification area, and relates them to the proposed framework. Section IV discusses possible directions and areas where the DR approach can be applied.

II. General Framework For Dissimilarity-Based Classification

Although the DR is a general approach, where dissimilarity measures can be derived directly from patterns, e.g., raw (sensor) measurements, graphs, etc., we discuss here a special case where the DR is build on top of a feature representation (FR). This approach is suitable for the offline signature classification task, as many techniques of feature extraction are already proposed [1].

Figure 1 illustrates a DR constituted on top of a FR. Assume a system is designed for M different users, where for any user m there are R prototypes (templates) $\{p_{mr}\}_{r=1}^{R}$. Also, a user n provides a set of J questioned signature images $\{Q_{nj}\}_{j=1}^{J}$. The dissimilarity between a questioned sample Q_{nj} and a prototype p_{mr} is $D^{Q_{nj}p_{mr}}$. In case that questioned and prototype samples belong to the same person, i.e., $n = m$, the dissimilarity sample is an intra-personal sample (black cells in Figure 1). On the other hand, if questioned and prototype samples belong to different persons, i.e., $n \neq m$, then the dissimilarity sample is an inter-personal sample (white cells in Figure 1).

Perfect dissimilarity representation implies that all of the intra-class distances have zero values, while all of the inter-class distances have large values. This occurs when the employed dissimilarity measure absorbs all of the intra-class variabilities, and detects all of the inter-class similarities.

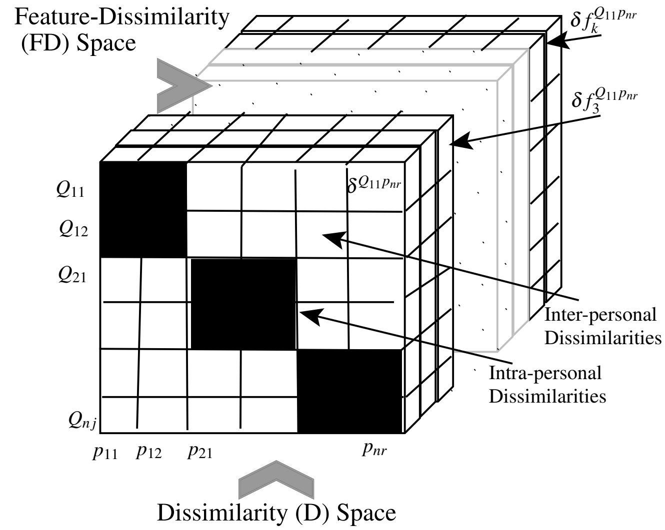

Fig. 1. Illustration of a Dissimilarity Representation (DR) built on top of a feature representation: black and white cells represent intra-personal and inter-personal dissimilarities, respectively. The third dimension represents the Feature Dissimilarity (FD) space, where dissimilarities between prototype and query signatures are measured by the distance between their feature representations. The dissimilarity cells may produce a simple dissimilarity matrix or a Dissimilarity (D) space, where distances to prototypes constitute the space dimensions. Values of FD-space vector elements control the value of corresponding dissimilarity cell.

To design a reliable classifier that works in a DR space, it is not mandatory to achieve a perfect representation, but only a discriminative one. The degree of ease to design a reliable classifier depends on the discriminative power of the representation. Accordingly, it is more important to carefully design the DR, then the classifier design comes in a next step.

In case of the DR is build on top of a FR, quality of the resulting DR relies on the quality of features that constitute the FR, and on the applied dissimilarity measure. For instance, assume the feature representations $F^{Q_{nj}} = \{f_k^{Q_{nj}}\}_{k=1}^{K}$ and $F^{p_{mr}} = \{f_k^{p_{mr}}\}_{k=1}^{K}$, are extracted from the query sample Q_{nj} (from user n) and a prototype p_{mr} (of user m), respectively. Also, consider the Euclidean distance $\delta^{Q_{nj}p_{mr}}$ as a measure of dissimilarity:

$$\delta^{Q_{nj}p_{mr}} = \sqrt{\sum_{k=1}^{K}(\delta f_k)^2}, where\ \delta f_k = \|f_k^{Q_{nj}} - f_k^{p_{mr}}\| \quad (1)$$

It is obvious that, the overall distance between feature representations of the two samples is controlled by the individual feature components, and on the reference prototypes. Accordingly, features and prototypes should be properly selected, in order to minimize the intra-personal dissimilarities and to maximize the inter-personal dissimilarities. Moreover, dissimilarity measures other than the Euclidean distance can be investigated for better dissimilarity representations.

