A Two-Stage Approach for English and Hindi Off-line Signature Verification

Srikanta Pal[1], Umapada Pal[2], and Michael Blumenstein[1]

[1] School of Information and Communication Technology, Griffith University, Australia
[2] Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, India
srikanta.pal@griffithuni.edu.au

Abstract. The purpose of this paper is to present an empirical contribution towards the understanding of multi-script off-line signature identification and verification using a novel method involving off-line Hindi (Devnagari) and English signatures. The main aim of this approach is to demonstrate the significant advantage of the use of signature script identification in a multi-script signature verification environment. In the 1st stage of the proposed signature verification technique a script identification technique is employed to know whether a signature is written in Hindi or English. In the second stage, a verification approach was explored separately for English signatures and Hindi signatures based on the script identification result. Different features like gradient feature, water reservoir feature, loop feature, aspect ratio etc. were employed, and Support Vector Machines (SVMs) were considered in our scheme. To get the comparative idea, multi-script signature verification results on the joint Hindi and English dataset without using any script identification technique is also computed. From the experiment results it is noted that we are able to reduce average error rate 4.81% more when script identification method is employed.

1 Introduction

Nowadays, civilized and advanced societies require secure means for personal authentication. Traditional authentication methods, which are based on knowledge (a password) and the utility of a token (photo ID cards, magnetic strip cards, keys) are less reliable because of forgetfulness, loss and theft. These issues direct substantial attention to biometrics as an alternative method for person authentication and identification. Consequently, handwritten signatures as a pure behavioral biometric are used widely and are well accepted as a personal authentication method.

On the basis of signature acquisition, signature verification methods can be categorized into two groups: static (offline) and dynamic (online) methods [1, 7]. Offline signature verification uses the shape of the signature to authenticate the signer [2]. Online signature verification uses dynamic characteristics of the signature (time-dependent signals) to authenticate the signer [3].

Although significant research has already been undertaken in the field of signature verification, particularly involving single-script signatures, conversely, less attention has

A. Petrosino, L. Maddalena, P. Pala (Eds.): ICIAP 2013 Workshops, LNCS 8158, pp. 140–148, 2013.

been devoted to the task of multi-script signature verification. Pal et al. [4] presented a bi-script signature identification technique involving Bangla (Bengali) and English off-line signatures based on background and foreground information. The purpose of that paper was to identify whether a claimed signature belongs to the group of Bengali or English signatures. In another contribution by Pal et al. [5], a multi-script off-line signature identification technique was proposed. In that signature identification scheme, the signatures involving Bangla (Bengali), Hindi (Devnagari) and English were considered for the identification process. An encouraging accuracy of 92.14% was obtained in those experiments. In another report by Pal et al. [6], multi-script off-line signature identification and verification involving English and Hindi signatures was presented. The verification phase considering multi-script signatures before the identification stage was not investigated, and it was the drawback of that technique. As a consequence, the comparison of experimental outcomes of two different verification stages (before identification and after identification) was not previously considered in that paper.

In this paper a two-stage approach is proposed for multi-script signature verification. In the 1st stage of the proposed signature verification technique a script identification technique is employed to know whether a signature is written in Hindi or English. In the second stage, a verification approach was explored separately for English signatures and Hindi signatures based on the script identification result. To get the idea about the advantage of this two-stage approach, multi-script signature verification results on the joint Hindi and English dataset without using any script identification technique is also computed and the diagram of such a system is shown in Fig.1.

2 Significance of Multi-script Signature Verification

In 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 a passport application form, an examination question paper, a money order form, bank account application form 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 expressly needed. Consequently, the idea of the proposed multi-script signature verification technique considering Hindi and English signatures are significant.

Development of a general multi-script signature verification system, which can verify signatures of all scripts, is very complicated and it is not possible to develop such a system in the Indian scenario. The verification accuracy in such multi-script signature environments will not be desirable compared to single script signature verification. To achieve the necessary accuracy for multi-script signature verification, it is first important to identify signatures based on the type of script and then use an individual

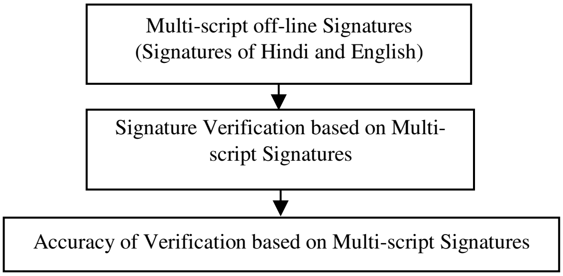


Fig. 1. Diagram of Signature Verification Considering Joint Dataset

single script signature verification system for the identified signature script. Based on this observation, in the proposed system the signatures of different scripts are separated to feed into the individual signature verification system. The diagram of such a system is shown in Fig.2. To the best of the authors' knowledge, a complete analysis of a multi-script signature verification technique, where signature identification factors affecting signature verification results with a large dataset, is still missing from the literature. This research work is the first important report towards such a direction in the area of signature verification.

