Writer-Independent Off-line Signature Verification using Surroundedness Feature

Rajesh Kumar[a,*], J D Sharma[b], Bhabatosh Chanda[c]

[a] *Directorate of Forensic Science, MHA, GOI, New Delhi, India*
[b] *Dr. HSG University, Sagar, M.P., India*
[c] *ECS Unit, Indian Statistical Institute, Kolkata, India*

Abstract

The paper presents a novel set of features based on surroundedness property of a signature (image in binary form) for off-line signature verification. The proposed feature set describes the shape of a signature in terms of spatial distribution of black pixels around a candidate pixel (on the signature). It also provides a measure of texture through the correlation among signature pixels in the neighborhood of that candidate pixel. So the proposed feature set is unique in the sense that it contains both shape and texture property unlike most of the earlier proposed features for off-line signature verification. Since the features are proposed based on intuitive idea of the problem, evaluation of features by various feature selection techniques has also been sought to get a compact set of features. To examine the efficacy of the proposed features, two popular classifiers namely, multilayer perceptron and support vector ma-

*Corresponding Author:- Image Processing Laboratory, ECSU, Indian Statistical Institute, 203, B.T. Road, Kolkata- 700 108, India, Tel:- +91 99032 18891, Fax:- +91 33 2577 3035

Email addresses: rajeshkumar512@gmail.com (Rajesh Kumar), jdsharma29@gmail.com (J D Sharma), chanda@isical.ac.in (Bhabatosh Chanda)

chine are implemented and tested on two publicly available database namely, GPDS300 corpus and CEDAR signature database.

Keywords: Signature verification, surroundedness, shape, texture, Feature Selection.

---

1. Introduction

Signature is one of the oldest biometric hallmarks used for authentication of an individual or a document. Even in the modern digital era, signature remains one of the popular means for the authentication of official documents like bank checks, credit card transactions, certificates, contracts and bonds. An automatic signature verification system aims to verify the identity of an individual based on the analysis of his or her signature. Depending on the mode of signature acquisition, such a system may be classified as off-line and on-line (Plamondon and Srihari, 2000). In general, on-line systems perform far better since they reckon with dynamic features extracted from the signature like time, speed, pressure and order of strokes (Nalwa, 1997). Off-line systems, on the other hand, rely only on static features (or pseudo-dynamic features) extracted from signature images. Although an efficient off-line signature verification system is comparatively difficult to design, as it fails to extract the dynamic information, its wide application in the area of forensics and biometrics has made it an active area of research.

During last few decades, several innovative concepts and technologies have been proposed and evaluated in the context of off-line signature verification, to name a few, projection and contour based systems (Bajaj and Chaudhury, 1996; Dimauro et al., 1997; Fang et al., 2003), direction profile based systems (Drouhard et al., 1996), geometric measure based approaches (Dimauro et al., 1997; Huang and Yan, 1997; Ferrer et al., 2005), grid based approaches (Huang and Yan, 1997; Madasu and Lovell, 2008), moment based approaches (Ramesh and Murthy, 1999), wavelet based approaches (Deng et al., 1999), graphometry based systems (Justino et al., 2005) and graph matching based approaches (Chen and Srihari, 2006). A survey on state-of-the-art methodologies is available in (Leclerc and Plamondon, 1994; Plamondon and Srihari, 2000; Impedovo and Pirlo, 2008).

Off-line signature verification problem may be attempted in either of the two different approaches, namely writer-dependent and writer-independent (Bertolinia et al., 2010). For the former case, the system is trained either only with the genuine signatures or with genuine and forged signatures both (depending on the model: one-class or two-class) of a particular writer. In the testing phase, the trained system (model) has to make a decision based on the (dis)similarity between the given signature and the genuine signatures of that particular writer. The most important disadvantage of a writer-dependent approach is that each time a new writer is introduced in the system the classifier has to be retrained (Oliveira et al., 2007). In writer-independent scenario also, signature verification system can be modeled either as one-class or two-class problem. But the main difference between the writer-independent and writer-dependent system is that the later makes a model for each writer separately while the former go for a generic system which can be tested on any writer. Thus a writer-independent system is more economic as far as maintenance of the classifier is concerned. However in both the two approaches, objective is the same, i.e., to differentiate a genuine signature from the forged one.

The most challenging job for an automatic signature verification system is to discriminate between genuine signature and skilled forgery (Batista et al., 2008), as it is usually done very carefully and accurately. Particularly for off-line signature it is more difficult as the dynamic features are lost. In this work, we consider only skilled forgery.

One of the main contributions of this paper is to propose a new feature set based on the surroundedness property that is experimentally found to be sufficiently representative as well as distinctive. The proposed feature set embeds in itself shape property of a signature by considering a distribution of surrounded signature (black) pixels with respect to each candidate signature pixel. It also considers the correlation between a candidate signature pixel and the other signature pixels in the neighborhood of that candidate pixel, which in turn formulates a measure of texture. Thus the proposed feature set is unique in the sense that it contains both shape and texture property unlike most of the earlier proposed features for off-line signature verification. Since the features are based on the intuitive idea of the problem, for further refinement of the feature set, here, we have employed feature analysis. An evaluation of various feature selection techniques for off-line signature verification may also be considered as a contribution of this paper. Finally, the performance evaluation of the system is done in terms of Receiver Operating Curve (ROC) (Fawcett, 2006) which provides a way to trade-off between False Rejection Rate (or FRR) and False Acceptance Rate (or FAR).

