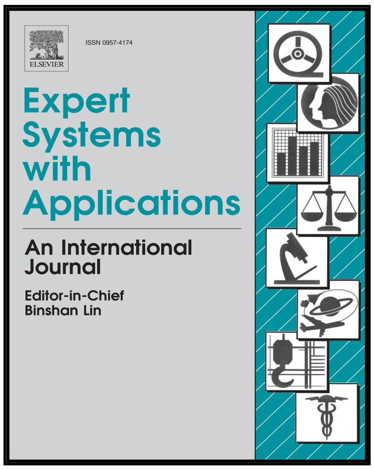
Accepted Manuscript



Interval Valued Symbolic Representation of Writer Dependent Features for Online Signature Verification

D.S. Guru , K.S. Manjunatha , S. Manjunath , M.T. Somashekara

PII: S0957-4174(17)30175-6
DOI: 10.1016/j.eswa.2017.03.024
Reference: ESWA 11179

To appear in: *Expert Systems With Applications*

Received date: 13 November 2016
Revised date: 10 March 2017

Please cite this article as: D.S. Guru , K.S. Manjunatha , S. Manjunath , M.T. Somashekara , Interval Valued Symbolic Representation of Writer Dependent Features for Online Signature Verification, *Expert Systems With Applications* (2017), doi: 10.1016/j.eswa.2017.03.024

Highlights

- An approach for online signature verification based on writer dependent parameters
- Interval valued symbolic representation of writer dependent features
- Verification based on both symbolic representation and conventional representation
- Lowest EER with symbolic representation and writer dependent parameters
- Obtained results indicate the superiority of the proposed approach

Interval Valued Symbolic Representation of Writer Dependent Features for Online Signature Verification

D.S. Guru[1], K. S. Manjunatha[1], S. Manjunath[1], M.T. Somashekara[2]

[1]Department of Studies in Computer Science, Manasagangothri, University of Mysore, Mysuru – 570 006, Karnataka, India
[2]Department of Computer Science and Applications, Bangalore University, Bengaluru – 560 056, Karnataka, India
dsg@compsci.uni-mysore.ac.in, kowshik.manjunath@gmail.com, manjunath.shantharamu@gmail.com, somashekara_mt@hotmail.com

Abstract

This work focusses on exploitation of the notion of writer dependent parameters for online signature verification. Writer dependent parameters namely features, decision threshold and feature dimension have been well exploited for effective verification. For each writer, a subset of the original set of features are selected using different filter based feature selection criteria. This is in contrast to writer independent approaches which work on a common set of features for all writers. Once features for each writer are selected, they are represented in the form of an interval valued symbolic feature vector. Number of features and the decision threshold to be used for each writer during verification are decided based on the equal error rate (EER) estimated with only the signatures considered for training the system. To demonstrate the effectiveness of the proposed approach, extensive experiments are conducted on both MCYT (DB1) and MCYT (DB2) benchmarking online signature datasets consisting of signatures of 100 and 330 individuals respectively using the available 100 global parametric features.

Keywords: Online signature verification, Writer dependent parameters, symbolic representation, symbolic feature vector, feature relevancy.

1. Introduction

Automatic signature verification is an interesting research problem in the area of biometrics. Signature verification is a process of determining whether a given signature truly belongs to a person who is claiming. It is the most widely accepted biometric trait for verifying the identity of a person in many day-to-day applications (Jain et al., 2002). During the last three decades, a significant progress has been made in the area of automatic signature verification. A number of models have been proposed which differ in the features, representation scheme adopted, dataset and the classifiers adopted. In spite of several models, finding an optimal set of discriminating features and also deciding upon a best classifier for verification are still open issues. The effectiveness of a verification system depends on the features used to discriminate genuine signatures from forgery signatures of each writer. Forgery may be either a skilled forgery which is imitated by professionals with sufficient practice or a random forgery with zero effort. As a result, for a verification system to be more effective, it is necessary to consider features which are most relevant for an individual writer rather than considering a common set of features for all writers. In addition, it is also uncommon to use the same number of features to verify signatures of every writer. This is due to the fact that some signatures are easy to forge when compared to other and also the consistency of signing varies from a writer to a writer. On the other hand, there will be variation in the signatures of a same writer and preserving these intra-writer variations is another challenging issue.

In recent years, the concept of symbolic data analysis has received a greater attention by researchers due to its ability in summarizing a large data of a specific domain. Symbolic data analysis has been exploited well for finding an effective solution for many pattern recognition applications such as data clustering (Carvalho 2007; Carvalho et al., 2009; Giusti and Grassini, 2008), text categorization (Harish et al., 2011), micro array data analysis (Hedjazi et al., 2013) and shape representation (Guru and Nagendraswamy, 2006; Daliri and Torre, 2008). It has been argued in these works that symbolic data in general and interval valued type data in particular has an ability to capture the intra class variability and thus have been capable of representing the reality. It has also been argued that a solution based on symbolic data outperforms a solution based on conventional crisp data (Gowda and Diday, 1991; Neto and De Carvalho, 2016; De Carvalho et al., 2016). Therefore even in this direction there is an attempt on symbolic representation of online signatures by means of interval-valued features (Guru and Prakash, 2007; Guru and Prakash, 2009).

In this paper, we propose a model for online signature verification based on the usage of various writer dependent parameters. Overall the following are the major contributions of this paper.

1. Exploring the notions of writer dependent features for online signature verification.
2. Preserving the intra-writer variations by representing the selected writer dependent features by means of an interval valued symbolic feature vector.
3. Fixing up of writer dependent feature dimension, similarity threshold based on the minimum error criterion.

4. Conduction of extensive experimentation for demonstrating the superiority of the proposed model when compared to many well-known models in achieving the lower error rate.

The remaining part of the paper is organized as follows. In section 2, a brief survey on related works is presented. In section 3, we discuss different stages of the proposed model. Description of the database used and details of experimentation conducted along with the results are presented in section 4. In section 5, we compare our model with other well-known existing models to bring out the superiority. Finally conclusions are drawn in section 6.

2. Related work

Authentication based on signatures finds numerous applications in many of our daily life such as banking transaction, financial transactions and attestation of documents etc. Depending on the acquisition mode, signature verification can be of two types namely offline and online (Jain et al., 2002). In online mode, when compared to offline mode, additional dynamic information such as pressure, velocity, speed etc., are also extracted in addition to the shape of a signature (Rashidi et al., 2012; Sae-bae and memon, 2014). As these dynamic characteristics are difficult to forge, online signature verification is more reliable than that of offline.

The approaches for online signature verification are categorized into parametric and function based approaches (Plamondon and Lorette, 1989; Impedovo and Pirlo, 2008). In a parametric based approach, a signature is stored in the knowledgebase by means of a few parameters. During testing, the corresponding parameters of a test signature are matched against that of a reference signature. Based on the estimated similarity, the authenticity of a test signature is decided. Generally, a parametric based approach takes a less time for matching and also a less memory for enrolling a writer. In a function-based approach, a signature is characterized by means of time functions of various dynamic properties such as pressure, speed etc. During matching, time functions of a test signature are matched against that of a reference signature directly on a point to point basis (Jain et al., 2002).

During verification, the authenticity of a test signature is decided by means of a suitable matching technique based on pattern recognition techniques such as Dynamic time warping (Jain et al., 2002; Khomatov and Yanikoglu, 2005), Hidden markov model (Aguilar et al., 2007; Enrique and Jose, 2012), Support vector machine (Khomatov and Yanikoglu, 2005; Christian et al., 2010), Neural network (Balzakis and Papamarkos, 2001), symbolic classifier (Guru and Prakash, 2009), random forest (Parodi and Gómez, 2014), neuro fuzzy (Cpalka and Zalasinki, 2014).

Further to reduce error rates and to enhance the reliability of a classifier, fusion based approaches (Aguiliar et al., 2005(a); Aguiliar et al., 2005(b); Nanni, 2006; Nanni et al., 2010, multi-stage approaches (Cordella et al., 1999; Sansone and vento, 2000; Zhang et al., 2001), multi-domain approaches (Pirlo, et al., 2014; Pirlo et al., 2015), ensemble of classifier approaches (Nanni and Lumini, 2005; Lumini and Nanni, 2009) have been proposed for signature verification. In a fusion based approach, the decision on the authenticity of a test signature is made based on the combined outcome of individual classifiers. In a multi-stage approach, the decision is organized into different stages and the final decision is made based on the outcome of the previous stages. In a multi-domain approach, a stability model is constructed for each writer based on the most stable segments of a signature which is represented in different domains. The authenticity of a test signature is decided by authenticating individual segments in which the given signature is most stable in the corresponding domain of representation. In case of ensemble of classifiers, the outcome of different classifiers trained with same data or different classifiers trained with a subset of data are combined using a suitable combination strategy.

Most of the above existing works share a common property that every writer is represented by means of either a common set of features or verification is done by means of a common classifier or a common classifier combination across all writers. Hence these models are referred to as writer independent models and they differ from the way a human expert performs verification. Generally, a human expert while verifying a signature looks for a different set of characteristics for different writers and also a different matching strategies for different writers. This demands the adoption of writer dependent characteristics for effective verification. In signature verification, writer dependency can be expected at three levels viz., feature level, decision threshold and classifier level. It is argued in the works of (Jain et al., 2002; Aguilar et al, 2005(a); Guru and prakash, 2009), usage of writer dependent threshold resulted in a better performance when compared to the usage of a common threshold. In addition, the number of features required for verification may not be same for all writers as the consistency of signing vary from a writer to a writer. Hence a verification system based on the usage of writer dependent features, writer dependent feature dimension and writer dependent threshold could be more effective than the one which is based on a common set of features with a common feature dimension.