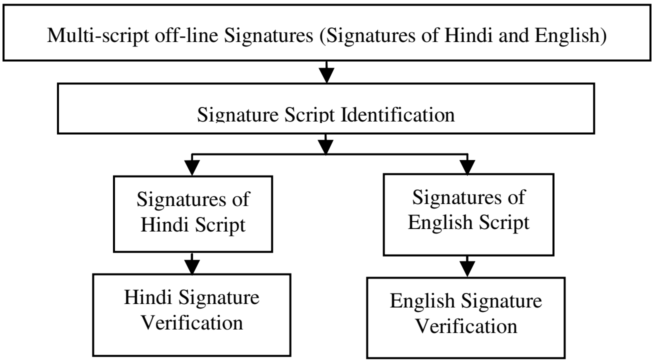


Fig. 2. Diagram of Multi-script Signature Identification and Verification Based on Hindi and English Signatures

3 Database Used for Experimentation

3.1 Hindi Signature Database

As there has been no public signature corpus available for Hindi script, it was necessary to create a database of Hindi signatures. This Hindi signature database consists of 100 sets. Each set consists of 24 genuine signatures and 30 skilled forgeries. From each individual, 24 genuine signatures were collected. A total number of 2400 genuine signatures from 100 individuals were collected. A total number of 3000 skilled forgeries were collected from the writers.

3.2 GPDS English Database

Another database consisting of 100 sets from GPDS-160 [8] was also utilised for these experiments. The reason 100 sets were used from the GPDS on this occasion, is due to the fact that the Hindi dataset described previously was comprised of 100 sets, and it was considered important to have equivalent signature numbers for experimentation and comparison between the two datasets.

4 Feature Extraction and Classifier Details

Before feature extraction, the signatures are extracted from the data collection form, and for this purpose some pre-processing is performed as follows. The signatures to be processed by the system needed to be in a digital image format. Each signature was handwritten on a rectangular space of fixed size of a white sheet of paper. It was necessary to scan all signature document pages. At the very beginning, the images were captured in 256 level grey-scale at 300 dpi and stored in TIFF format (Tagged Image File Format) for the purpose of future processing. In the pre-processing step, a histogram-based threshold technique was applied for binarization. Then the signature images were extracted from the signature-collecting document forms.

Feature extraction is a crucial step in any pattern recognition system. Three different types of feature extraction techniques such as: 576 dimensional gradient feature extraction (described in [9]), water reservoir-based technique (described in [10]), the aspect ratio-based technique [11] and the loop feature [12] are considered here.

Based on gradient feature, Support Vector Machines (SVMs) are applied as the classifier for verification experiments. Other features are used in a Tree classifier for signature script identification.

5 Experimental Settings

5.1 Settings for Verification Used Prior to Script Identification

In this experimental method of verification, skilled forgeries were not considered for training, and genuine signatures were considered for both training and testing purposes. For the experiments in the proposed research, the Hindi database developed and 100 sets from the GPDS described in Section 3, were used. For each signature set, an SVM was trained with 12 randomly chosen genuine signatures. The number of genuine samples (24) in a set was divided in two parts for training and testing purposes. The negative samples for training (random forgeries) were the genuine signatures of 199 other signature sets. Two signatures were taken from each set. In total, there were 199x2=398 random forgeries 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 1.

Table 1. No. of Signatures Used Per Set in 1st Phase of Verification

	Genuine Signature	Random Forgeries	Skilled Forgeries
Training	12	398	n/a
Testing	12	n/a	30

Table 2. No. of Signatures used per Set in the 2nd Phase of Verification

	Genuine Signature	Random Forgeries	Skilled Forgeries
Training	12	198	n/a
Testing	12	n/a	30

5.2 Settings for Verification Used after Script Identification

The same experimental settings were also used for this verification method, except the number of negative (random) samples used for the training phase. In total 99x2=198 negative samples for training were employed from 99 other genuine signature sets. The number of samples for training and testing for these experiments are shown in Table 2.

5.3 Script Identification Strategies

Feature set selection for signature script identification was decided based on the knowledge of observing the signature samples used during experimentation. As the physical characteristics and structural behaviours of the Hindi and English signatures were well known, the feature set selection for the script identification method was straightforward. Viewing the samples in the combined dataset it was noted that all the Hindi samples were written as full signatures and the English samples were written primarily as ‘initial’ signatures. As a consequence, the height of Hindi signature samples was smaller than the width of samples. Conversely, the height and width of most of the English signatures were nearly same. So, it was easy to apply the aspect ratio feature to identify those two different scripts of signatures. But there were a few samples that were not as easy to identify by the aspect ratio (H/W) feature. So, in the next stage, the top water reservoir (WR) (for details about this feature see [10]), number of loops (L) and reservoir height (RH) was further applied to identify those signatures. Applying these three simple feature sets, all the samples were correctly identified (100% accuracy). Naturally, these three heuristic features may not necessarily yield the same accuracy levels for identification using other datasets. Two signatures of Hindi and English with their existing top water reservoirs are shown in Fig.3 and Fig.4 respectively. The graphical representation of the script identification strategy is shown in Fig.5.