Rest of the paper is as follows: Section 2 demonstrates the signature verification scheme including preprocessing, computation of features and feature selection strategies. In Section 3, we discuss experimental results along with the experimental setup and the description of databases used. Discussion regarding performance of the proposed systems and its comparison with the state-of-the-art methodologies is placed in Section 4. Section 5 concludes the paper with the scope of future work.

2. Proposed Signature Verification Scheme

Here, we formulate off-line signature verification as a two-class pattern recognition problem. As shown in Figure 1, proposed signature verification system is based on a pair of signature images (one is authentic and the other to be tested) as input. The output of the system is one of the two classes (class-I: both the signatures are of the same writer and class-II: the signatures are of different writers). The input images (if in grey-level) are first binarized and preprocessed. Then from each preprocessed image, relevant features are extracted and the extracted feature vectors of the two signature images are combined in the 'pairing' module. The feature vectors may be combined in many different ways. Here, we have adopted one of the simplest methods, i.e., absolute difference of corresponding elements of the feature vectors as the pairing of signatures. Finally, verification is done using a two-class classifiers namely, RBF-SVM (Support Vector Machine with RBF kernel) or MLP (Multilayer perceptron). Next, we discuss preprocessing, feature extraction and analysis in detail as follows.

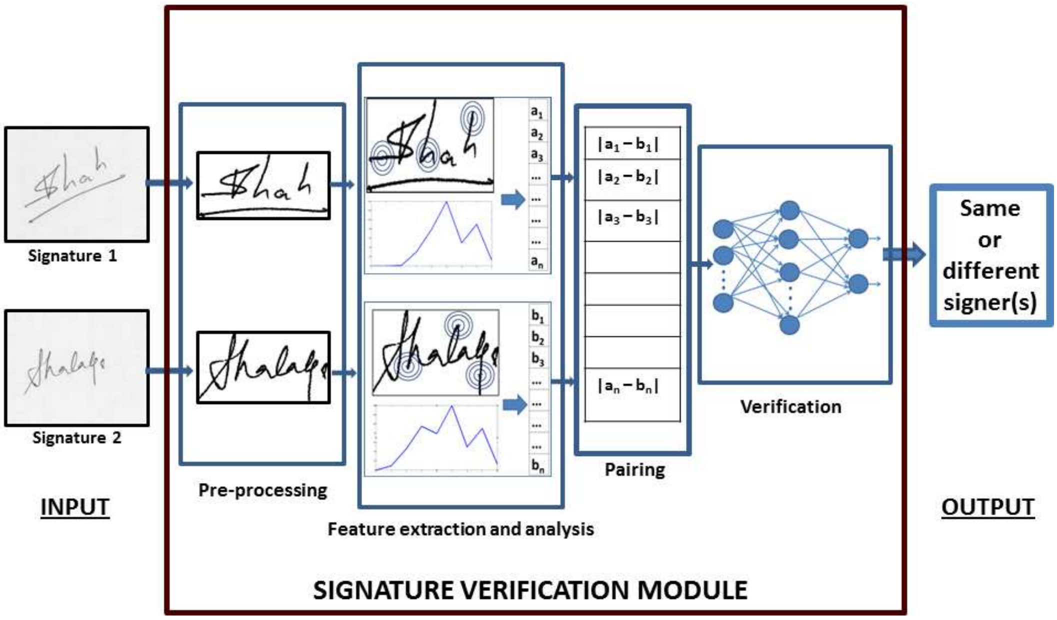


Figure 1: A schematic diagram of the proposed signature verification system. A pair of scanned signature images (grey-level/binary) is fed to the system to get an inference for their (dis)similarity. Both the images are processed individually at different stages (like preprocessing, feature extraction and analysis). Before going to a verifier (here, MLP), feature vectors corresponding to the two signatures are got paired. Decision is taken based on receiver operating curve analysis on the outputs of MLP/RBF-SVM.

2.1. Preprocessing

Preprocessing is one of the crucial stages for solving any document analysis problem. Scanned signature images may be available (as some of the images in the database being used as testbed in this experiment) in the form of grey-level images. Hence the signature images are binarized using Otsu's method (Otsu, 1979). Grain noise, if any highlighted through binarization, is cleaned using connected component analysis choosing a threshold T ($T = 7$, is selected experimentally). Thus isolated particles of size less than T pixels are removed. Since the signature might be skewed during scanning or due to some external factors during signing, we use a skew correction method based on the principal component analysis (Kalera et al., 2004). Finally, in the skew corrected image the thickness of stroke is normalized, first, by thinning and then by dilating the thinned image with a structuring element of size 2×2. Result of different preprocessing stages are shown in Figure 2.

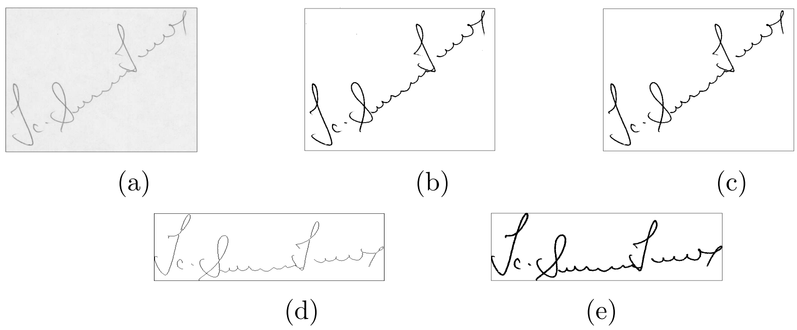


Figure 2: Results of preprocessing on an image of CEDAR database. (a) original image, (b) binarized image, (c) noise-cleaned image, (d) rotated and thinned image, (e) thinkness normalized image.