After designing a discriminative representation, classifiers can be designed in the resulting space. Different forms of dissimilarity representation spaces can be employed. More

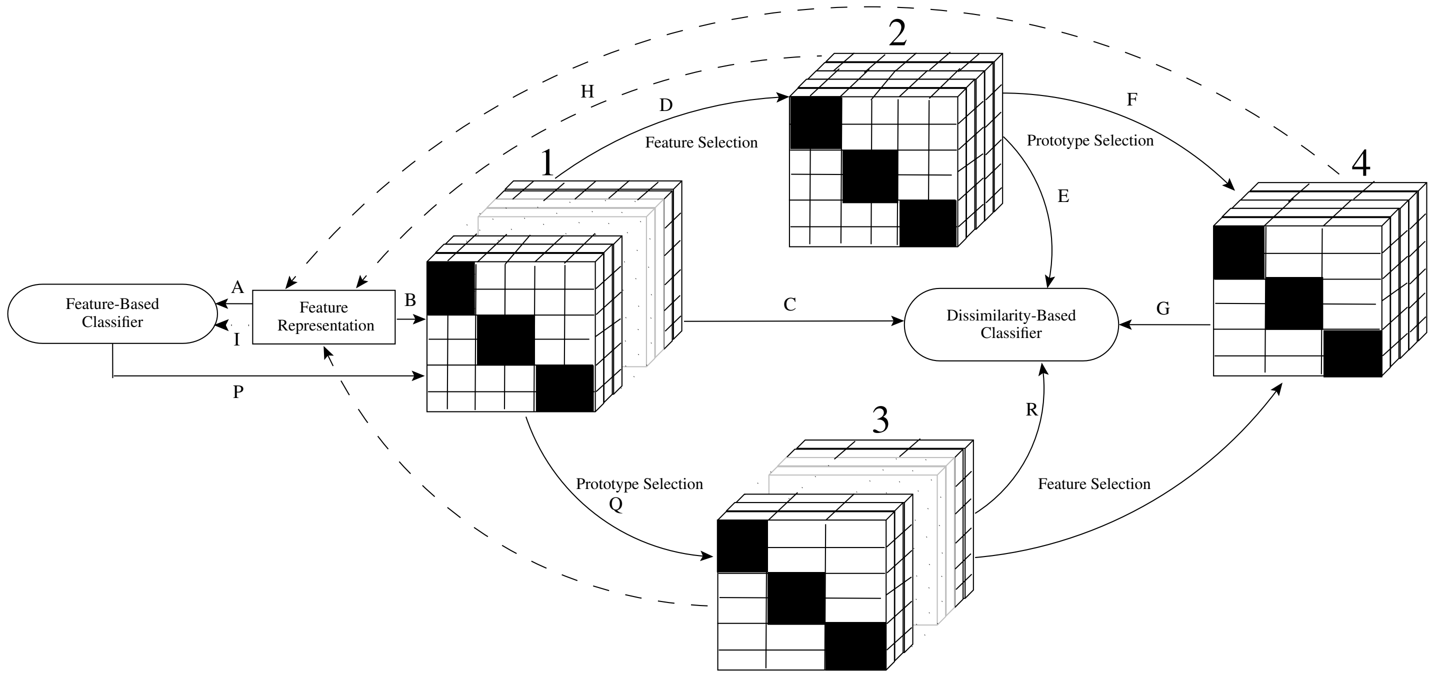

Fig. 2. General framework for designing classification systems based on the Dissimilarity Representation (DR) approach: Block 1–full FD-space-full D-space, Block 2–reduced FD-space-full D-space, Block 3–full FD-space-reduced D-space, Block 4–reduced FD-space-reduced D-space.

specifically, three different forms of dissimilarity representations (DR) can be constituted:

- Dissimilarity matrix: the matrix of all distances, where a row D^{nj} represents distances between a query j that belongs to a specific user n, with respect to the prototypes of all users:

$$D^{nj} = \{\delta^{Q_{nj}p_{11}}, .., \delta^{Q_{nj}p_{mr}}, .., \delta^{Q_{nj}p_{MR}}\}. \quad (2)$$

where $m \in [1, M]$ and $r \in [1, R]$.
- Dissimilarity space (D-Space): the dissimilarity matrix is projected on a space, where each row of the matrix is represented as a vector D^{nj} in this space. By other words, each dimension of the D-space is the distance to a specific prototype.
- Feature-Dissimilarity space (FD-Space): the dissimilarity matrix is embedded in an Euclidean space, where dimensions of this space are the dissimilarities of feature values. In the FD-space, a vector $d^{Q_{nj}p_{mr}}$, has same dimensionality as that of the original feature space, where $d^{Q_{nj}p_{mr}} = \{\delta f_k^{Q_{nj}p_{mr}}\}_{k=1}^{K}$. The length of a vector $d^{Q_{nj}p_{mr}}$ is equivalent to $\delta^{Q_{nj}p_{mr}}$, given by Eq. 1.

