Multi-script Off-line Signature Verification: A Two Stage Approach

Srikanta Pal
School of Information and
Communication Technology,
Griffith University, Gold Coast
Australia, Email:
srikanta.pal@griffithuni.edu.au

Umapada Pal
Computer Vision and Pattern
Recognition Unit, Indian Statistical
Institute, Kolkata, India,
Email: umapada@isical.ac.in

Michael Blumenstein
School of Information and
Communication Technology,
Griffith University, Gold Coast,
Australia, Email:
m.blumenstein@griffith.edu.au

Abstract—**Signature identification and verification are of great importance in authentication systems. The purpose of this paper is to introduce an experimental contribution in the direction of multi-script off-line signature identification and verification using a novel technique involving off-line English, Hindi (Devnagari) and Bangla (Bengali) signatures. In the first evaluation stage of the proposed signature verification technique, the performance of a multi-script off-line signature verification system, considering a joint dataset of English, Hindi and Bangla signatures, was investigated. In the second stage of experimentation, multi-script signatures were identified based on the script type, and subsequently the verification task was explored separately for English, Hindi and Bangla signatures based on the identified script result. The gradient and chain code features were employed, and Support Vector Machines (SVMs) along with the Modified Quadratic Discriminate Function (MQDF) were considered in this scheme. From the experimental results achieved, it is noted that the verification accuracy obtained in the second stage of experiments (where a signature script identification method was introduced) is better than the verification accuracy produced following the first stage of experiments. Experimental results indicated that an average error rate of 20.80% and 16.40% were obtained for two different types of verification experiments.**

Keywords—Biometrics; off-line signature verification; multi-script signature identification.

I. INTRODUCTION

Biometrics are the most widely used approaches for personal identification and verification. Among all of the biometric authentication systems, handwritten signatures, a pure behavioral biometric, have been accepted as an official means to verify personal identity for legal purposes on such documents as cheques, credit cards and wills [1].

In general, automated signature verification is divided into two broad categories: static (off-line) methods and dynamic (on-line) methods [2], depending on the mode of handwritten signature acquisition. If both the spatial as well as temporal information regarding signatures are available to the systems, verification is performed using on-line [3] data. In the case where temporal information is not available and the system can only utilize spatial information gleaned through scanned or even camera-captured documents, verification is performed on off-line data [4].

Considerable research has previously been undertaken in the area of signature verification, particularly involving single-script signatures. On the other hand, less attention has been devoted to the task of multi-script signature verification. Very few published papers involving multi-script signatures, including non-English signatures, have been communicated in the field of signature verification.

Pal et al. [5] introduced a signature verification system employing Hindi Signatures. The direction of the paper was to present an investigation of the performance of a signature verification system involving Hindi off-line signatures. In that study, two important features such as: gradient feature, Zernike moment feature and SVM classifiers were employed. Encouraging results were obtained in this investigation. In a different contribution by Pal et al. [6], a multi-script off-line signature identification technique was proposed. In that report, the signatures involving Bangla (Bengali), Hindi (Devnagari) and English were considered for the signature script identification process. A multi-script off-line signature identification and verification approach, involving English and Hindi signatures, was presented by Pal et al. [7]. In that paper, the multi-script signatures were identified first on the basis of signature script type, and afterward, verification experiments were conducted based on the identified script result.

Development of a general multi-script signature verification system, which can verify signatures of all scripts, is very complicated. The verification accuracy in such multi-script signature environments will not be as successful when compared to single script signature verification [10]. To achieve the necessary accuracy for multi-script signature verification, it is important to identify signatures based on the type of script and then use an individual single script signature verification system for the identified script [10]. Based on this observation, in the proposed system, the signatures of three different scripts are separated to feed into the individual signature verification system. On the other hand to get a comparative idea, multi-script signature verification results on a joint English, Hindi and Bangla dataset, without using any script identification, is also investigated.

The remainder of this paper is organized as follows. The multi-script signature verification concept is described in Section II. Section III introduces the notable properties of Hindi and Bangla script. The Hindi, Bangla and English signature database used for the current research is described in Section IV. Section V briefly presents the feature extraction techniques employed in this work. The classifier details are described in Section VI. The experimental

settings are presented in Section VII. Results and a discussion are provided in Section VIII. Finally, conclusions and future work are discussed in Section IX.

II. MULTI-SCRIPT SIGNATURE VERIFICATION CONCEPT

When a country deals with two or more scripts and languages for reading and writing purposes, it is known as a multi-script and multi-lingual country. In India, there are officially 23 (Indian constitution accepted) languages and 11 different scripts.