2.2. Feature extraction

In off-line signature verification, a static signature image is treated as a two dimensional arrangement of black pixels, over white background (say), having distinct shape characteristics (Bajaj and Chaudhury, 1996). An efficient feature extraction technique, therefore, should extract information like connectivity among pixels, curvilinear nature of strokes and local density of black pixels that can adequately describe the signature. Keeping these criteria in mind, we propose a feature set based on neighboring pixel surroundedness. For example, if all the eight neighbors of a candidate pixel (the black pixel for which surroundedness is being considered) are black, then the candidate pixel is called totally surrounded, on the other hand, if this number is zero then the pixel is called open. So, any number between 0 and 8 gives a measure of surroundedness at a distance one. Number of black pixels being totally varies with distance, a more formal presentation of the proposed feature follows.

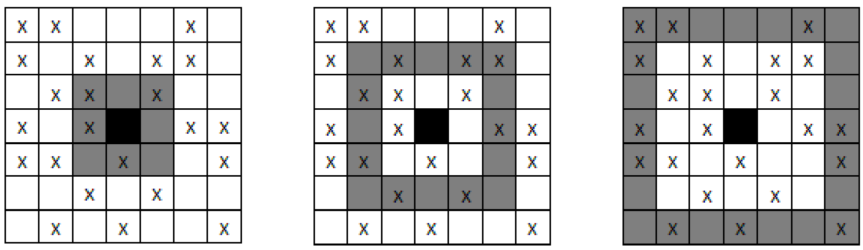
For each black pixel, we measure surroundedness at different distances. To measure surroundedness at a distance r, the number of pixels lying on the circle with radius r centering the candidate pixel are counted. The circle of radius r is determined using Chebyshev distance (Cantrell, 2000) which is a special case of Minkowski distance. Minkowski distance of order p between two points $P = (x_1, x_2, ..., x_m)$ and $Q = (z_1, z_2, , ..., z_m)$ both belong to $\mathcal{R}^m$ is defined as:

$$D_{Minkowski}(P, Q) = (\sum_{i=1}^{m} |x_i - z_i|^p)^{1/p} \tag{1}$$

In the limiting case, when p tends to ∞ , Minkowski distance becomes Chebyshev distance which is given by:

$$D_{Chebyshev}(P,Q) = \lim_{p\to\infty}(\sum_{i=1}^{n}|x_i - z_i|^p)^{1/p} = max_i|x_i - z_i| \qquad (2)$$

Under this metric in discrete domain, a circle of radius r is a square whose sides consists of $(2r+1)$ pixels. Some examples of such circles are shown in Figure 3. The black pixel at the center represents the candidate signature pixel and the grey shaded pixels surrounding it constitute the encompassing circles of different radii, e.g., $r = 1, 2$ and 3. The 'x' marks show the positions of other black signature pixels.



(a) (b) (c)

Figure 3: Circles (composed of grey pixels) in terms of Chebyshev metric in discrete domain (a) $r = 1$, (b) $r = 2$, (c)$r = 3$

For each candidate signature pixel, number n_r of signature pixels at Chebyshev distance r is counted, where $0 \leq n_r \leq 8r$. For example, it can be seen from Figure 3 that around the center pixel, number of pixels at distance 1 is 4, at distance 2 is 8 and at distance 3 is 11. Thus n_r provides a measure of surroundedness. Now, for the entire image, frequency distribution of n_r is calculated for different values of r.

Here, we consider the correlation of signature pixels (black pixels) only at different distances because the distribution of background pixel (white pixel) is complementary to the signature pixel. Thus the proposed feature is a special case of color correlogram (Huang et al., 1999) where only one color (black) is considered. On the other hand, the proposed feature is more general in a sense that it provides frequency distribution, while correlogram provides with the total number of pixels of a particular distance. Color correlogram or auto-correlogram, however, reveals the texture feature of the image, so does the proposed feature but in more detail. The proposed feature set also bears some similarity to shape context (Belongie et al., 2002) for object recognition and shape matching. Shape context considers 30° non overlapping bins at various distance ranges in a logarithmic-polar space, whereas the proposed feature considers whole 360° at a time with a linear scale for distance. In this sense the proposed feature is a special case of shape context feature. Like the later, the proposed feature also represents the shape characteristics of a signature. On the other hand, the proposed feature is different from the shape context feature, as the former provides in general the frequency distribution of surroundedness while the later reveals the count of black pixels at different locations or bins. A comparative results for these three feature sets, namely color correlogram, shape context and the proposed one are placed in Section 4.

In our experiment, we have used $r = 1, 2, \cdots, 11$, so, we have too many feature values with less useful minute details. Hence, to reveal more general and representative characteristics, we summarize them using some statistical measures. For each discrete distance, the following four statistical measures are calculated using each of the frequency distribution tables.