We argue that, classifiers can be designed in any of the aforementioned dissimilarity representation spaces. Moreover, the different tasks for feature selection, prototype selection, and classifiers design, can be done in different spaces, whenever translation between spaces is possible. This strategy permits applying a massive number of pattern recognition techniques, with multiple combinations of space transitions. We propose that new techniques for pattern recognition might be developed based on this strategy. In this context, the DR approach is employed either as a tool for enhancing the standard FR-based systems (for feature/prototype selection), or to design reliable dissimilarity-based classification systems (when classifiers are designed in a DR space).

Figure 2 illustrates a general framework for designing classification systems based on the DR approach. The standard approach is to extract feature representations from the training samples, and design classifiers in the feature space (path A in the Figure). However, the DR approach can be employed in different scenarios for either build new family of classifiers in DR-based spaces, or to enhance the performance of standard feature-based classifiers. More specifically, dissimilarities can be computed on top of a feature representation, and are used to constitute different types dissimilarity representations (DR), e.g., dissimilarity matrix, D-space, or FD-space (path B). The resulting representation could be constituted on top of a huge number of feature extractions, and based on large number of prototypes. The intra-personal (black cells) and inter-personal (white cells) dissimilarities, should be discriminative enough in order to design a DR-based classifier (path C). In case that the DR is not enough informative, feature selection and/or prototype selection can be applied for enhanced representation. For instance, feature selection can be employed in a FD-space (path D). In literature, there are many methodologies of feature selection that can be applied to select the most discriminative and stable features. The resulting DR is constituted on top of a sparser feature representation, however, redundancy in prototypes may exist (block 2). A classifier can be then designed in the resulting space (path E), or a prototype selection step is done (path F) producing a more compact representation (block 4). Surely, classifiers designed in the sparse and compact representation, are lighter and more accurate (path G). Also, order of the feature/prototype selection processes can be reversed (see the bottom part of the Figure). It is obvious that, it is more logical to run the feature selection process in the FD-space, however, the D-space is more suitable for prototype selection task. The classifier design process can be implemented based on different DR (dissimilarity matrix, D-space, or FD-space).

Besides that the DR approach can be employed to design dissimilarity-based classifiers, it can be considered as an intermediate tool for building reliable feature-based classifiers. Good features and/or prototypes can be selected in a dissimilarity-based space, then the representation is translated back to a sparser and more informative feature space (dotted paths, like path H-I). On contrary, FR-based classifiers can be designed and they are considered as an intermediate tool, to design reliable DR-based classifiers. In such case, multi-classifier systems can be designed, where FR-based classifiers are used to produce the dissimilarity measures, that are needed to build the DR (path P).

III. Current Implementations to Offline Signature Systems

The first application of the dissimilarity learning to biometrics, and more specifically, to the behavioral handwritten biometrics is proposed by Jain, A.K. et al., [10]. The dissimilarity between handwritten digits is measured by the amount of deformation required to restore a query sample to its stored prototype. This approach is extended to the author identification problem by Cha and Srihari [11], where distance statistics are used for classification. Later, similar concept is applied to the handwritten signature images. Here we list and categorize some of these implementations, and relate them to the proposed framework for DR-based classification shown in Figure 2.

A. Writer-Dependent Systems

The Writer-Dependent (WD) approach seeks to build a single classifier for each user based on his enrolling signatures. The DR concept is first introduced to design WD-SV systems, by Siteargur N. Srihari et al., [8]. Correlation between high dimensional (1024-bits) binary feature vectors, is employed as a dissimilarity measure. For a specific user, distances among every pair of his training samples, are determined to represent the intra-class samples. Also, distances between samples of the specific user and some forgeries are computed to represent the inter-class samples. The authors tried different classification strategies: one-class, two-class, discriminative, and generative classifiers. This implementation is a realization of the path B-C in Figure 2, where classifiers are designed based on the statistics of the dissimilarity matrix.