A few attempts have been made towards the usage of writer dependent features which characterize an individual writer (Wijesoma, 2000; Kim et al., 2012). An initial attempt (Guru et al., 2013) is made on the usage of writer dependent features for online signature verification (Guru et al., 2013). In this work, even though the selected features are different for different writers, feature dimension and threshold for every writer are the same. The current work differs from our earlier work in many ways. In this work, writer dependent features are selected by means of simple dispersion measures used as relevance criteria. Further, in addition to the usage of writer dependent features, significance of writer dependent feature dimension and threshold are also studied.

In the current work, our intuition is to consider the problem of signature verification as a pattern recognition problem where the main objective is to perform signature verification based on the parameters selected for each writer individually. Therefore, the idea proposed in the current work can be applied on any pattern recognition problem in general and biometric verification in particular.

Recently an attempt on dynamic signature verification based on identifying the stable regions in different segments of a signature represented in different domains is made (Pirlo et al., 2015). Another challenging issue in signature verification is to preserve variations among the signature samples of a writer with a suitable representation. A few attempts can be traced towards capturing these variations (Marcos, 2007; Guru and Prakash, 2007; Guru and Prakash, 2009).

With this backdrop, we propose a verification model by the use of writer dependent features which are later effectively represented by means of a symbolic feature vector. In addition, this work is based on the usage of other writer dependent parameters such as feature dimension, similarity threshold and feature selection method. Suitable features, the number of features and the threshold to be used for each writer during verification are decided based on the error rate estimated with the training samples. At the outset, the contributions of this paper are of two folds. One is exploring the notion of writer dependent parameters for online signature verification and the other one is to demonstrate the superiority of symbolic representation and usage of a symbolic classifier over conventional counterpart in accomplishing a lower EER.

3. Proposed model

The proposed model basically has four important stages namely

1. Selection of writer dependent features based on feature relevance.
2. Representation of the selected writer dependent features in the form of an interval valued symbolic feature vector.
3. Fixation of writer dependent parameters
4. Signature verification based on the writer dependent parameters.

The block diagram of the proposed model is as shown in Fig 1. Here, in our work, we don't focus on extraction of features and preprocessing of features of a signature. As our intuition is to look for writer dependent feature selection, we just assume that a common set of features are available for each signature sample of every writer.

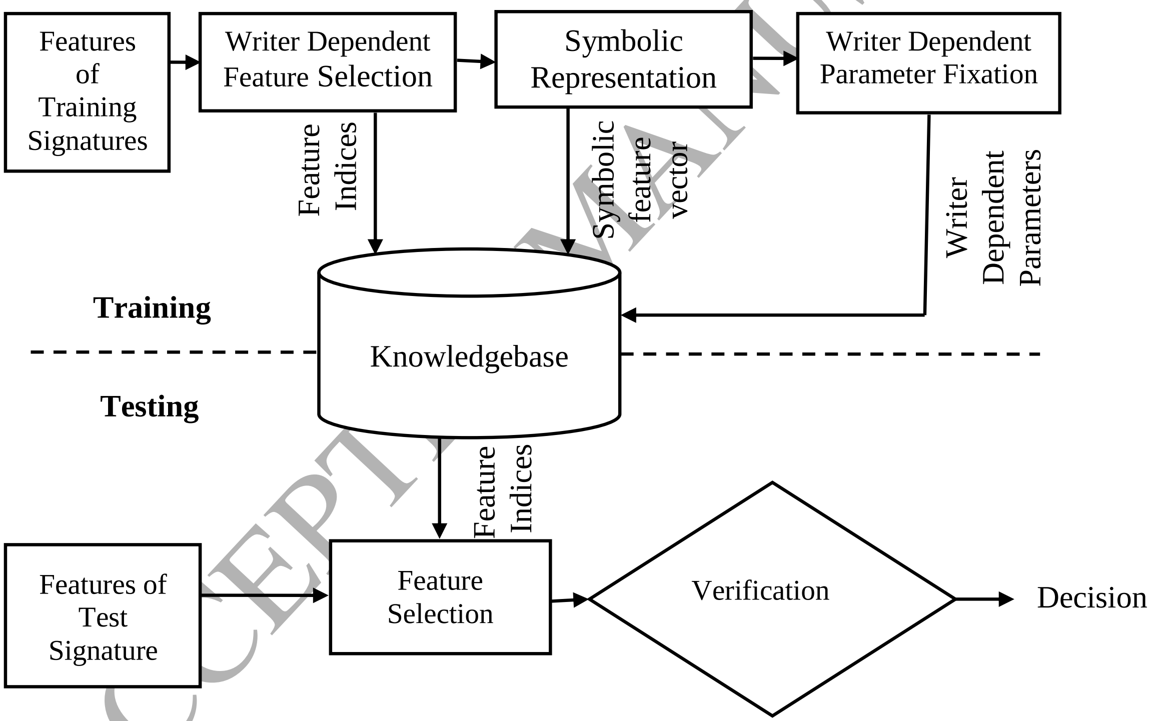


Fig. 1 Block diagram of the proposed model

3.1. Selection of writer dependent features

Let $\{S_1, S_2, S_3, ..., S_n\}$ be the n training signature samples of the i^{th} writer and let $\{f_{j1}, f_{j2}, ..., f_{jP}\}$ be the feature vector characterizing j^{th} signature ($j = 1, 2, ..., n$) of the i^{th} writer ($i = 1, 2, ..., N$). Here,

N is the number of writers, n is the number of training signature samples of each writer and P be the common dimension of the feature vector representing each writer.

We recommend selecting only d writer dependent features out of P available features based on their relevance for each writer. In this work, we considered the following 3 different dispersion measures for selecting writer dependent features (Artur et al., 2012).

1. **Mean Absolute Difference** (MAD_s)= $\frac{1}{n}\sum_{j=1}^{n}\left| f_{js} - \overline{f_s} \right|$, $s = 1, 2, \ldots, P$. (1)

   Here, $\overline{f_s}$ is the mean of the s^{th} feature of the writer W_i and f_{js} is the feature value of the s^{th} feature for the j^{th} sample of the writer W_i. MAD_s is computed for each feature i.e., $s = 1, 2, \ldots, P$. This statistical dispersion measure is more robust when compared to standard deviation as in case of standard deviation, outliers have more influence, but not in case of MAD_s.

2. **Mean Median Difference** (MM_s) = $\left| \overline{f_s} - Median(s) \right|$, $s = 1, 2, \ldots, P$ (2)

   The mean median difference MM_s is the absolute difference between the mean and the median of the s^{th} feature of the writer W_i. This dispersion measure is computationally more efficient than variance.

3. **Modified Arithmetic and Geometric Mean ratio** ($AGMR_s$) $= \frac{1}{n\exp(\overline{f_s})}\sum_{j=1}^{n} f_{js}$,

   $s = 1, 2, \ldots, P$ (3)

For every feature a score is computed by applying each of the above three dispersion measures. The computed score denotes the relevance of a feature. After computing the relevancy score of all P features, the features are ranked based on the scores computed and the top d features are selected. The top features selected by the three dispersion measures are represented by means of the following three feature vectors.

$$FS_1 = \left\{ f_1^1, f_2^1, f_3^1, \ldots, f_d^1 \right\} \quad (4)$$

$$FS_2 = \left\{ f_1^2, f_2^2, f_3^2, ..., f_d^2 \right\} \quad (5)$$

$$FS_3 = \left\{ f_1^3, f_2^3, f_3^3, ..., f_d^3 \right\} \quad (6)$$

Here, subscript denotes the indices of the selected features and superscript denotes the dispersion measure used. For instance f_a^b denotes the index of the a^{th} feature selected from the b^{th} dispersion measure. Here $FS_{cr}, (cr = 1, 2, 3)$ represents the set of indices of the d features selected from the cr dispersion measure.

Further, we fused the feature vectors FS_1, FS_2, and FS_3 by means of union and intersection operations to create additional feature vectors as follows.

$$FS_1 \cup FS_2 \cup FS_3 = \left\{ f_1^1, f_2^1, f_3^1, ..., f_d^1 \right\} \cup \left\{ f_1^2, f_2^2, f_3^2, ..., f_d^2 \right\} \cup \left\{ f_1^3, f_2^3, f_3^3, ..., f_d^3 \right\}$$

$$= \{ f_1^{cr}, f_2^{cr}, f_3^{cr} ..., f_{t1}^{cr} \}, \quad (7)$$

where $t1 \geq d$ and each f_x^{cr} in $FS_1 \cup FS_2 \cup FS_3 \in FS_1$ or FS_2 or FS_3, $x = 1, 2, ..., t1$ and cr =1 or 2 or 3.