In such a multi-script and multi-lingual country like India, languages are not only used for writing/reading purposes but also applied for reasons pertaining to signing and signatures. In such an environment in India, the signatures of an individual with more than one language (regional language and international language) are essentially needed in official transactions (e.g. in passport application forms, examination question papers, money order forms, bank account application forms etc.). To deal with these situations, signature verification techniques employing single-script signatures are not sufficient for consideration. Therefore in a multi-script and multi-lingual scenario, signature verification methods considering more than one script are necessarily required.

Towards this direction of verification, the contribution of this paper is twofold: First, multi-script signature verification considering joint datasets as shown in Figure 1, the second is identification of signatures based on script, and subsequent verification for English, Hindi and Bangla signatures based on the identified script result. A diagram of this second verification mode is shown in Figure 2.
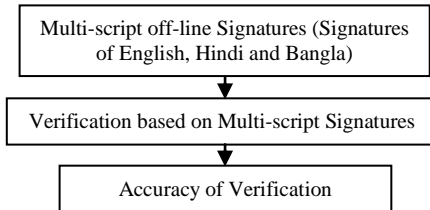
Multi-script off-line Signatures (Signatures of English, Hindi and Bangla)

Verification based on Multi-script Signatures

Accuracy of Verification

Figure 1. Diagram of signature verification considering a joint dataset.

Multi-script Signatures (English, Hindi and Bangla)

Signature Script Identification

Signatures of English Script

Signatures of Hindi Script

Signatures of Bangla Script

English Signature Verification

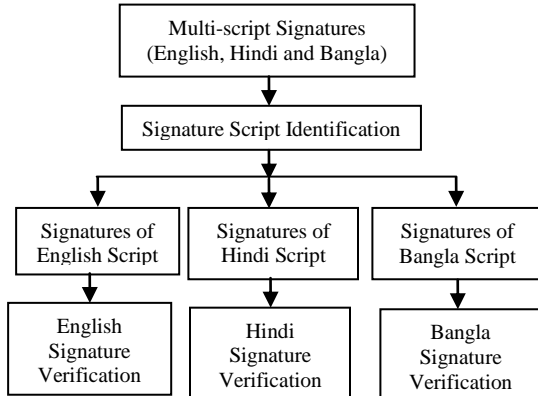Hindi Signature Verification

Bangla Signature Verification

Figure 2. Diagram of multi-script signature identification and verification based on English, Hindi and Bangla signatures.

III. PROPERTIES OF HINDI AND BANGLA SCRIPT

Most of the Indian scripts including Bangla and Devanagari have originated from ancient Brahmi script through various transformations and evolution [8]. Bangla and Devanagari are the two most accepted scripts in India. In both scripts, the writing style is from left to right and there is no concept of upper/lower case. These scripts have a complex composition of their constituent symbols. The scripts are recognizable by a distinctive horizontal line called the 'head line' that runs along the top of full letters, and it links all the letters together in a word. Both scripts have about fifty basic characters including vowels and consonants.

IV. DATABASE USED FOR EXPERIMENTATION

A. Hindi and Bangla Signature Database

As there has been no public signature corpus available for Hindi and Bangla script, it was necessary to create a database of Hindi and Bangla signatures. The Hindi and Bangla signature databases used for experimentation consisted of 50 sets per script type. Each set consists of 24 genuine signatures and 30 skilled forgeries. Some genuine signature samples of Hindi and Bangla, with their corresponding forgeries, are displayed in Table 1 and Table 2.

B. GPDS English Database

Another database, consisting of 50 sets from GPDS-160 [9], was also utilised for these experiments. Each signature set of this corpus consists of 24 genuine signatures and 30 simple forgeries. The reason 50 sets were used from the GPDS on this occasion, is due to the fact that the Bangla and Hindi datasets described previously were comprised of 50 sets each, and it was considered important to have equivalent signature numbers for experimentation.

TABLE 1. SAMPLES OF HINDI GENUINE AND FORGED SIGNATURES

Genuine Signatures	Forged signatures
नीन पटनायक	जीन पटनायक
अद्विती भुखर्जी	अद्विती मुखर्जी

TABLE 2. SAMPLES OF BANGLA GENUINE AND FORGED SIGNATURES

Genuine Signatures	Forged signatures
অতুন্ত্রমান কে	অতুল্লমান কে
বন্দিনী টৈল	বন্দিনী টিন

V. FEATURE EXTRACTION

Feature extraction is a crucial step in any pattern recognition system. Two different types of feature extraction techniques such as: gradient feature extraction and the chain code feature are considered here.

A. Computation of 576-dimensional gradient Features

576-dimensional gradient features were extracted for this research and experimentation, which are described in paper [7].