Later, Batista et al., [13] applied the dissimilarity learning concept to produce reliable WD-SV systems. A feature-based one-class classifier is built by producing user-specific generative models using Hidden Markov models (HMMs). To increase the system accuracy, a two-class discriminative model is build in DR space. The HMMs models are considered as prototypes, and samples are projected to a D-space by considering the likehood to each HMM generative model as a similarity measure. SVM classifies are then designed in the produced D-space. This implementation is a realization of the path APC in Figure 2. Also, the authors employed the AdaBoost method for classifier design in the D-space. This later implementation achieves prototype selection, while building the classifier, which is a realization of the path APQR in the Figure.

B. Writer-Independent Systems

Instead of building a single writer-dependent (WD) classifier for each user using his enrolling signatures, a single writer-independent (WI) classifier is designed by learning the dissimilarities between signatures of all users. This concept is impossible to realize by means of the standard FR approach. However, it is possible to model the class distributions of intra-class and inter-class dissimilarities, by employing the DR approach. A single "global" classifier can be designed to model, or to discriminate between, these classes. If a huge number of samples are used to build the global DR-based classifier, it is statistically valid that the resulting model generalizes for users whose samples are not included in the training set.

The WI concept is proposed by Siteargur N. Srihari et al., [8], and Santos and Sabourin et al., [4]. 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. This implementation is a realization of the path BC in Figure 2, where the classifiers are designed in the FD-space. Improved implementation of this concept is proposed where different dissimilarity spaces are generated based on different feature representations, and classification decisions taken in each space are fused to produce the final decision [5], [6]. This scenario can be considered as generation of different instances for path BC, and fusion is done in the score or decision levels.

More recently, Rivard et al., [7] extended the idea to perform multiple feature extraction and selection. In this work, information fusion is also performed at the feature level. Multiple graphometric features are extracted based on multiple size grids. Then, the features are fused and pairwise distances between corresponding features are computed to constitute a high dimensional feature-dissimilarity space, where each dimension represents dissimilarity of a single feature. This complex representation is then simplified by applying the boosting feature selection approach (BFS) [12]. A sparser and more discriminative FD-space is produced by applying BFS with multi-feature extraction. This scenario can be considered as realization of path BDE in Figure 2. As the resulting WI classifier recognizes all users, even the users who are enrolled after the design phase, so the feature representation embedded in the WI classifier is considered as a global "population-based" representation.

C. Adaptation of Writer-Independent Systems

Recently, some work is done to combine advantages of both WI and WD approaches. Eskander et al., [14] extends on the system in [7] by adapting the population-based representation to each specific user, with the aim of reducing the classification complexity. While the first WI stage is designed in a FD-space, the following WD stage is designed in a standard feature space. Accordingly, the final WD classifier is FR-based classifier, that avoids storing reference signatures for enhanced security. Simulation results on two real-world offline signature databases (the Brazilian DB and GPDS public DB) confirm the feasibility and robustness of the proposed approach. Only a single compact classifier produced similar level of accuracy (Average Error Rate of about 5.38% and 13.96% for the Brazilian and the GPDS databases, respectively) as complex WI and WD systems in literature. This scenario is a realization of path BDHI in Figure 2.

IV. Research Directions

The aforementioned implementations represent a subset of large number of possible FR/DR combinations. Future research may investigate the unvisited scenarios of the proposed framework. For instance, combinations of global/user-specific, generative/discriminative, one-class/two-class systems can be designed. Also, all of the tasks for feature selection, prototype selection, classifier design, etc., can be employed in either feature space, dissimilarity matrix, FD-space, and D-space. Selection of the working space for each step, should depend on the specific requirements and constraints of the design problem and on the application itself. For example, in [14], features are selected in a FD-space as that provides a way to select reliable feature representations. Then, the classifiers are designed in a standard feature space, to avoid the need for storing signature templates for verification. Besides the large number of possible combinations and translations between the different spaces, there is also a wide range of pattern recognition techniques and tools that can be tested with the proposed framework. This includes different methods for feature extraction and selection, prototype selection, classifiers, etc.