- Entropy: $E = -\sum_i p_i(\log_2 p_i)$, where $p_i = \frac{f_{ri}}{\sum_{i=0}^{i=8r} f_{ri}}$ = normalized frequency and f_{ri} is the frequncy of i^{th} bin at r distance.
- First order moment : $M_1(\bar{n}) = \sum_i n_i p_i$
- Second order moment : $M_2(\sigma) = \sqrt{\sum_i p_i(n_i - \bar{n})^2}$
- Third order moment : $M_3 = \sqrt[3]{\sum_i p_i(n_i - \bar{n})^3}$

It is experimentally observed that higher moments, of more than third order, does not provide much additional information. Thus $11 \times 4 = 44$ features are extracted initially from each signature image. Number of features is finally reduced by retaining only relevant features through feature analysis as described in Section 2.3.

2.3. Feature Analysis

It is well known that use of more features is not necessarily good because there may be some derogatory as well as redundant/correlated features in a given set. Use of a set of just adequate features usually makes the system identification easier and such a system is likely to yield better generalization. A small set of features also requires less memory and usually reduces the computational costs. The aim of feature selection is to find a subset (as small as possible) of features while simultaneously optimizing the performance of the system. We employ feature analysis technique for the same.

Usually, feature selection techniques may be divided broadly into two categories : filter and wrapper methods (Guyon and Elisseeff, 2003; Theodoridis and Koutroumbas, 2009). A filter method looks at each feature independently and rank them based on some criteria (e.g., correlation coefficient, mutual information, hypothesis testing etc.). One of the main disadvantages of these methods, as they look at each feature independently, lie in the fact that they may yield into a number of redundant features. At the same time, filter methods are comparatively faster as well as robust against over-fitting. Wrapper methods, on the other hand, utilize the induction algorithm as black box to score subsets of features according to their predictive power. These methods are often criticized due to its massive computation. There is a third category of feature selection techniques, called embedded methods, where feature selection is performed in the process of training and are usually specific to the learning mechanism (Guyon and Elisseeff, 2003).

Here, we also evaluate various feature selection techniques, using the proposed features, for off-line signature verification problem. For this analysis, we consider two filter methods namely, T-test based method (Theodoridis and Koutroumbas, 2009) and entropy based method (Li and Wong, 2001), three wrapper methods namely, Sequential Forward Selection (SFS), Sequential Backward selection (SBS) and Sequential Forward Floating Selection (SFFS) (Theodoridis and Koutroumbas, 2009) and one embedded method namely, FSMLP (Pal and Chintalapudi, 1997; Kumar et al., 2009).

The feature selection is done on the training data and the data with the selected features are fed to the classifier to get the inference. Unlike the other five feature selection techniques, FSMLP is run 10 times on the training data as the feature selected by FSMLP depends on the initialization condition. Each run R of FSMLP generates a gate opening value for each feature, f as $g_f^R; R = 1, \cdots, 10; f = 1, \cdots, 44$. Now, we compute the composite importance for feature f as $g_f = \sum_{R=1}^{10} g_f^R$ and use these values to select a set of features. A comparative analysis of these six feature selection methodologies have been described in Section 3.2.

3. Experimental Results

3.1. Experiment Design

Experiments are carried out on two widely used databases, namely GPDS300 corpus and CEDAR signature database, separately. Description of data and experimental setup are as follows.

3.1.1. GPDS300 signature corpus

GPDS300 corpus[1] (Vargas et al., 2007) contains 24 genuine signatures and 30 forgeries of each of 300 individuals. So, there are 7200 genuine signatures and 9000 forged signatures. The 24 genuine specimens of each signer were collected in a single day writing sessions. The genuine signatures are shown to

[1]GPDS300 corpus is available on http://www.gpds.ulpgc.es/download/ index.htm/.

each forger and are chosen randomly from the 24 genuine ones to be imitated. All the signature images in this corpus are in binary form.

Since the database is large, we divide it equally to make training and test set. That means genuine signatures as well as corresponding forgeries of randomly selected 150 individuals are used for training and the rest is used for testing. Since for each individual there are 24 genuine and 30 forged signatures, we get ${}^{24}C_2 = 276$ genuine-genuine pairs of signatures and ${}^{24}C_1 \times {}^{30}C_1 = 720$ genuine-forged pairs of signatures for each individual. Since the number of genuine-forged pairs is significantly larger than that of genuine-genuine pairs, we randomly select as many genuine-forged pairs as genuine-genuine pairs to avoid any bias towards any class during training as well as verification. Thus each of the training and test data set contains $2 \times 150 \times 276 = 82800$ signature pairs. Pairing is done as an absolute difference of corresponding elements in the feature vector of the two signatures of each pair.

3.1.2. CEDAR signature database

CEDAR database[2] contains signatures of 55 volunteer signers belonging to versatile cultural backgrounds. Each writer signed 24 genuine signatures, taken 20 minutes apart. Each of the forgers simulated the signatures of 3 persons, 8 times each, to produce 24 forged signatures. Thus for each genuine signer, 24 forged signatures are generated. The database comprises of 1320

[2]CEDAR Signature database is available on http://www.cedar.buffalo.edu/NIJ/publications.html/.

genuine and 1320 forged signatures corresponding to 55 signers (Chen and Srihari, 2006). This database contains the signature images in both grey-level and binary forms.

Since the CEDAR database contains signatures of only 55 persons, we select 5 individuals randomly and keep them aside as test data and the signatures of remaining 50 individuals comprise the training data. Since for each individual there are 24 genuine and 24 forged signatures, we get ${}^{24}C_2 = 276$ genuine-genuine pairs and ${}^{24}C_1 \times {}^{24}C_1 = 576$ genuine-forged pairs of signatures for each individual. Similarly, as discussed for GPDS300 corpus, we get the number of training samples equals to $2 \times 50 \times 276 = 27600$ and the number of test samples equal to $2 \times 5 \times 276 = 2760$. Since 5 individuals are selected randomly for testing, to avoid the effect of selection, we repeat the experiment 10 times and report the average performance.