Similarly

$$FS_1 \cap FS_2 \cap FS_3 = \left\{ f_1^1, f_2^1, f_3^1, ..., f_d^1 \right\} \cap \left\{ f_1^2, f_2^2, f_3^2, ..., f_d^2 \right\} \cap \left\{ f_1^3, f_2^3, f_3^3, ..., f_d^3 \right\}$$

$$= f_1^{cr}, f_2^{cr}, f_3^{cr} ..., f_{t2}^{cr}, \quad (8)$$

where $t2 \leq d$ and each f_x^{cr} in $FS_1 \cap FS_2 \cap FS_3 \in FS_1$, FS_2 and FS_3, $x = 1, 2, ..., t2$ and cr =1 or 2 or 3.

We also recommend fusing these feature vectors two at a time thereby totally resulting with 11 different combinations of feature vectors which are denoted as $FS_1, FS_2, FS_3, U_{123}, I_{123}, U_{12}, U_{13}, U_{23}, I_{12}, I_{13}$ and I_{23} where the labels U and I denote union and intersection combinations respectively. In Table 1, indices of the top 10 most relevant features selected from each of the three dispersion measures are shown for the first 5 writers of MCYT (DB1) dataset as examples.

Table 1 Indices of the ten most relevant features selected from the three dispersion measures for the first 5 writers of MCYT (DB1) as examples

Writer Id	Dispersion Measure	Indices of the selected features									
1	FS_1	10	20	6	39	21	85	9	99	76	5
	FS_2	6	20	10	21	5	9	39	83	47	99
	FS_3	20	39	85	10	21	9	99	47	11	4
2	FS_1	20	10	39	6	76	85	33	83	21	3
	FS_2	10	6	76	9	85	3	20	33	21	44
	FS_3	6	76	85	20	10	33	78	44	99	83
3	FS_1	20	44	33	3	76	39	85	10	83	99
	FS_2	3	33	44	20	8	39	85	76	12	10
	FS_3	3	33	85	20	21	39	44	8	76	12
4	FS_1	33	44	20	85	6	3	39	10	21	8
	FS_2	39	44	20	85	33	10	8	3	21	76
	FS_3	33	10	20	8	44	85	76	6	9	39
5	FS_1	6	10	39	20	85	83	5	21	76	9
	FS_2	6	21	39	10	20	5	85	83	76	3
	FS_3	6	39	20	21	9	83	78	5	3	98

3.2 Symbolic representation

Once the writer dependent features are selected, training signatures of each writer are stored in the knowledgebase in the form of an interval valued symbolic feature vector (Guru and Prakash, 2007, 2009). The symbolic feature vector for the i^{th} writer is created as follows.

Let $\{S_1, S_2, S_3, ..., S_n\}$ be n training signatures of the i^{th} writer ($i = 1, 2, ..., N$), where N is the number of writers. Let $\{f_{i1}, f_{i2}, ..., f_{id}\}$ be the feature vector representing the i^{th} writer, where d is the number of features selected. To compute the interval-valued feature vector to represent the i^{th} writer, we compute the mean and standard deviation of each of the d features selected. Let $Mean(f_p^i)$ and

$Std(f_p^i)$ be the mean and the standard deviation of the p^{th} feature due to all n samples of the i^{th} writer ($p = 1, 2, \ldots, d$). That is,

$$Mean(f_p^i) = \frac{1}{n}\sum_{s=1}^{n} f_{sp} \text{ and } Std(f_p^i) = \left(\frac{1}{n}\sum_{s=1}^{n}(f_{sp} - Mean(f_p^i))^2\right)^{\frac{1}{2}} \tag{9}$$

After computing the mean and the standard deviation of all the selected features, each feature of the writer W_i is represented in the form of an interval. For example, the p^{th} feature of the i^{th} writer is represented as $\left[f_{ip}^-, f_{ip}^+\right]$ where f_{ip}^- and f_{ip}^+ denote the lower limit and the upper limit of the p^{th} feature of the i^{th} writer respectively which are computed as in (10).

$$f_{ip}^- = Mean(f_p^i) - Std(f_p^i) \text{ and } f_{ip}^+ = Mean(f_p^i) + Std(f_p^i)\,. \tag{10}$$

Thus, the interval $\left[f_{ip}^-, f_{ip}^+\right]$ depends on the mean and the standard deviation of the p^{th} feature values of the i^{th} writer. In general, each of the d features selected is represented in the form of an interval which results in the creation of an interval valued symbolic feature vector say RF_i for the i^{th} writer, given by

$$RF_i = \left\{\left[f_{i1}^-, f_{i1}^+\right], \left[f_{i2}^-, f_{i2}^+\right] \ldots \left[f_{id}^-, f_{id}^+\right]\right\},\ i = 1, 2, \ldots, N \tag{11}$$

This symbolic feature vector is stored in the knowledgebase as the representative of the i^{th} writer. In this representation, instead of storing all signatures of a writer, it is sufficient to store only one symbolic feature vector characterizing the writer. Hence the total number of reference feature vectors to be stored in the knowledgebase is only N . In (Jain et al., 2002; Aguilar et al., 2005(a)), for each writer, all n training signatures are stored in the knowledgebase. Hence the total number of signatures to be stored in the knowledgebase is $N \times n$. Even though, templates are stored instead of all training signatures of a writer (Sae-bae and Memon, 2014; Cpalka et al., 2016), it is necessary to store multiple templates for each writer to assimilate intra-writer variations effectively (Liu and Wang, 2008; Garcia et al., 2014). In Table 2, interval valued symbolic feature vectors for the first 5 writers of the MCYT (DB1) dataset are shown as examples along with the indices of the top 5 most relevant features selected for each writer.

Table 2. Interval valued symbolic feature vectors (top 5 features) for 5 writers of MCYT (DB1) along with the indices of the features selected

Writer ID	Interval values of selected features					
1	Indices	10	20	6	39	21
	Interval	[197.70, 243.86]	[121.88, 152.51]	[121.17, 139.09]	[51.83, 62.11]	[60.75, 73.56]
2	Indices	20	10	39	6	76
	Interval	[99.92, 123.48]	[61.05, 77.27]	[26.30, 51.01]	[53.97, 59.09]	[25.89, 43.12]
3	Indices	20	44	33	3	76
	Interval	[97.62, 136.45]	[21.38, 35.42]	[23.39, 55.81]	[112.30, 135.30]	[24.97, 36.12]
4	Indices	33	44	20	85	6
	Interval	[8.01, 37.19]	[5.26, 30.34]	[86.90, 102.93]	[26.76, 46.91]	[46.83, 54.50]
5	Indices	6	10	39	20	85
	Interval	[72.19, 82.50]	[61.18, 66.30]	[58.82, 78.61]	[92.89, 108.81]	[23.11, 32.20]

3.3 Writer dependent parameter fixation

In this section, we discuss the procedure adopted for fixing up of writer dependent parameters namely writer dependent feature dimension and similarity threshold. Feature dimension and the threshold to be used for each writer during verification are determined empirically as follows. For each writer, an interval valued symbolic reference feature vector is computed considering the training signatures only (genuine signatures) as explained in section 3.2. Then the feature values of each of the training signatures are compared with the reference feature vector to decide the number of features of the training signature that lie within the corresponding interval value of the reference feature vector. After computing the score for all features for all writers, we vary the number of features to be selected (d) from 5 to 75 in step of 5 for each writer. For each d, the similarity threshold is varied from 0.1 to 1.0 in step of 0.1 and the FAR and the FRR are estimated. The FRR is calculated considering the genuine signatures used for training and the FRR is calculated considering equal number of random forgeries (genuine signatures of other writers). Finally, the EER is estimated for each d from the receiver operating characteristics (ROC) curve. The decision on the feature dimension for each writer is arrived based on the minimum EER criterion. That is, d which results in lowest EER is decided to be the suitable feature dimension and the corresponding threshold as the suitable threshold for the respective writer. In case of symbolic representation, similarity threshold is the percentage of the number of features of a test signature that should lie within the corresponding interval valued features of the reference signature. For example, similarity threshold value equal to 0.5 indicates that 50% of the features of a test signature should lie within the corresponding intervals of the reference signature for accepting it as a genuine signature. In case of conventional representation, threshold is the normalized distance estimated among the genuine training signatures of the corresponding writer scaled to the range 0 to 1. We arrive at these parameters based on 20 trials conducted with randomly selected training signature samples in each trial. The same procedure is repeated with all the 11 different feature selection combinations. Finally, the feature selection method which gives the lowest EER is decided to be the suitable feature selection method for the respective writer.

3.4. Signature Verification

Given is a test signature characterized by its P dimensional feature vector say $\mathrm{F_q} = \{\mathrm{f_{q1}}, \mathrm{f_{q2}}, \ldots, \mathrm{f_{qP}}\}$, its authenticity is decided by comparing it with the symbolic reference feature vector of the claimed writer. It is interesting to note that all the features of the test signature are of crisp type while the corresponding features of the reference signature are of of interval valued. For authentication, we compare only d features ($d < P$) of the test signature with the corresponding d interval valued features of the reference signature. The indices of the d features to be compared are available in the knowledgebase. To keep track of the number of features of a test signature that lie within the corresponding interval valued feature of the reference signature we use a counter (A_{cp}). If a feature of a test signature lies within the corresponding interval-valued feature of a reference signature, the A_{cp} is incremented by one. If the value of A_{cp} is greater than the predefined similarity threshold computed for the corresponding writer, then the test signature is considered as a genuine otherwise the test signature is rejected as a forgery.

4. Experimentation and results

In this section, we discuss about the dataset used for experimentation, training and testing details, experimental protocol along with the results obtained.