From the application perspective, the proposed framework can be utilized for other applications, rather than the standard SV systems. For example, the Signature Identification (SI) systems that identify a producer of a signature sample, can be designed based on the DR-approach. Prototypes of all system users can be considered to build a classification D-space. Another example of systems, that imply a challenging design problem, is the signature-based bio-cryptographic systems. In these systems, cryptographic keys of encryption and digital signatures, are secured by means of handwritten signatures. It is a challenging to select informative features, signature prototype, and system parameters, for encoding reliable signature-based bio-cryptographic systems, based on the standard FR approach. Instead, recently, we proposed a methodology to design such systems, by means of the DR approach [15]. Features are selected in the FD-space and prototypes are selected in the D-space. Some of the system parameters such as length of the cryptographic key, are optimized in the different spaces.

V. Conclusions

In this paper, the dissimilarity approach for pattern recognition is considered to design signature verification (SV) systems. A general framework is proposed, for designing classification system based on a mixture of feature and dissimilarity representations. This framework imparts additional flexibility to the pattern recognition (PR) area. Combinations of transitions between different feature and dissimilarity spaces are suggested. Some of the existing implementations to the SV problem, are surveyed and related to the proposed framework. There are, however, a wide range of methodologies and applications that might benefit from the proposed approach, that opens a door for new research directions.

Acknowledgment

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

References

[1] D. Impedovo and G. Pirlo., Automatic signature verification: the state of the art. *IEEE Transactions on SMC, Part C: Applications and Reviews*, vol.38, no.5, pp.609-635,2008.

[2] Elzbieta Pekalska , Robert P.W. Duin. Dissimilarity representations allow for building good classifiers. *PR Letters*, vol.23, no.8, pp.161-166, 2002.

[3] Robert P.W. Duin , Marco Loog, Elzbieta Pekalska , and David M.J. Tax. Feature-Based Dissimilarity Space Classification. *Proceedings of the 20th International conference on Recognizing patterns in signals, speech, images, and videos (ICPR'10)*, pp.46-55, 2010.

[4] C. Santos, E. Justino, F. Bortolozzi, R. Sabourin. An off-line signature verification method based on document questioned experts approach and a neural network *Proceedings of 9Th IWFHR International Workshop on Frontiers in Handwriting Recognition (IWFHR'04)*, pp.498502, 2004.

[5] L. Oliveira, E. Justino, R. Sabourin. Off-line signature using writer-independent approach. *IJCNN*, pp.25392544, 2007.

[6] D. Bertolini, L. Oliveira, E. Justino, R.Sabourin. Reducing forgeries in writer-independent off-line signature verification through ensemble of classifiers. *PR*, vol.43, no.1, pp.387396, 2010.

[7] Rivard, D, Granger, E and Sabourin, R., Multi-Feature extraction and selection in writer-independent offline signature verification. *IJDAR*, vol.16, no.1, pp.83-103, 2013.

[8] Sargur N. Srihari, Aihua Xu and Meenakshi K. Kalera. Learning Strategies and Classification Methods for Off-Line Signature Verification. *Proceedings of the Ninth International Workshop on Frontiers in Handwriting Recognition (IWFHR'04)*,

[9] Elzbieta pekalska, Robert P.W. Duin, Pavel Paclk Prototype selection for dissimilarity-based classifiers. *PR*, vol.39, pp.189208, 2006.

[10] Jain, A.K. and Zongker, D. Representation and recognition of handwritten digits using deformable templates. *IEEE Transactions on PAMI*, vol.19, no.12, pp.1386-1390, 1997.

[11] S. Cha., Use of distance measures in handwriting Analysis. *PhD Thesis, State University of New York at Buffalo*, USA, 2001.

[12] K. Tieu and P. Viola., Boosting image retrieval. *International Journal of Computer Vision*, vol.56, no.1, pp.17-36, 2004.

[13] L. Batista, E. Granger and R. Sabourin. Applying Dissimilarity Representation to Off-Line Signature Verification . *International Conference on PR (ICPR)*, pp.1293-1297, 2010.

[14] Eskander, G.S., Sabourin, R. and Granger, E., Hybrid Writer-Independent–Writer-Dependent Offline Signature Verification System. *IET-Biometrics Journal, Special issue on Handwriting Biometrics*, doi: 10.1049/iet-bmt.2013.0024, in press, 2013.

[15] Eskander, G.S., Sabourin, R. and Granger, E., On the Dissimilarity Representation and Prototype Selection for Signature-Based Bio-Cryptographic Systems. *2nd Intel. Workshop on Similarity-Based Pattern Analysis and Recognition (SIMBAD2013)*, York, UK, 3-5 July 2013, LNCS, vol.7953, pp.265-280.