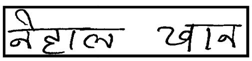
a)  b) 

Fig. 3. (a) A Hindi Sample and (b) its Top Reservoirs

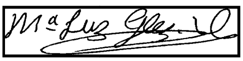
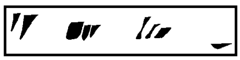
a)  b) 

Fig. 4. (a) An English Sample and (b) its Top Reservoirs

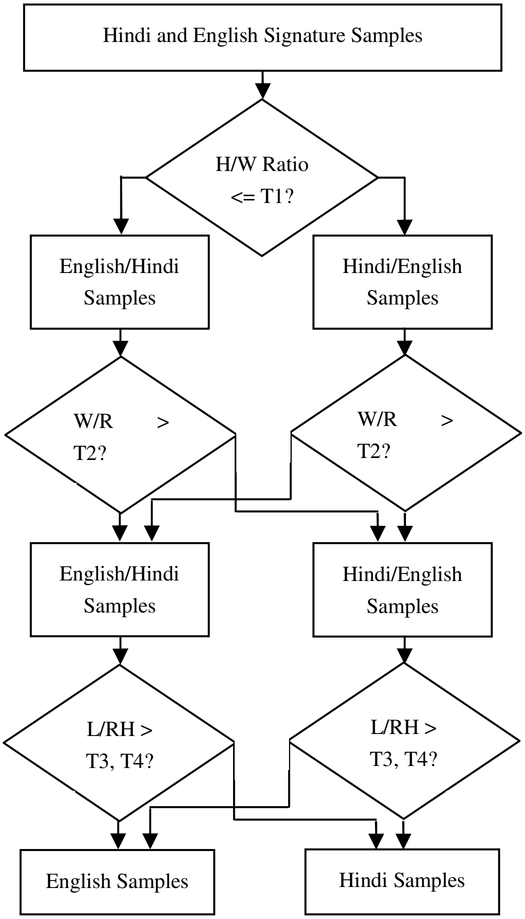


Fig. 5. Flow Chart of the proposed Identification Method (Here T1=4, T2=3, T3=7 and T4=30 are different threshold values obtained empirically)

6 Results and Discussion

6.1 Experimental Results

The experimental results of any signature verification system are evaluated based on FRR (False Rejection Rate), FAR (False Acceptance Rate) and the AER (Average Error Rate).

6.1.1 First Method of Verification(Without Script Identification)

In this experiment, 10800 signatures involving English and Hindi scripts were employed for training and testing purposes. Among these signatures, 4800 (24x200) samples were genuine and 6000 (30x200) samples were skilled forgeries. Using the gradient features, the FRR, FAR and AER are calculated. At this operational point, the SVM classifiers produced an FRR of 20.12 % and an FAR of 14.36 %. An encouraging accuracy of 82.76 % is achieved for this verification experiment. The graphical representation of different accuracies with different values of gamma settings for SVMs is shown in Table 3.

Table 3. Accuracies with Different Values of Gamma Using Joint Dataset

Gamma	FRR(%)	FAR(%)	AER(%)
17000	20.12	14.36	17.24
18000	21.44	16.24	18.84
19000	21.90	16.34	19.12

6.1.2 Second Method of Verification

For this experiment, the signatures are first identified based on their script and subsequently the identified signatures are sent separately for verification. All signature samples are identified based on the features discussed in Section 4. An accuracy of 100% is achieved at the identification stage.

Based on the outcomes of the identification phase, verification experiments are conducted as follows. In this phase of experimentation, the SVMs produced an FRR of 18.12 %, FAR of 12.18 % and AER of 15.15% using English signatures. On the other hand, an FRR of 12.17%, FAR of 7.25% and AER of 9.71% were achieved employing Hindi signatures. Three different accuracies for Hindi and English signatures are calculated based on three different values of gamma for the SVM classifier. The values of these three different experimental outcomes achieved for different values of SVM settings for gamma are shown in Tables 4 (a) and (b).