Dataset: We conducted an experimentation using MCYT online signature data sets (both DB1 and DB2), standard benchmarking datasets for online signatures. The DB1 dataset is consisting of signatures of first 100 writers and the DB2 is consisting of signatures of 330 writers (Garcia et al., 2003). Both datasets consist of 25 genuine and 25 skilled forgeries for each writer where the skilled forgeries are collected from 5 different professionals. For experimentation, we considered the available 100 preprocessed global parametric features, the details of which can be seen in (Aguilar et al., 2005(a).

Experimental protocol: During experimentation, a data set is divided into training and testing subsets. We tested the performance of the proposed model with both skilled and random forgeries. Skilled forgeries are nothing but forgeries created with sufficient practice. Random forgeries are nothing but the genuine signatures of other writers. We considered 4 different categories of training and testing named as Skilled_05, Skilled_20, Random_05 and Random_20. The details of training and testing signatures used in these four categories are shown in Table-3. In Table 3, the notations G and SF denote genuine signatures and skilled forgeries respectively. For identifying the common feature dimension and the common threshold to be used for all writers, we varied the number of features selected (d) from 5 to 75 in step of 5. For each d , the similarity threshold is varied from 0.1 to 1.0 in step of 0.1 and estimated the false acceptance rate (FAR) and the false rejection rate (FRR). The FAR is the percentage of forgery samples wrongly treated as genuine signatures and the FRR is the percentage of genuine signatures wrongly rejected as forgery. The point at which these two values are equal in the ROC curve is the EER of the system. In Fig 2(a) to Fig 2(e), the variations of both FAR and FRR are shown as examples with the number of feature selected being equal to 25 for different combinations of feature selection with 5 training signatures.

Table 3. Details on number of training and testing samples for each writer under four categories of experiments.

Category	Training	Testing	
		DB1	DB2
Skilled_05	05 G	20 G (For FRR) 25 SF (For FAR)	20 G (For FRR) 25 SF (For FAR)
Skilled_20	20 G	05 G (For FRR) 25 SF (For FAR)	05 G (For FRR) 25 SF (For FAR)
Random_05	05 G	20 G (For FRR) 99 G of other writers (For FAR)	20 G (For FRR) 329 G of other writers (For FAR)
Random_20	20 G	05 G (For FRR) 99 G of other writers (For FAR)	05 G (For FRR) 329 G of other writers (For FAR)

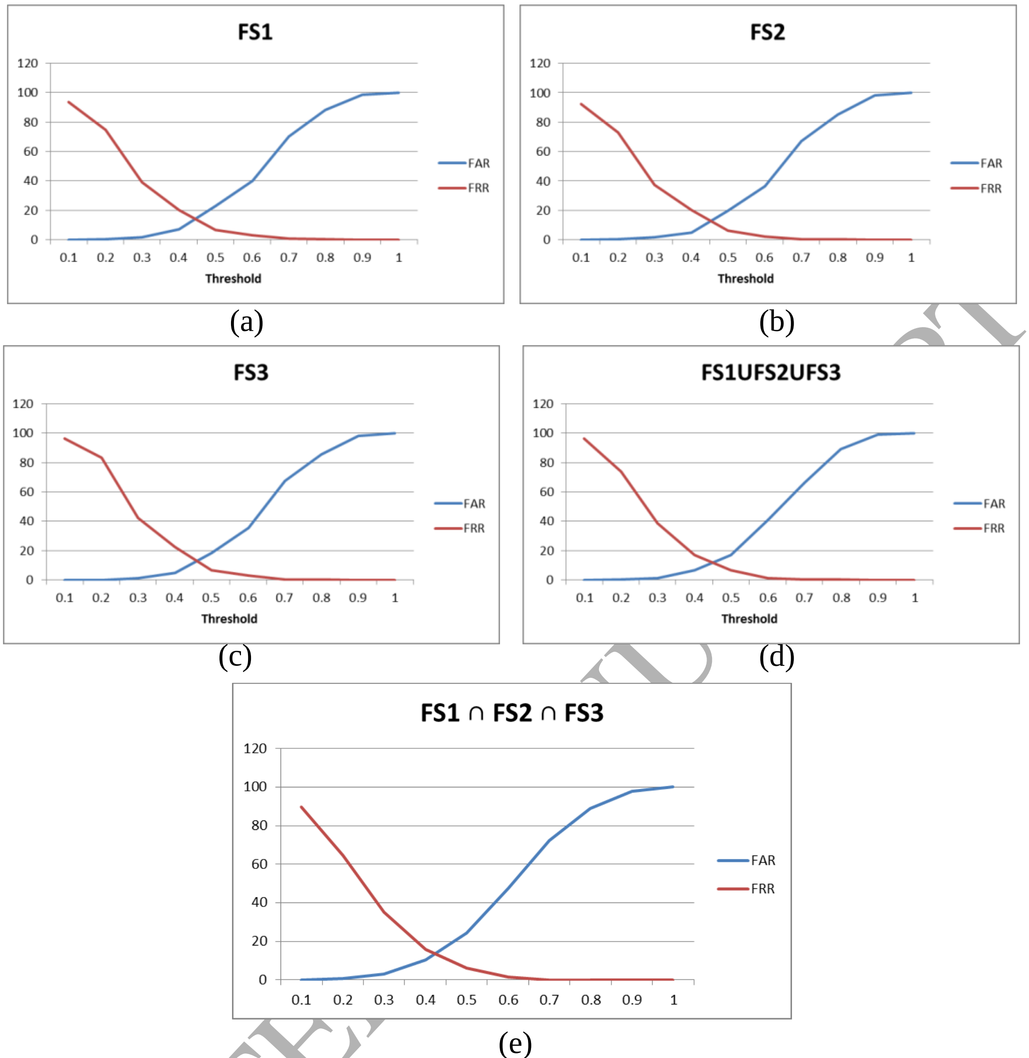


Fig. 2 Variations of FAR and FRR for various combinations of Feature Selection
(a) FS1 (b) FS2 (c) FS3 (d) FS1UFS2UFS3 (e) FS1∩FS2∩FS3

In Fig 3(a) to Fig 3(e) the EER obtained for the first 10 writers of the DB1 dataset are shown when 05 signatures are used for training as examples.

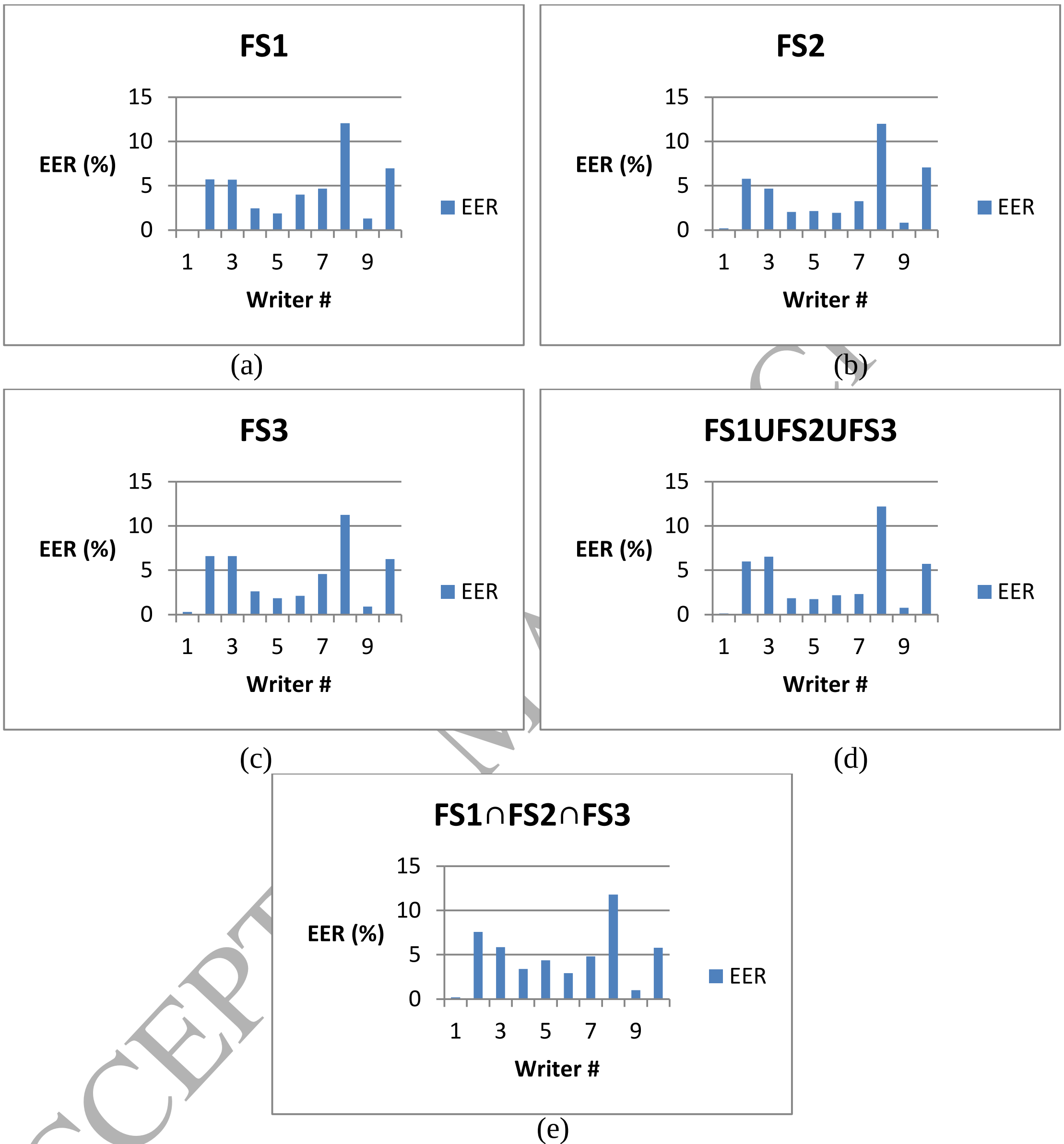


Fig. 3. EER of Individual Writers of DB1 (Skilled_05)

(a) FS1 (b) FS2 (c) FS3 (d) FS1UFS2UFS3 (e) FS1∩FS2∩FS3

We conducted similar experiments on DB2 also. The minimum, maximum and average EER due to twenty trials are shown in Table 4 and Table 5 for DB1 and DB2 respectively.