B. 64-Dimensional Chain Code Feature Extraction

The 64-dimensional Chain Code feature is determined as follows. In order to compute the contour points of a two-tone image, a 3 x 3 window is considered surrounding the object point. If any one of the four neighbouring points (as shown in Fig. 3 (a)) is a background point, then this object point (P) is considered as a contour point. Otherwise it is a non-contour point.

The bounding box (minimum rectangle containing the character) of an input character is then divided into 7 x 7 blocks. In each of these blocks, the direction chain code for each contour point is noted and the frequency of the direction codes is computed. Here, the chain code of four directions only [directions 1 (horizontal), 2 (45 degree slanted), 3 (vertical) and 4 (135 degree slanted)] is used. Four chain code directions are shown in Fig. 3 (b). It is assumed that the chain code of directions 1 and 5, 2 and 6, 3 and 7, 4 and 8, are the same. Thus, in each block, an array is obtained of four integer values representing the frequencies, and those frequency values are used as features. Thus, for 7 x 7 blocks, 7 x 7 x 4= 196 features are obtained. To reduce the feature dimensions, after the histogram calculation into 7 x 7 blocks, the blocks are down-sampled with a Gaussian filter into 4 x 4 blocks. As a result, 4 x 4 x 4 = 64 features are obtained for recognition. To normalize the features, a maximum value of the histograms from all the blocks, is computed. Each of the above features is divided by this maximum value to obtain the feature values between 0 and 1.
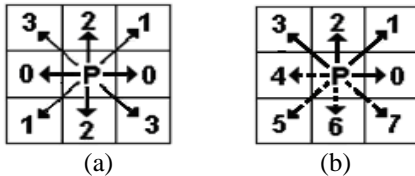

Figure 3. Eight neighbours (a) For a point P and its neighbours (b) For a point P the direction codes for its eight neighbouring points.

VI. CLASSIFIER DETAILS

Based on these features, Support Vector Machines (SVMs) and the Modified Quadratic Discriminant Function (MQDF) are applied as the classifiers for the experiments.

A. SVM Classifier

For this experiment, a Support Vector Machine (SVM) classifier is used. The SVM is originally defined for two-class problems and it looks for the optimal hyper plane, which maximizes the distance and the margin, between the nearest examples of both classes, named support vectors (SVs). Given a training database of M data: $\{x_m | m=1,..., M\}$, the linear SVM classifier is then defined as:

$$f(x) = \sum_j \alpha_j x_j \cdot x + b$$

where $\{x_j\}$ are the set of support vectors and the parameters α_j and b have been determined by solving a quadratic problem [11]. The linear SVM can be extended to various non-linear variants; details can be found in [11, 12]. In these proposed experiments, the Gaussian kernel SVM outperformed other non-linear SVM kernels, hence identification/verification results based on the Gaussian kernel are reported only.

B. MQDF Classifier

The Modified Quadratic Discriminant Function is defined as follows [13].

$$D(X) = (N + N_0 + n - 1)In\left[1 + \frac{1}{N_0\sigma^2}\left[\|X - M\|^2 - \sum_{i=1}^k \frac{\lambda_i}{\lambda_i\frac{N_0}{N}\sigma^2}\right]\right] + \sum_{i=1}^k In\left(\lambda_i + \frac{N_0}{N}\sigma^2\right)$$

where X is the feature vector of an input character; M is a mean vector of samples; Φ_i^T is the ith eigen vector of the sample covariance matrix; λ_i is the ith eigen value of the sample covariance matrix; k is the number of eigen values considered here; n is the feature size; σ^2 is the initial estimation of a variance; N is the number of learning samples; and N_0 is a confidence constant for s and N_0 is considered as 3N/7 for experimentation. All the eigen values and their respective eigen vectors are not used for classification. Here, the eigen values are stored in descending order and the first 60 (k=60) eigen values and their respective eigen vectors are used for classification. Compromising on trade-off between accuracy and computation time, k was determined as 60.

VII. EXPERIMENTAL SETTINGS

A. Settings for Verification used in 1st Stage of Experiments

The skilled forgeries were not considered for training purposes. For experimentation, random signatures were considered for training purposes. For each signature set, an SVM was trained with 12 randomly chosen genuine signatures. The negative samples for training (random signatures) were the genuine signatures of 149 other signature sets. Two signatures were taken from each set. In total, there were 149x2=298 random signatures employed for training. For testing, the remaining 12 genuine signatures and 30 skilled forgeries of the signature set being considered were employed. The number of samples for training and testing for these experiments are shown in Table 3.