3.1.3. Model Selection

Selection of optimal parameters for MLP (or RBF-SVM) is always been a challenging task. For this particular problem, a 10-fold cross-validation is done on training data with a set of choices for number of hidden nodes (regularization parameter C and σ of a RBF-kernel, in case of RBF-SVM (Theodoridis and Koutroumbas, 2009)) to get the optimal parameters. In case of MLP, we take only one hidden layer. The optimal parameters obtained from cross-validation is used to design the MLP (or RBF-SVM) using the entire training data. The trained network is then tested with the test data.

In MLP, with sigmoid function as activation function, the output of the network lies in the range [0,1]. FAR, FRR and the accuracy are very much dependent on suitable selection of threshold values. ROC analysis (Fawcett, 2006) is consulted to get a trade-off between FAR and true positive rate (1-FRR) by choosing a suitable threshold. We draw an ROC curve for different number of best selected features and determine the equal error rate (EER). Similarly for RBF-SVM, the probability of a data to belong to a class is considered for ROC analysis. In this paper, we always measure accuracy of the proposed system at EER, which is nothing but (100-EER).

3.2. Performance of the classifiers on different databases

Figure 4 depicts performance of the classifiers using six types of feature selection techniques mentioned in Section 2.3. Figure 4 (a) and 4 (b) shows a comparison of MLP and RBF-SVM for FSMLP selected features on CEDAR and GPDS 300 database respectively. One can see from the figure that MLP and RBF-SVM are giving more or less similar performance. Since MLP gives the best performance (using 34 features for CEDAR and 29 features for GPDS 300) using FSMLP selected features, to show the comparative analysis of other feature selection techniques, we display the performance of MLP in Figure 4 (c) and 4 (d). From both the figures, one can see that in general wrapper method (SFFS, SFS and SBS) gives the best performance followed by embedded method (FSMLP). If we look at wrapper methods, all the three methods (SFFS, SFS and SBS) give similar performances. Although all the

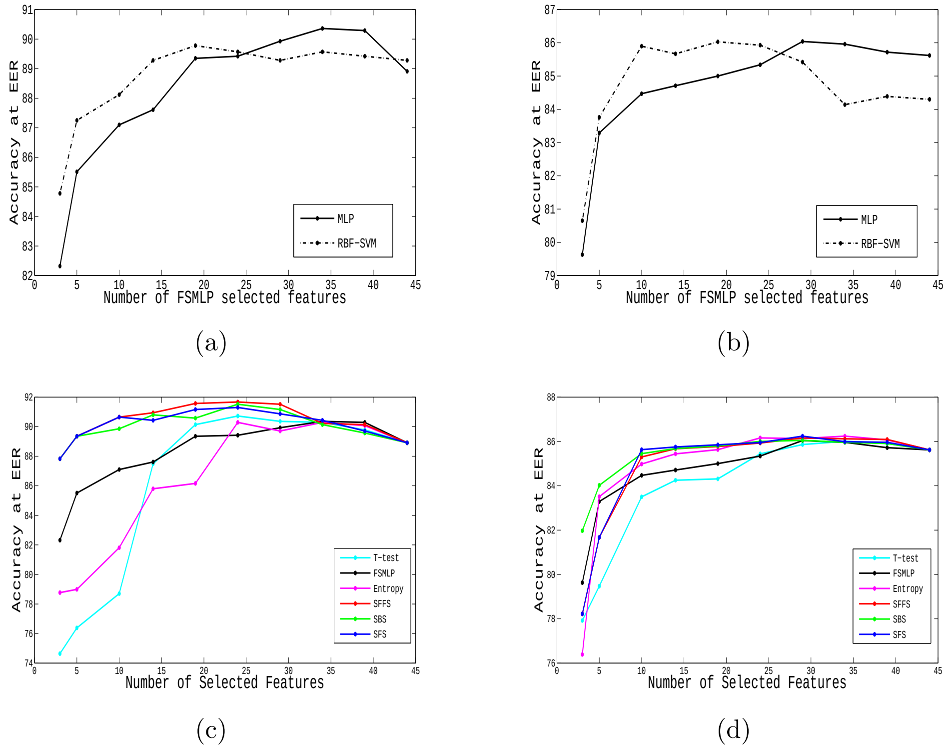
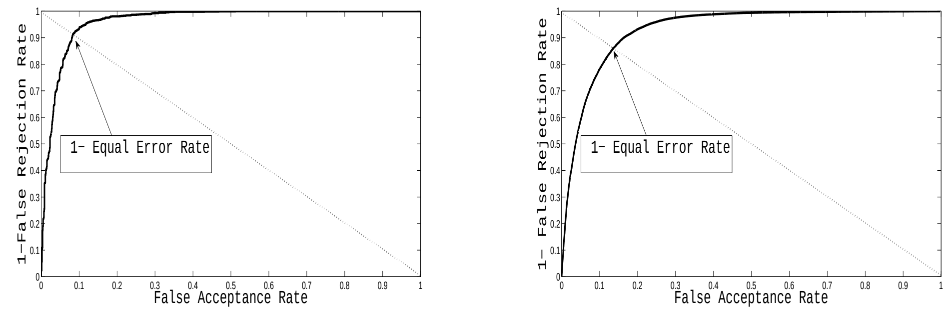


Figure 4: A graphical representation of performance with different numbers of selected features. Comparison of RBF-SVM and MLP for FSMLP selected features on (a) CEDAR database and (b) GPDS300 corpus. Comparison of performances of MLP with various feature selection techniques on (c) CEDAR database and (d) GPDS300 corpus.

six feature selection techniques perform more or less in the same fashion, best performance (an accuracy of 91.67%) on CEDAR is achieved with just 24 SFFS selected features while best performance (an accuracy of 86.24%) on GPDS 300 is achieved with 29 SFS selected features.