Table 4. The Minimum, Maximum and Average EER for different combinations of feature selection methods with common threshold and common feature dimension on DB1

Feature selection Method	Skilled_05 Threshold = 0.5			Skilled_20 Threshold = 0.5			Random_05 Threshold = 0.4			Random_20 Threshold = 0.4		
	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg
FS1	5.7 (70)	9.8 (75)	7.9 (68)	4.5 (75)	6.8 (70)	5.3 (75)	2.7 (75)	5.0 (70)	3.6 (75)	1.1 (55)	3.7 (75)	2.7 (69)
FS2	5.7 (70)	9.6 (75)	7.8 (67)	4.5 (75)	6.8 (75)	5.5 (72)	2.4 (70)	4.3 (75)	3.4 (74)	1.2 (55)	3.3 (70)	2.4 (65)
FS3	6.0 (75)	9.1 (70)	7.9 (70)	4.1 (75)	7.7 (70)	5.0 (75)	2.8 (70)	4.1 (75)	3.4 (75)	1.5 (70	3.4 (75)	2.5 (73)
U123	6.1 (70)	8.4 (75)	7.4 (71)	3.7 (75)	7.6 (55)	5.0 (72)	2.6 (75)	4.8 (65)	3.1 (72)	1.8 (60)	3.0 (75)	2.3 (71)
I123	6.7 (70)	9.0 (75)	8.4 (75)	4.9 (75)	6.9 (75)	5.7 (75)	3.1 (75)	4.3 (75)	3.8 (75)	1.7 (75)	3.8 (75)	2.9 (70)
U12	6.9 (75)	9.1 (65)	8.0 (73)	4.6 (75)	7.4 (75)	5.6 (73)	2.4 (75)	4.2 (65)	3.3 (72)	1.7 (75)	3.3 (75)	2.5 (72)
U13	6.7 (75)	8.8 (75)	7.6 (71)	4.0 (75)	7.6 (55)	5.1 (73)	2.3 (75)	4.1 (60)	3.2 (67)	1.2 (70)	2.9 (75)	2.2 (72)
U23	6.2 (65)	8.6 (75)	7.5 (73)	4.0 (75	5.3 (70)	4.6 (75)	2.4 (75)	4.4 (65)	3.1 (70)	1.5 (65)	2.9 (75)	2.2 (72)
I12	6.8 (65)	9.1 (75)	8.0 (71)	4.7 (75)	7.3 (75)	5.5 (75)	2.8 (70)	5.2 (75)	3.4 (74)	1.6 (60)	3.5 (75)	2.6 (71)
I13	7.7 (65)	9.5 (75)	8.6 (73)	5.3 (75)	7.7 (65)	6.0 (74)	2.9 (75)	4.7 (70)	3.8 (74)	1.0 (75)	3.6 (75)	2.5 (71)
I23	6.4 (75)	8.9 (75)	7.8 (72)	4.7 (75)	7.0 (75)	5.8 (75)	3.0 (75)	4.5 (75)	3.8 (75)	1.1 (70)	3.9 (75)	2.5 (71)

Entries in Table 4 and Table 5 correspond to the EER obtained with a common threshold and a common feature dimension. The feature dimension which resulted in a lowest EER is also shown within parenthesis. The notations U and I denote the feature vectors obtained with union and intersection of the feature indices obtained from different criterion. For instance U123 denotes the feature vector obtained from the union of FS_1, FS_2 and FS_3. Even though it is a general statement that verification based on writer dependent parameters performs better than that of a common set of parameters, Table 4 and Table 5 are provided as an empirical proof. Further, it also helps in comparing the results obtained based on writer dependent parameters.

Table 5**.** The Minimum, Maximum and Average EER for different combinations of feature selection methods with common threshold and common feature dimension on DB2

Feature selection Method	Skilled_05 Threshold = 0.5			Skilled_20 Threshold = 0.5			Random_05 Threshold = 0.4			Random_20 Threshold = 0.4		
	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg
FS1	7.0 (60)	10.9 (70)	8.9 (69)	5.2 (75)	9.5 (70)	6.1 (74)	3.3 (75)	4.4 (75)	3.9 (73)	1.9 (60)	3.6 (75)	2.8 (70)
FS2	7.2 (70)	10.2 (70)	8.7 (71)	5.1 (75)	7.6 (70)	5.8 (75)	3.2 (75)	5.8 (70)	3.9 (74)	2.2 (75)	3.7 (75)	3.0 (72)
FS3	7.2 (60)	11.0 (75)	8.9 (70)	5.3 (75)	7.7 (75)	6.0 (75)	3.1 (75)	4.1 (65)	3.5 (73)	1.7 (60)	3.6 (75)	2.7 (71)
U123	7.7 (75)	9.1 (75)	8.5 (73)	5.2 (75)	6.4 (75)	5.8 (74)	2.8 (75)	4.0 (70)	3.4 (71)	1.7 (60)	3.6 (75)	2.7 (72)
I123	6.9 (70)	11.4 (75)	9.0 (74)	5.8 (75)	8.9 (75)	6.7 (74)	3.6 (75)	5.1 (75)	4.1 (75)	2.4 (50)	3.6 (75)	2.9 (72)
U12	7.4 (75)	9.7 (75)	8.6 (73)	5.2 (75)	6.4 (75)	5.8 (75)	3.0 (75)	5.3 (75)	3.8 (74)	1.9 (75)	3.3 (75)	2.6 (74)
U13	7.1 (65)	10.1 (75)	8.6 (72)	5.1 (75)	8.6 (75)	5.8 (75)	2.6 (75)	4.0 (65)	3.2 (71)	1.6 (60)	3.1 (75)	2.6 (74)
U23	7.4 (75)	9.3 (75)	8.6 (74)	5.3 (75)	7.3 (70)	5.8 (75)	2.9 (75)	5.6 (75)	3.6 (72)	1.5 (75)	3.2 (75)	2.6 (74)
I12	7.8 (70)	9.9 (75)	8.9 (74)	5.4 (70)	8.2 (70)	6.2 (75)	3.0 (70)	4.7 (75)	3.7 (74)	1.8 (55)	3.4 (75)	2.7 (74)
I13	7.4 (75)	12.5 (75)	9.1 (74)	5.5 (75)	7.9 (70)	6.5 (75)	3.2 (75)	5.0 (70)	3.9 (75)	2.0 (75)	3.9 (75)	3.0 (74)
I23	7.2 (75)	11.1 (75)	8.7 (74)	5.9 (75)	7.9 (75)	6.5 (75)	3.3 (75)	4.8 (70)	4.1 (75)	2.1 (75)	3.9 (75)	3.1 (74)

We have also conducted experiments to estimate the EER using writer dependent parameters. In this case, the average EER of the system is estimated by taking the average of the EER of each individual writer. We conducted 20 trials with different training and testing samples in each trial and average EER of 20 trials is considered to be the EER of an individual writer. The Minimum, Maximum and average EER obtained with writer dependent parameters are shown in Table 6 and Table 7 for DB1 and DB2 respectively. Since the threshold and the feature dimensions vary from a writer to a writer, threshold and the feature dimension are not shown in Table 6 and Table 7.