Table 3. No. of Signatures used per set in 1st Phase of Verification

	Genuine Signature	Random Signatures	Skilled Forgeries
Training	12	298	n/a
Testing	12	n/a	30

B. Settings for Verification used in 2nd Stage of Experiments

1) Settings for Signature Script Identification

150 sets of signatures (50 sets of English, 50 sets of Hindi and 50 sets of Bangla) were used for signature script identification. 30 sets of signatures from each script were considered for training, and the remaining 20 sets were considered for testing purposes. The number of samples for training and testing used in experimentation of the identification approach are shown in Table 4.

TABLE 4. SIGNATURE SAMPLES USED FOR SCRIPT IDENTIFICATION PHASE.

	English Signatures		Hindi Signatures		Bangla Signatures	
	Genuine	Forged	Genuine	Forged	Genuine	Forged
Training	720	900	720	900	720	900
Testing	480	600	480	600	480	600

2) Settings for Signature Verification after Signature Script Identification

The verification task in the second stage was explored separately for English signatures, Hindi signatures and Bangla signatures based on the identified script result. Signature samples (30 sets from each script) that were considered for training purposes in signature script identification were not used for the individual verification task. Only the correctly identified samples from 20 sets (used for the testing part in identification) were considered for verification. For each signature set, an SVM was trained with 12 genuine signatures. The negative samples for training were 95 (19x5) genuine signatures of 19 other signature sets.

VIII. RESULTS AND DISCUSSION

A. Experimental Results

1) First Verification Experiments

In this stage of experimentation, 8100 (150x54) signatures involving English, Hindi and Bangla scripts were employed for training and testing purposes. At this operational point, the SVMs produced an AER of 20.80%, and an encouraging accuracy of 79.20% was achieved in this first mode of verification.

2) Second Verification Experiments

In this stage of verification the signatures are identified based on their script and subsequently, the identified signatures are applied separately for verification. In the signature script identification stage, only 64-dimensional chain code features were used because a slightly better accuracy was obtained when compared to the gradient feature. The MQDF classifier was also taken into account in the script identification step applying chain code features for a better accuracy, but MQDF did not achieve the better result as compared to SVMs in this study. To get a comparative idea, script identification results using two different classifiers with chain code features are shown in Table 5. An accuracy of 93.08% is achieved at the script identification stage by using the SVM classifier. The accuracy of Bangla, English and Hindi are 85.19, 95.74 and 98.33% respectively. Confusion matrices obtained using SVM classifiers, and the 64-dimensional chain code features investigated, are shown in Table 6.

TABLE 5. ACCURACY OBTAINED USING SVM AND MQDF CLASSIFIERS

Classifiers	Identification Accuracy (%)
SVMs	93.08
MQDF	82.45

TABLE. 6. CONFUSION MATRIX OBTAINED USING THE CHAIN CODE FEATURE AND SVM CLASSIFIER

	Bangla	English	Hindi
Bangla	920	19	141
English	27	1034	19
Hindi	10	8	1062

Based on the outcomes of the identification phase, verification experiments subsequently followed. Verification results obtained for individual scripts were calculated on 93.08% (identification rate) accuracy levels. In this phase of experimentation, the SVMs produced an overall AER of 21.10%, 13.05% and 15.05% using English, Hindi and Bangla signatures respectively. The overall verification accuracy obtained for the second major experiments (identification plus verification) was 83.60% (average of 78.90% of English, 86.95% of Hindi and 84.94% of Bangla).

B. Comparision of Performance

From the experimental results obtained, it was observed that the performance of signature verification in the second set of experiments (identification and verification) was encouraging compared to the signature verification accuracy from the first experiment set (verification only). Table 7 demonstrates the accuracies attained in the first experiment set as well as separate verification results for English, Hindi and Bangla from the second experiment set.

TABLE 7. VERIFICATION ACCURACIES RESULTING FROM DIFFERENT EXPERIMENTS

Verification Techniques		Accuracy (%)
Experiment Sets	**Dataset Used**	
1st experiment	English, Hindi and Bangla	79.20
2nd experiment	English	78.90
	Hindi	86.95
	Bangla	84.94

In the second stage of verification, the overall accuracy is 83.60% (Avg. of 78.90%, 86.95% and 84.94%) which is 4.40 (83.60-79.20) higher than the accuracy in the first

stage. The comparison of these two accuracies is shown in Table. 8.

TABLE 8. ACCURACY IN DIFFERENT PHASES OF VERIFICATION

Verification Experiment	Verification Accuracy (%)
Without Script Identification	79.20
With Script Identification	83.60

From the above table it is evident that verification accuracy with script identification is much higher than without script identification. This increased accuracy is achieved because of the proper application of the identification stage. This research clearly demonstrates the importance of using identification in multi-script signature verification techniques.