For better understanding, in Figure 5, we show ROC analysis on 29 features selected through SFFS for both the databases. In the figure, one can see the variation of FAR with true positive rate (1-FRR) obtained by varying threshold. The point of intersection of ROC curve with the diagonal joining

(1,0) and (0,1) represents the accuracy at EER.



(a) (b)

Figure 5: A typical ROC curve for 29 best selected features by SFFS for (a) CEDAR database, and (b) GPDS300 corpus.

A careful study of both Figure 4 (c) and 4 (d) reveals the fact that the proposed features perform in similar fashion on both the databases while number of selected features is varied between 20 and 30 (specially in case of wrapper methods). Variation in accuracies in this range is almost insignificant for both the databases, which gives a hint that the feature set does not have derogatory features rather there may be some redundant (correlated) features. Incorporation and deletion of such features may result in little variation in accuracy according to their interactions among themselves. To further analyze the redundancy (correlation) between the features, we implement Random Subspace ensemble (RSE), mentioned in (Nanni and Lumini, 2005), taking 29 random feature 10 times. RSE gives an accuracy of 90.71% and 85.31% for CEDAR and GPDS300 database respectively which is similar to the result of a single set of features selected by any of the six feature selection techniques. The Results obtained from RSE also verifies the

Table 1: A comparative analysis of performances of similar features

Feature	Accuracy at EER	
	CEDAR database	GPDS 300 corpus
Black correlogram [3]	71.38	67.09
Shape context	65.94	61.62
Proposed feature	91.67	86.24

redundancy of the features.

4. Discussions

In Section 2.2, we observe that the proposed feature set has got some similarity with color correlogram and shape context feature. Here, we compare the performance of those features for the targeted task. Corresponding results using GPDS 300 and CEDAR databases, along with that of the proposed feature set through feature analysis, are shown in Table 1, which establishes superiority of the proposed feature set. The reason may be incorporation of both local texture and shape information to some extent simultaneously into the proposed feature set.

Results of the proposed system show a consistency on both the databases as far as accuracy is concerned. For CEDAR and GPDS300 databases, we get an accuracy of 91.67% and 86.24% respectively. One should keep in mind that the number of test samples in CEDAR database (=2760 without considering repetition) is much less than the test samples in GPDS300 corpus

[3]Here, color correlogram (Huang et al., 1999) is calculated on black pixels (only) and hence termed as black correlogram.

(=82800). Thus size of the databases affects the performance to some extent (but not much) due to variability in data. Due to lack of a standard database of signature, it is very difficult to compare different signature verification systems as people use different databases. Even on the same database, people follow different strategies to respect accuracy. For the sake of comparative evaluation, we compare the performances of the proposed system with that of other systems on both GPDS and CEDAR signature database, reported in literature, as follows.

4.1. CEDAR signature database

Performance of different systems reported on the CEDAR database is summarized in Table 2. From the table, it is clear that the proposed system performs better than the word shape (GSC) (Kalera et al., 2004), Zernike moment (Chen and Srihari, 2005) and signature morphology (Kumar et al., Dec. 2010). It is also comparable to graph matching system (Chen and Srihari, 2006) which is comparatively very expensive. One of the appealing features about our system is that it uses just 24 features which is significantly less in compared to the number of features used by the other systems. Less number of features makes our system computationally inexpensive compared to the other systems. This characteristics is very much suitable for real-time signature verification system.

Though the accuracy of the graph matching system (Chen and Srihari, 2006) is 0.4 % more than that of the proposed system, it uses lots of com-

Table 2: Comparison of the proposed system with the state-of-the-art on CEDAR database and GPDS signature corpus

Databases	SYSTEMS	No. of signers	Accuracy	FAR	FRR
CEDAR signature database	Word shape(GSC) (Chen and Srihari, 2006; Kalera et al., 2004)	55	78.50	19.50	22.45
	Zernike moments (Chen and Srihari, 2005, 2006)	55	83.60	16.30	16.60
	Signature Morphology (Kumar et al., Dec. 2010)	55	88.41	11.59	11.59
	Graph Matching (Chen and Srihari, 2006)	55	92.10	08.20	07.70
	Proposed system	55	91.67	08.33	08.33
GPDS corpus	Ferrer et al. (Ferrer et al., 2005)	160	86.65	12.60	14.10
	Vargas et al. (Vargas et al., 2008)	160	87.67	14.66	10.01
	Solar et al. (Solar et al., 2008)	160	84.70	14.20	16.40
	Proposed system	300	86.24	13.76	13.76

putationally expensive operations. First, they extract chain-code of exterior and interior contours (a complex algorithm of order N^2, where N is number of signature pixels) followed by locating extrema on these contours and categorizing them into 16 parts. Once the extrema have been located, they go for point matching method based on eigen-analysis of proximity matrix (again a complex algorithm with greater than $(4 \times m^2 \times n + 8 \times m \times n^2 + 9 \times n^3)$ operations where m and n are number of extrema in two signatures). Then they calculate bending energy using thin-plate spline warping (another expensive algorithm of order $(n^3 + n^2)$). Further, 1024 dimensional GSC features are calculated to get dissimilarity between two signatures and final decision is given on the weighted sum of bending energy and dissimilarity score of GSC algorithm. On the other hand, the proposed algorithm uses black pixel counting and arithmetic operators while extracting the feature and use a trained MLP (or RBF-SVM) network to give the final decision. Thereby restrict the order of complexity to linear (in terms of number of signature pixels).