Table 6. The Minimum, Maximum and Average EER for different combinations of feature selection methods with writer dependent parameters on DB1

Feature selection Method	Skilled_05			Skilled_20			Random_05			Random_20		
	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg
FS1	2.7	4.3	3.3	0.7	2.0	1.3	1.1	2.3	1.4	0.1	1.2	0.7
FS2	2.5	4.0	3.2	0.9	2.6	1.7	1.1	1.8	1.5	0.2	0.9	0.5
FS3	2.8	4.4	3.5	0.8	2.6	1.6	1.0	1.8	1.3	0.2	1.0	0.6
U123	2.2	3.9	3.1	0.8	2.2	1.5	1.1	1.8	1.4	0.3	1.0	0.6
I123	3.3	4.7	3.8	0.9	2.1	1.5	1.3	1.9	1.6	0.4	1.0	0.7
U12	2.9	3.9	3.3	0.7	2.4	1.8	1.2	2.0	1.5	0.3	1.0	0.6
U13	2.4	3.8	3.1	0.9	2.5	1.6	1.0	1.7	1.4	0.3	1.0	0.6
U23	2.6	4.0	3.1	0.6	2.1	1.3	1.0	1.8	1.3	0.2	0.9	0.5
I12	3.0	4.2	3.5	1.0	2.8	1.6	1.0	1.8	1.4	0.3	1.2	0.7
I13	3.3	4.9	3.9	1.2	2.2	1.6	1.0	2.2	1.6	0.2	1.1	0.7
I23	3.0	4.2	3.6	1.4	2.6	1.9	1.4	2.1	1.7	0.1	1.3	0.7

Table 7. The Minimum, Maximum and Average EER for different combinations of feature selection methods with writer dependent parameters on DB2

Feature selection Method	Skilled_05			Skilled_20			Random_05			Random_20		
	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg
FS1	3.5	4.3	3.9	1.2	2.3	1.7	1.4	2.1	1.8	0.4	1.1	0.6
FS2	3.4	4.5	3.9	1.5	2.6	2.1	1.5	2.7	1.9	0.6	1.4	0.8
FS3	3.4	4.8	4.0	1.7	3.4	2.2	1.4	2.0	1.6	0.2	1.1	1.0
U123	3.3	4.1	3.8	1.5	2.4	1.8	1.4	2.0	1.6	0.4	0.9	0.7
I123	3.7	5.0	4.3	0.9	2.3	1.6	1.7	2.5	2.0	0.4	0.9	0.6
U12	3.3	4.2	3.9	1.3	2.8	2.1	1.4	2.2	1.8	0.4	1.1	0.6
U13	3.1	4.4	3.6	1.0	2.3	1.8	1.4	1.9	1.6	0.4	0.9	0.6
U23	3.4	4.3	3.7	1.2	3.5	2.0	1.4	2.2	1.6	0.2	1.2	0.7
I12	3.4	4.9	4.0	1.3	2.8	1.9	1.7	2.2	1.9	0.4	1.0	0.8
I13	3.7	4.9	4.4	1.1	3.4	2.0	1.7	2.3	1.9	0.5	1.5	0.9
I23	3.6	5.4	4.4	1.2	2.6	2.1	1.6	2.3	1.9	0.6	0.9	0.8

From Table 4 to Table 7, it is clear that usage of writer dependent threshold and writer dependent feature dimension resulted in a considerable reduction in the EER when compared to the usage of a common threshold and a common feature dimension. This shows the superiority of writer dependent parameters for online signature verification.

For the sake of comparison between symbolic and other existing conventional representation schemes, we conducted verification experiments with a conventional representation also. In a conventional representation, every feature is of crisp type unlike a symbolic feature where every feature is of interval valued type. In case of conventional representation, verification is done by means of a minimum distance classifier. Here also we conducted experimentation with both common set of parameters and writer dependent parameters. Table 8 and Table 9 show the minimum, maximum and average EER obtained for conventional representation on DB1 with a common set of parameters and writer dependent parameters respectively.

Table 8. The Minimum, Maximum and Average EER for different combinations of feature selection methods for conventional representation along with common threshold and common feature dimension on DB1

Feature selection Method	Skilled_05 Threshold = 0.4			Skilled_20 Threshold = 0.3			Random_05 Threshold = 0.5			Random_20 Threshold = 0.4		
	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg
FS1	9.2	10.9	9.9	7.8	9.9	8.7	5.7	6.4	6.0	3.9	5.3	4.6
	(75)	(50)	(51)	(15)	(40)	(27)	(40)	(25)	(38)	(50)	(30)	(42)
FS2	9.5	10.7	10.1	5.7	8.6	7.6	5.7	6.4	6.0	4.3	5.3	4.6
	(60)	(35)	(46)	(50)	(20)	(26)	(40)	(25)	(38)	(40)	(35)	(42)
FS3	9.5	10.7	10.1	6.4	8.6	7.2	5.7	6.5	6.2	4.2	5.3	4.8
	(60)	(35)	(47)	(45)	(70)	(48)	(65)	(30)	(48)	(50)	(55)	(57)
U123	9.4	10.9	10.3	6.7	9.7	8.0	5.7	6.4	6.0	4.0	5.0	4.4
	(30)	(40)	(46)	(40)	(25)	(33)	(45)	(40)	(41)	(35)	(50)	(40)
I123	9.4	10.3	9.8	7.4	8.9	7.9	5.8	6.5	6.1	4.0	5.4	4.6
	(75)	(50)	(65)	(50)	(50)	(36)	(30)	(60)	(53)	(55)	(60)	(56)
U12	9.2	11.2	10.1	6.5	9.3	7.9	5.6	6.5	5.9	4.2	5.1	4.6
	(60)	(50)	(48)	(20)	(50)	(25)	(35)	(40)	(40)	(30)	(45)	(37)
U13	9.6	11.1	10.3	7.0	9.7	8.4	5.7	6.7	6.1	4.2	5.3	4.6
	(40)	(40)	(46)	(25)	(40)	(26)	(45)	(40)	(44)	(45)	(40)	(40)
U23	9.5	11.0	10.0	7.8	9.5	8.3	5.6	6.6	5.9	4.0	5.3	4.7
	(45)	(50)	(47)	(40)	(45)	(30)	(25)	(35)	(36)	(55)	(35)	(39)
I12	9.4	11.1	9.9	7.3	9.5	8.2	5.5	6.4	6.1	4.0	5.0	4.6
	(45)	(40)	(42)	(15)	(15)	(25)	(35)	(40)	(41)	(30)	(40)	(42)
I13	9.7	11.3	10.2	7.3	9.1	8.2	5.8	7.0	6.1	4.0	5.4	4.7
	(60)	(45)	(61)	(15)	(15)	(23)	(75)	(65)	(61)	(65)	(55)	(56)
I23	9.6	10.7	10.1	7.0	9.3	8.1	5.6	6.8	6.1	4.3	5.0	4.6
	(35)	(70)	(59)	(30)	(25)	(31)	(35)	(50)	(45)	(30)	(50)	(62)

Table 9. The Minimum, Maximum and Average EER for different combinations of feature selection methods for conventional representation with writer dependent parameters on DB1

Feature selection Method	Skilled_05			Skilled_20			Random_05			Random_20		
	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg
FS1	6.3	7.8	6.7	4.2	5.2	4.5	3.3	3.6	3.4	1.5	2.5	2.0
FS2	6.2	7.7	6.7	4.2	5.4	4.8	3.5	3.7	3.6	1.7	2.1	1.9
FS3	5.5	6.9	6.3	4.1	4.9	4.4	3.2	3.7	3.5	1.7	2.1	1.9
U123	6.3	7.6	6.9	4.2	5.0	4.7	3.3	3.9	3.5	1.8	2.2	2.0
I123	6.0	7.6	6.5	4.3	4.8	4.6	3.2	3.9	3.4	1.6	2.0	1.7
U12	6.2	8.4	6.9	3.9	5.7	4.5	3.3	3.8	3.5	1.4	2.5	1.9
U13	6.3	8.0	7.0	4.6	5.2	4.8	3.2	3.6	3.4	1.8	2.9	2.1
U23	6.2	8.2	7.0	4.9	5.5	5.2	3.3	4.1	3.7	1.6	2.3	2.1
I12	6.0	8.1	6.7	3.8	5.7	4.5	3.2	4.2	3.5	1.5	2.5	1.9
I13	6.4	7.0	6.7	4.1	5.6	4.8	3.3	3.5	3.6	1.7	2.3	2.1
I23	6.2	7.7	6.8	3.9	4.7	4.4	3.3	3.4	3.4	1.7	2.1	1.8

In a conventional representation also, usage of writer dependent parameters yielded lower EER when compared to a common set of parameters. Fig 4 and Fig 5 show the EER obtained with symbolic and conventional representation with a common set of parameters and writer dependent parameters respectively for DB1.

We obtained the lower EER in all categories for symbolic representation scheme when compared to a conventional representation. This clearly indicates the superiority of symbolic representation when compared to a conventional representation. Results also demonstrate the superiority of writer dependent parameters when compared to a common set of parameters for signature verification with both symbolic and conventional representation. Finally, complete details of writer specific parameters for the first 5 writers from DB1 in Skilled_05 category are shown in Table 10.