C. Error Analysis

Most of the methods used for signature verification generate some erroneous results. In these experiments, a few signature samples were mis-identified in both the identification and verification stages. Few of the confusing signature samples obtained in the signature script identification stage using the SVM classifier are shown in Figures 4, 5 and 6. Three categories of confusing samples are generated by the classifier. The first category illustrates a Bangla signature sample treated as a Hindi signature sample. The second one illustrates an English signature sample treated as a Bangla signature sample and the third one illustrates a Hindi signature sample treated as a Bangla signature sample.
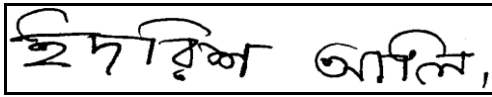

Figure 4. Bangla sample treated as Hindi

Figure 5. English treated as Bangla

Figure 6. Hindi Signature treated as Bangla

IX. CONCLUSIONS AND FUTURE WORK

This paper provides an investigation of the excellent performance of a multi-script signature verification technique involving English, Hindi and Bangla off-line signatures. The novel approach used in a multi-script signature verification environment with the combination of a custom Hindi and Bangla off-line signature dataset provides a substantial contribution to the field of signature verification. In such a verification environment, the proper utilization of a script identification technique, which

substantially affects the verification accuracy, indicates an important step in the process. The comparatively higher verification accuracy obtained in the second experimental approach is likewise a substantial contribution. The gradient feature, chain code feature as well as SVM and MQDF classifiers were employed for experimentation. The idea of a multi-script signature verification approach, which deals with an identification phase, is a very important contribution to the area of signature verification. The proposed off-line multi-script signature verification scheme is a new investigation in the field of off-line signature verification. In the near future, we plan to extend our work considering further sets of signature samples, which may include different languages/scripts.

X. ACKNOWLEDGMENTS

REFERENCES

[1] R. Plamondon and G. Lorette, "Automatic signature verification and writer identification - the state of the art", Pattern Recognition, pp.107–131, 1989.

[2] S. Madabusi, V. Srinivas, S. Bhaskaran and M. Balasubramanian, "On-line and off-line signature verification using relative slope algorithm", in proc. International Workshop on Measurement Systems for Homeland Security, pp. 11-15, 2005.

[3] D. Impedovo, G. Pirlo, " Automatic signature verification: The state of the art", IEEE transactions on Systems, Man, and Cybernetics part-C, vol. 38, no. 5, pp. 609–635, 2008.

[4] M. Kalera, S. Srihari and A. Xu. "Offline signature verification and identification using distance statistics", International Journal on Pattern Recognition and Artificial Intelligence, pp.1339-1360, 2004.

[5] S. Pal, U. Pal, M. Blumenstein, "Hindi Off-line Signature Verification", in proc. International Conference on Frontiers in Handwritten Recognition, ICFHR 2012, Bari, Italy, pp. 371-376.

[6] S. Pal, A. Alaei, U. Pal, M. Blumenstein, "Multi-Script off-line signature identification" , in proc. International Conference on Hybrid Intelligent Systems, pp. 236-240, 2012.

[7] S. Pal, U. Pal, M. Blumenstein, "Hindi and English off-line signature identification and verification", in proc. International Conference on Advances in Computing. pp. 905–910, 2012.

[8] B. B. Chaudhuri and U. Pal, "An OCR system to read two Indian language scripts: Bangla and Devnagari (Hindi)", in proc. International Conference on Document Analysis and Recognition, pp. 1011–1015, 1997.

[9] M. A. Ferrer, J. B. Alonso, and C. M. Travieso, "Offline geometric parameters for automatic signature verification using fixed-point arithmetic", IEEE transactions on Pattern Analysis and Machine Intelligence, 27:993–997, 2005.

[10] S. Pal, U. Pal and M. Blumenstein, "A Two-Stage Approach for English and Hindi Off-line Signature Verification", International workshop on Emerging Aspects in Handwritten signature processing, 2013(Acceoted).

[11] V.Vapnik, "The Nature of Statistical Learning Theory", Springer Verlang, 1995.

[12] C. Burges, "A Tutorial on support Vector machines for pattern recognition", Data Mining and Knowledge Discovery, pp.1-43, 1998.

[13] F. Kimura, K. Takashina, S. Tsuruoka and Y. Miyake, "Modified quadratic discriminant function and the application to Chinese character recognition", IEEE transactions on Pattern Analysis and Machine Intelligence, Vol. 9, pp 149-153, 1987.