Second, unlike the proposed system (except signature morphology), all the three systems reported in the Table 2 are writer-dependent. Basically, in all those systems a model for *each* writer is made and the opinion over unknown signature is given based on the threshold determined for each model separately. If a new signer (client) is added, these systems have to be updated. On the other hand, the proposed system can be used even for any newly added signer without re-training the system. This can be seen as one of the major advantage of the proposed system over those three systems.

4.2. GPDS300 signature corpus

Table 2 also shows the performance of different systems on GPDS signature database with different numbers of signing individuals. One can see from the table that the accuracy of the proposed system is comparable (numerically) to the earlier reported systems, while the difficulty level the proposed system is dealing with is much higher. First, the other systems used earlier version of the GPDS signature corpus containing less number of signers (only 160 individuals), that means less variety. Moreover, the results of these systems are not reported at EER. Second, systems presented in (Vargas et al., 2008) and (Solar et al., 2008) have reported the combined accuracy of random forgery (which is comparatively much simpler to detect) and skilled forgery. (Armand et al., 2006) has reported an accuracy of 91.12% on partial GPDS signature corpuses (39 writers, 1560 signatures for training and 576 signatures for testing) without giving any information regarding FAR and FRR . On the other hand, we report the performance of our system for skilled forgery only and that too on all 300 writers (82,800 signature pairs for training and 82,800 for testing). Experimentally, we have seen that the best accuracy of the proposed system may go up to 93.46% for some randomly selected 160 signers out of 300 signers, which is significantly higher than the earlier reported systems. Since we are working on a reasonably large database, the performance of the proposed system seems to be quite stable and better than the results of other systems on GPDS corpuses.

5. Conclusion and Future work

In this paper, we propose a writer-independent off-line signature verification scheme based on surroundedness features extracted from the binary image of signatures. The feature set based on surroundedness property of a signature, is supposed to represent both shape and texture attributes of the signature. From the results obtained, it is clear that the proposed feature set has got some edge over related features like shape context and auto-correlogram. A subset of features (20-30 in numbers) selected by any of the six feature selection methodologies performs with almost similar accuracy. This establishes the fact that features are not derogatory rather may be redundant ensuring robustness against over-fitting. As far as performance on both the CEDAR and GPDS300 databases is concerned, it has been seen that the proposed system is superior either in terms of accuracy or the time complexity or both, when compared to the state-of-the-art methodologies.

Moreover, the performance of the proposed system on both GPDS300 and CEDAR database is comparable, which indicates that the proposed approach and the feature set are sufficiently general to handle data of varied standards.

Proposal of a new set of features, utilization of other classifiers and fusion of different classifiers may indeed be a future scope for further improvement of the system.

Acknowledgements

The first author would like to thank Directorate of Forensic Science, Ministry of Home Affairs, Govt. of India for the financial support.

References

Armand, S., Blumenstien, M., Muthukkumarasamy, V., Sept. 2006. Off-line Signature Verification based on the Modified Direction Feature. In: Proceedings of the International Conference on Pattern Recognition (ICPR'06). pp. 509–512.

Bajaj, R., Chaudhury, S., April 1996. Signature Verification using Multiple Neural Classifiers. Pattern Recognition 3 (1), 1–7.

Batista, L., Rivard, D., Sabourin, R., Granger, E., Maupin, P., 2008. Pattern Recognition Technologies and Applications: Recent Advances. IGI Global snippet, Ch. III:state of the art in off-line signature verification, pp. 39–62.

Belongie, S., Malik, J., Puzicha, J., April 2002. Shape Matching and Object Recognition Using Shape Contexts. IEEE Transaction on Pattern Analysis and Machine Intelligence 24 (24), 509–522.

Bertolinia, D., Oliveira, L. S., Justinoa, E., Sabourinc, R., Jan. 2010. Reducing Forgeries in Writer-Independent Off-line Signature Verification through Ensemble of Classifiers. Pattern Recognition 43 (1), 387–396.

Cantrell, C. D., 2000. Modern Mathematical Methods for Physicists and Engineers. Cambridge University Press.

Chen, S., Srihari, S., 2005. Use of Exterior Contours and Shape Features in Off-line Signature Verification. In: Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'05). pp. 1280–1284.

Chen, S., Srihari, S., 2006. A New Off-line Signature Verification Method based on Graph Matching. In: Proceedings of the International Conference on Pattern Recognition (ICPR'06. Vol. 2. pp. 869–872.

Deng, P. S., Liao, H. Y. M., Ho, C. W., Tyan, H. R., 1999. Wavelet based offline handwritten signature verification. Computer vision and image understanding. 76 (3), 173–190.