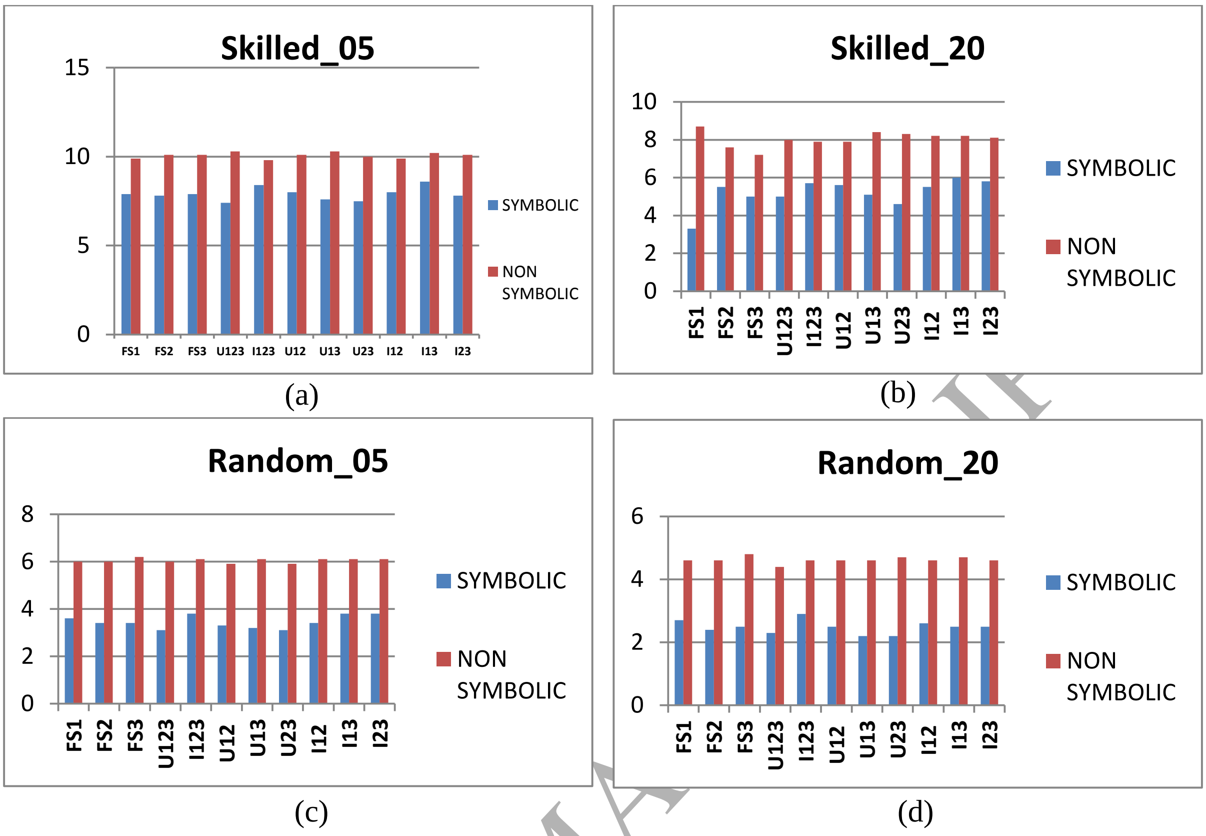


Fig. 4. EER obtained with common set of parameters for symbolic and conventional approach.
a. Skilled_05 b. Skilled_20 c. Random_05 d. Random_20

Fig. 5. EER obtained with writer dependent parameters for symbolic and conventional approach.
a. Skilled_05 b. Skilled_20 c. Random_05 d. Random_20

Table 10. Detail of Class specific parameters for first 5 writers of DB1

Writer ID	EER	Threshold	Feature Dimension	Feature Selection Method	Feature Indices
1	0.0	0.5	46	FS1	10,20,6,39,21,85,9,99,76,5,47,83,11,4,31,3,25,33,44,8,80,87,55,77,78,100, 96,92,90,69,91,54,95,23,97,28, 16,43,68,79,14,35, 74, 36, 49, 62
2	2.0	0.4	34	FS2	10,6,76,9,85,3,20,33,21,44,78,39,83,5,72,77,84,98,11,4,8,99,2,12,47,31,80, 92,15,25,30,93,13,52
3	1.7	0.4	37	FS3	3,33,85,20,21,39,44,8,76,12,99,6,98,4,5,10,9,11,2,83,55,31,94,62,100,92,91, 87,68,82,47,38,93,74,90,78,97
4	1.9	0.5	44	FS3	33,10,20,8,44,85,76,6,9,39,83,3,11,12,90,84,4,98,5,47.25,78,100,97,55,93, 67,86,70,21,43,94,49,75,30,99,79,24,15,60,13,72,28,38
5	1.7	0.4	62	FS3	6,39,20,21,9,83,78,5,3,98,85,44,4,10,11,62,92,8,33,25,74,93,89,97,47,87, 91,55,82,88,30,80,84,100,36,90,31,99,96,77,72,19,43,27,75,69,18,35,23, 37,76,54,14,65,40,58,34, 57,67,71,24,32

6. Comparative study

In this section, we compare the performance of our model with other well-known online signature verification models reported in the literature.

Table 11. EER of various online signature verification approaches on DB1

Method	Skilled_05	Skilled_20	Random_05	Random_20
1. Proposed Model				
a. With writer dependent parameters (Symbolic)	2.2	0.6	1.0	0.1
b. With Common feature dimension and threshold (Symbolic)	5.7	3.7	2.3	1.1
c. With writer dependent parameters (conventional)	5.5	3.8	3.2	1.4
d. With Common feature dimension and threshold (conventional)	9.2	5.7	5.5	3.9
2. User dependent features [21]	14.9	5.0	7.9	2.2
3. Symbolic classifier [20]	5.8	3.8	1.9	1.7
4. Cluster based symbolic representation [22]	15.4	4.2	3.6	1.2
5. Linear Programming Description(LPD) [34]	9.4	5.6	3.6	2.5
6. Principal Component Analysis Description(PCAD) [34]	7.9	4.2	3.8	1.4
7. Support Vector Description (SVD) [34]	8.9	5.4	3.8	1.6
8. Nearest Neighbour Description (NND) [34]	12.2	6.3	6.9	2.1
9. Random Ensemble of Base (RS) [35]	9.0	-	5.3	-
10. Random Subspace Ensemble with Resampling of Base (RSB) [35]	9.0	-	5.0	-
11. Base Classifier (BASE) [35]	17.0	-	8.3	-
12. Parzen Window Classifier (PWC) [34]	9.7	5.2	3.4	1.4
13. Mixture of Gaussian Description_3(MOGD_3) [34]	8.9	7.3	5.4	4.3
14. Mixture of Gaussian Description_2 (MOGD_2)[34]	8.1	7.0	5.4	4.3
15. Gaussian Model Description [34]	7.7	4.4	5.1	1.5
16. Kholmatov Model (KHA) [35]	11.3	-	5.8	-
17. Fusion model [35]	7.6	-	2.3	-
18. Regularized Parzen Window classifier RPWC [35]	9.7	-	3.4	-
19. Thumwarin et al., [45]	7.0	-	-	-
20. Quio et al., [43]	3.3	-	-	-
21. Maiorana et al., [32]			4.2	
22. Porwik et al., [42]	0.71			
23. Aguilar et al., [2]	2.12	0.55	0.24	0.00
24. Doroz et al., [12]		0.0		0.0
25. Fischer et al., [14]	3.94		1.06	

Comparison of different verification models is difficult due to variations in the dataset used for experimentation, variations in the training and testing set, different performance measures used etc. Hence for a comparative study, we consider only those models which are validated on MCYT (DB1). Table 11 shows the EER of various models along with our model.

All the models reported in Table 11 used DB1 data corpus for experimentation. Some entries in the table are filled with (–) mark as the respective authors have not reported the results for the corresponding category. Further, the model proposed by Doroz et al., (2015), used 10 genuine signatures for training while the other models have used 5 or 20 signatures for training.

From Table 11, it is noticed that our proposed model with writer dependent parameters and symbolic representation has the EER which is lower than that of most of the existing models especially in case of skilled_20 and Random_20. In case of skilled_20 and Random_20, the EER that we achieved is lowest (except Aguilar et al., 2005(a); Doroz et al., 2016). Even in case of training with 05 signatures the EER that we achieved is lower than that of all other models except Porwik et al., (2016) and Aguilar et al., 2005(a) with skilled forgery testing and lower than the other models except Aguilar et al., 2005(a) and Fischer et al., (2015) with random forgery testing.

6. Conclusion

In this paper, a new approach for online signature verification with writer dependent parameters is proposed. Writer dependent features are selected using different filter based feature selection methods and represented in the form of an interval valued symbolic feature vector. An experimentation is conducted on standard benchmarking data set. Results obtained establish the effectiveness of writer dependent parameters for signature verification and also the effectiveness of symbolic representation over conventional representation. The EER that we obtained with the writer dependent parameters and the symbolic representation is lowest when compared to that of many well-known existing models for online signature verification on the MCYT benchmarking dataset. In addition, our model works in a lower dimensional space but yet resulted in an EER, which is lowest when compared to state of the art works reported in literature.

Acknowledgement: We acknowledge the support rendered by Prof. Anil K Jain and also his help in getting MCYT dataset through Julian Fierrez Aguilar, Biometric Research Lab-AVTS, Madrid, Spain when the author D.S. Guru was doing his Postdoctoral work at PRIP Lab, Michigan State

University. Authors also thank Dr. Julian Fierrez Aguilar for his support. Author K.S. Manjunatha, thank University Grants Commission (UGC) for supporting this research under the faculty improvement programme. Further, we acknowledge the timely reviews comments by the reviewers which really helped a lot to improve significantly the quality and readability of the content of the paper.

References

[1]. Aguilar, J.F., Garcia, J. O., Ramos, D., & Rodriguez, J.G. (2007). HMM-Based On-Line Signature Verification: Feature extraction and signature modeling, Pattern recognition letters, 28 (2007), 2325-2334.

[2]. Aguilar, J.F., Nanni, L., Penalba, J. L., Garcia, J.O., & Maltoni, D. 2005(a). An On-Line Signature Verification System Based on Fusion of Local and Global Information. AVBPA. LNCS 3546. 523-532.

[3]. Aguiliar, J.F., Krawczyk, S., Garcia , J.O., & Jain, A.K. (2005(b)). Fusion of local and regional approaches for on-line signature verification. International Workshop on Biometric Recognition System (IWBRS). LNCS 3781. 188-196.

[4]. Artur, J.F., & FMáRio, A.T. (2012). Efficient feature selection filters for high dimensional data, Pattern Recognition Letters. 33. 1794-1804.

[5]. Baltzakis , H., & Papamarkos, N. (2001). A new signature verification technique based on a two-stage neural network classifier. Engineering Applications of Artificial Intelligence. 14. 95-103.

[6]. Carvalho, F.A.T. D., Csernel, M., & Lechevallier, Y. (2009). Clustering constrained symbolic data. Pattern Recognition Letters. 30. 1037-1045.

[7]. Carvalho, F.A.T.D. (2007). Fuzzy c-means clustering methods for symbolic interval data. Pattern Recognition Letters. 28. 423-437.

[8]. Christian, G., Gruber, T., Krinninger, S., & Sick, B. (2010). Online Signature Verification with Support Vector Machines Based on LCSS Kernel Functions. IEEE Trans. on Systems, Man and Cybernetics—PART B: Cybernetics. 40. 1088-1100.

[9]. Cordella, L.P., Foggia, P., Sansone, C., Tortorella, F., & Vento, M. (1999). Reliability Parameters to Improve Combination Strategies in Multi-Expert Systems. Pattern Analysis and Applications. 2. 205–214.

[10]. Cpałka, K., Zalasiński, M. (2014). On-line signature verification using vertical signature partitioning. Expert Systems with Applications. 41. 4170-4180.

[11]. Cpałka, K., Zalasi´nski, M. and Rutkowski, L. (2016). A new algorithm for identity verification based on the analysis of a handwritten dynamic signature. Applied Soft Computing. 43. 47-56.

[12]. Daliri, M. R., & Torre, V. (2008). Robust symbolic representation for shape recognition and retrieval. Pattern recognition. 41. 1782-1798.

[13]. De Carvalho, F.A.T., Bertrand, P. and Simões, E.C. (2016). Batch SOM algorithms for interval-valued data with automatic weighting of the variables. Neurocomputing. 182. 66-81.

[14]. Doroz, R., Porwik, P and Orczyk, T. (2016). Dynamic signature verification method based on association of features with similarity measures. 171. 921-931.

[15]. Enrique, A.R., & José, L.A.C. (2012). Online signature verification based on generative models. IEEE Trans. on Systems, Man, and Cybernetics—PART B: Cybernetics. 42. 1231 – 1242.

[16]. Fischer, A., Diaz§, M. and Plamondon, R. (2015). Robust Score Normalization for DTW-Based on-Line Signature Verification. ICDAR. pp. 241-245.

[17]. Garcia, O.J., Aguilar, J.F., & Simon, D. (2003). MCYT baseline corpus: A bimodal database, IEEE proceedings on vision, image and signal processing. 150. 3113-3123.

[18]. García, M. L., Lara, R. R., Hurtado, O.M. and Navarro, E.C. (2016). Embedded System for Biometric Online Signature Verification. 10. 491-501.

[1]. Giusti, A., & Grassini, L. (2008). Cluster analysis of census data using the symbolic data approach. Advances in data analysis and classification. 2. 163-176.

[2]. Gowda, K.C. and Diday, E. (1991). Symbolic clustering using a new dissimilarity measure. Pattern recognition. 24. 567-578.

[3]. Guru, D.S., & Nagendraswamy, H.S. (2006). Symbolic representation of Two-dimensional Shapes. Pattern Recognition Letters. 28. 144–155.

[4]. Guru, D.S., & Prakash, H.N. (2007). Symbolic representation of On-line signatures. International Conference on Computational Intelligence and Multimedia Applications (ICCIMA). 313-317.

[5]. Guru, D.S., & Prakash, H.N. (2009). Online signature verification and recognition: An approach based on Symbolic representation. IEEE Trans. Pattern Analysis and Machine Intelligence. 31. 1059-1073.

[6]. Guru, D.S., Manjunatha, K.S., & Manjunath, S. (2013). User dependent features in online signature verification. First International Conference on Multimedia Processing, Communication and Computing Applications (ICMCCA). LNEE 213.229-240.

[7]. Guru, D.S., Prakash, H.N., Manjunath, S. (2009). Online Signature Verification: An approach based on Cluster Representation of Global Features. Seventh International Conference on Advances in Pattern Recognition. 209-212.

[8]. Harish, B.S., Manjunath, S., & Guru, D.S. (2011). Cluster based Text Classification: A Symbolic Approach. International J. of Computer Applications. 9. 38-44.

[9]. Hedjazi, L., Lann, M., Kempowsky, T., Dalenc, F., Martin , J. A., & Favre, G. (2013). Symbolic Data Analysis to Defy Low Signal-to-Noise Ratio in Microarray Data for Breast Cancer Prognosis. J. Computational biology. 20. 610-620.

[10]. Impedovo, D., & Pirlo, G. (2008). Automatic signature verification: The state of the art. IEEE Trans. Systems Man and Cybernetics. 38. 609-635.

[11]. Jain, A.K., Griess, F.D., & Connel, S.D. (2002). On-line signature verification. Pattern Recognition. 35. 2963-2972.

[12]. Kholmatov, A., & Yanikoglu, B. (2005). Identity authentication using improved online signature verification method. Pattern Recognition Letters. 26. 2400-2408.

[13]. Kim, S.A., Park, M.S., & Kim, J. (1995). Applying Personalized Weights to a Feature Set for On-line Signature Verification. Proc. Third International Conference on Document Analysis and Recognition. 882-885.

[14]. Lei, H. and Govindaraju, V. (2005). A comparative study on the consistency of features in on-line signature verification. Pattern Recognition Letter. 26. pp. 2483-2489.

[15]. Liu, N and Wang, Y. (2008). Template selection for on-line signature verification. ICPR. 1-4.

[16]. Lumini, A., & Nanni, L. (2009). Ensemble of on-line signature matchers based on Over Complete feature generation. Expert systems with application. 36. 5291-5296.

[17]. Maiorana, E., Campisi, P., Rocca, D.L., & Scarano, G. (2012). Use of Polynomial Classifiers for On-line Signature Recognition, IEEE Fifth International Conference on biometrics: Theory, Applications and Systems (BTAS). 265-270.

[18]. Marcos, F.Z. (2007) On-line signature recognition based on VQ-DTW, Pattern Recognition, 40 (2007), 981-992.

[19]. Nanni, L. (2006). Experimental comparison of one-class classifiers for on-line signature verification. Neurocomputing. 69. 869-873.

[20]. Nanni, L., & Lumini, A. (2006). Advanced methods for two-class problem formulation for on-line signature verification. Neurocomputing. 69. 854-857.

[21]. Nanni, L., & Lumini, A. (2005). Ensemble of parzen window classifiers for on-line signature verification. Neurocomputing. 68. 217-224.

[22]. Nanni, L., Maiorana, E., Lumini, A., & Campisi, P. (2010). Combining local, regional and global matchers for a template protected on-line signature verification system. Expert Systems with Applications. 37. 3676-3684.

[23]. Neto, NAL. And De Carvalho, F.A.T., (2016). Nonlinear regression applied to interval-valued data. Pattern Analysis and Applications. 1-16.

[24]. Parodi, M., & Gomez, J.C. (2014). Legendre polynomials based feature extraction for online signature verification. Consistency analysis of feature combinations. Pattern Recognition. 47. 128-140.

[25]. Pirlo, G., Cuccovillo, V., Cabrera, M. D., Impedovo, D., & Mignone, P. (2015). Multidomain Verification of Dynamic Signatures Using Local Stability Analysis. IEEE Transactions on Human Machine-Systems. 45. 805-810.

[26]. Pirlo, G., Cuccovillo, V., Impedovo, D., Mignone, P. (2014). On-line Signature Verification by Multi-Domain Classification. 14th International Conference on Frontiers in Handwriting Recognition (ICFHR). 2014. 67-72.

[27]. Plamondon, R., & Lorette, G. (1989). Automatic signature verification and writer identification: the state of the art. Pattern Recognition. 2. 63-94.

[28]. Porwik, P., Doroz,R. and Orczyk, T. (2016). Signatures verification based on PNN classifier optimised by PSO algorithm. Pattern Recognition. 60. 998-1014.

[29]. Quio, Y., Wang, X., Xu, C. (2011). Learning Mahalonobis distance for DTW based online signature verification. IEEE international conference on Information and Automation. 333-338.

[30]. Sae-bae, N & Memon, N. (2014). Online Signature Verification on Mobile Devices. IEEE Transactions on Information Forensics and Security, 9. 933-947.

[31]. Samsone, C., & Vento, M. (2000). Signature verification: Increasing performance by a multi-stage system. Pattern analysis and application. 3. 169-181.

[32]. Thumwarin, P., Pernwong, J., Wakayaphattaramanus, N., & Matsuura, T. (2010). On-line signature verification based on FIR system characterizing velocity and direction change of barycenter trajectory. IEEE International Conference on progress in Informatics and Computing(PIC). 30-34.

[33]. Wijesoma, W.S. (2000). Selecting Optimal Personalized Features for Online Signature Verification using GA. IEEE International Conference on Systems, Man, and Cybernetics. 2740-2745.

[34]. Zhang, K., Nyssen, E., & Sahli, H. (2001). Multi-stage Online Signature Verification System. Pattern Analysis and Applications. 5. 288-295.