Dimauro, G., Impedovo, S., Pirlo, G., Salzo, A., 1997. A multi-expert signature verification for bankcheck processing. International Journal on Pattern Recognition and Artificial Intelligence 11 (5), 827–844.

Drouhard, J. P., Sabourin, R., Godbout, M., 1996. A neaural network approach to offline signature verification using direction pdf. Pattern Recognition 29 (3), 415–424.

Fang, B., Leung, C. H., Tang, Y. Y., Tse, K. W., Kwok, P. C., Wong, Y. K., 2003. Offline signature verification by tracking of feature and stroke positions. Pattern Recognition 36 (1), 91–101.

Fawcett, T., 2006. An Introduction to ROC Analysis. Pattern Recognition Letters 27, 861–874.

Ferrer, M., Alsono, J., Travieso, C., 2005. Off-line Geometric Parameters for Automatic Signature Verification using Fixed Point Arithmatic. IEEE Transaction on Pattern Analysis and Machine Intelligence 27 (6), 993–997.

Guyon, I., Elisseeff, A., 2003. An Introduction to Variable and Feature Selection. Journal of Machine Learning Research 3, 1157–1182.

Huang, J., Kumar, S. R., Mitra, M., Zhu, W.-Z., Zabih, R., 1999. Spatial Color Indexing and Application. International Journal of Computer Vision 35 (3), 245–268.

Huang, K., Yan, H., 1997. Offline signature verification based on geometric feature extraction and neaural network classification. Pattern Recognition 30 (1), 9–17.

Impedovo, D., Pirlo, G., Sept. 2008. Automatic Signature Verification: the state of the art. IEEE Transaction on System Man and Cybernatics-Part C 38 (5), 609–635.

Justino, E. J. R., Bortolozzi, F., Sabourin, R., Dec. 2005. An offline signature verification based on svm classifier and graphometric features. In: Proceedings of the InternationaL Conference on Advances on Pattern Recognotion (ICAPR'05). pp. 134–141.

Kalera, M. K., Srihari, S., Xu, A., 2004. Off-line Signature Verification and Identification using Distance Statistics. International Journal on Pattern Recognition and Artificial Intelligence 18 (7), 1339–1360.

Kumar, R., Kundu, L., Sharma, J. D., Chanda, B., Dec. 2010. A Writer Independent Off-line Signature Verification System based on Signature Morphology. In: Proceedings of the IITM'2010. ACM, New York, USA, pp. 261–265.

Kumar, R., Pal, N. R., Sharma, J. D., Chanda, B., December 2009. A Novel Approach for Detection of Alteration in Ball Pen Writings. In: Lecture Notes in Computer Science,PReMI. Springer, pp. 400–405.

Leclerc, F., Plamondon, R., June 1994. Automatic Signature Verification: the state of the art–1989-1993. International Journal on Pattern Recognition and Artificial Intelligence 8 (3), 643–660.

Li, J., Wong, L., 2001. Emerging Patterns and Gene Expression Data. Genome Informatics 12, 3–13.

Madasu, V. K., Lovell, B., 2008. Pattern recognition technologies and applications: recent advances. IGI Global snippet, Ch. IV:An Automatic Off-line Signature Verification and Forgery Detection System, pp. 63–89.

Nalwa, V. S., Feb. 1997. Automatic On-line Signature Verification. In: Proceedings of the IEEE. Vol. 85. pp. 215–239.

Nanni, L., Lumini, A., 2005. Ensemble of Parzen Window Classifiers for Online Signature Verification. NeuroComputing 68 (6), 217–224.

Oliveira, L. S., E.Justino, R.Sabourin, Aug. 12-17 2007. Off-line Signature Verification using Writer-Independent Approach. In: Proceedings of Internation Joint Conference on Neural Networks'07. Orlando,Florida, pp. 2539–2544.

Otsu, N., 1979. A Threshold Selection Method from Gray-level Histogram. IEEE Transaction on System, Man and Cybernatics 6, 62–66.

Pal, N. R., Chintalapudi, K. K., 1997. A Connectionist System for Feature Selection. Neural, Parallel and Scientific Computations 5, 359–382.

Plamondon, R., Srihari, S. N., Jan. 2000. On-line and Off-line Handwriting Recognition:A Comprehensive Survey. IEEE Transaction on Pattern Analysis and Machine Intelligence 22 (1), 63–84.

Ramesh, V. E., Murthy, M. N., 1999. Offline signature verification using genetically optimized weighted features. Pattern Recognition 32 (2), 217–233.

Solar, J. R., Devia, C., Loncomilla, P., Concha, F., Sept. 2008. Off-line Signature Verification using Local Interest Points and Descriptors. In: Proceeding of the 13th Iberoamerican Congress on Pattern Recognition (CIARP'08). pp. 22–29.

Theodoridis, S., Koutroumbas, K., 2009. Pattern Recognition, 4th Edition. Academic Press, San Diego.

Vargas, F., Ferrer, M., Travieso, C., Alsono, J., Sept. 2007. Off-line Handwritten Signature GPDS-960 Corpus. In: Proceedings of the International conference on Document Analysis and Recognition (ICDAR'07). pp. 764–768.

Vargas, J. F., Ferrer, M. A., Travieso, C. M., Alsono, J. B., August. 2008. Offline Signature Verification based on High Pressure Polar Distribution. In: Proceedings of the International Conference on Frontiers of Handwriting Recognition (ICFHR'08). pp. 1–6.