Table 4. Accuracies using Different Values of Gamma for (a) English and (b) Hindi Datasets

(a)

Gamma	FRR(%)	FAR(%)	AER(%)
19000	18.12	12.18	15.15
18000	18.66	12.20	15.43
17000	18.90	16.20	17.55

(b)

Gamma	FRR(%)	FAR(%)	AER(%)
19000	12.17	7.25	09.71
18000	12.65	7.85	10.25
17000	17.25	8.25	12.75

6.2 Comparision of Performance

From the experimental results obtained, it was observed that the performance of signature verification in the second method is very encouraging compared to signature verification using the first method. Table 5 demonstrates the accuracies achieved in the first and second methods discussed in sub-Sections 6.1.1 and 6.1.2.

For the second method of verification, the overall accuracy is 87.57 % (Avg. of 84.85 and 90.29) which is higher than the accuracy obtained from the first method.

Table 5. Accuracies for Different Methods

Verification Methods		Accuracy (%)
Stages	**Dataset Used**	
1st Method	**English & Hindi**	82.76
2nd Method	**English**	84.85
	Hindi	90.29

Table 6. Accuracies for Different Schemes

Verification Method	Accuracy (%)
Without Script Identification	82.76
With Script Identification	87.57

The comparison of these two accuracies is shown in Table 6. It can be noted from the table that the average error can be reduced 4.81% (87.57-82.76) when the script identification method is employed.

From the above table and chart it is easily understood that verification accuracy after script identification is much higher than before. This increased accuracy is achieved because of the proper application of the identification stage. Thus, this research clearly demonstrates the importance of using identification in multi-script signature verification techniques.

6.3 Error Analysis

Most of the methods used for signature verification generate some erroneous results. In the proposed approach, confusing signature samples obtained using the SVM classifier is shown in Fig.6 and Fig.7. Two categories of confusing samples are given by the classifier. The first category illustrates genuine signature samples treated as forged signature samples. The second one illustrates forged signature samples treated as genuine signature samples.

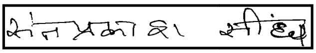


Fig. 6. Genuine Signatures Treated as Forgeries

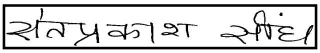


Fig. 7. Forged Signatures Treated as Genuine

7 Conclusions and Future Work

This paper demonstrates an investigation of the excellent performance of a multi-script signature verification technique involving Hindi and English off-line signatures. Actually, the novel approach used in a multi-script signature verification environment with the combination of a large newly-prepared Hindi off-line signature dataset provides a substantial contribution in the field of signature verification. In such a verification environment, the proper utilization of a script identification phase, which substantially affects the verification accuracy, indicates an important step in the process. The comparatively higher verification accuracy obtained for the second method of experimentation (identification plus verification) is likewise a substantial contribution.

The gradient feature, water reservoir feature, loop feature and aspect ratio feature as well as SVM classifiers were employed for experimentation. To the best of the authors' knowledge, the idea of multi-script signature verification, which deals with an identification phase, has not previously been used for the task of signature verification, and hence this is the first important report detailing such a process in 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 groups of signature samples, which may include different languages/scripts.

References

1. Madabusi, S., Srinivas, V., Bhaskaran, S., Balasubramanian, M.: On-line and off-line signature verification using relative slope algorithm. In: International Workshop on Measurement Systems for Homeland Security, pp. 11–15 (2005)
2. Kalera, M., Srihari, S., Xu, A.: Offline signature verification and identification using distance statistics. International Journal on Pattern Recognition and Artificial Intelligence, 1339–1360 (2004)
3. Plamondon, R., Lorette, G.: Automatic signature verification and writer identification - the state of the art. Pattern Recognition, 107–131 (1989)
4. Pal, S., Alaei, A., Pal, U., Blumenstein, M.: Off-line signature identification using background and foreground information. In: International Conference on Digital Image Computing, pp. 672–677 (2011)
5. Pal, S., Alaei, A., Pal, U., Blumenstein, M.: Multi-Script off-line signature identification. In: 12th International Conference on Hybrid Intelligent Systems, pp. 236–240 (2012)
6. Pal, S., Pal, U., Blumenstein, M.: Hindi and English off-line signature identification and verification. In: International Conference on Advances in Computing, pp. 905–910 (2012)
7. Impedovo, D., Pirlo, G.: Automatic Signature Verification: The State of the Art. IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews 38, 609–635 (2008)
8. Ferrer, M.A., Alonso, J.B., Travieso, C.M.: Offline geometric parameters for automatic signature verification using fixed-point arithmetic. IEEE PAMI 27, 993–997 (2005)
9. Ito, T., Ohyama, W., Wakabayashi, T., Kimura, F.: Combination of signature verification techniques by SVM. In: International Conference on Frontiers in Handwriting Recognition, pp. 428–431 (2012)
10. Pal, U., Belaid, A., Choisy, C.: Touching numeral segmentation using water reservoir concept. Pattern Recognition Letters 24(1-3), 261–272 (2003)
11. Schafer, B., Viriri, S.: An off-Line signature verification system. In: International Conference on Signal and Image Processing Applications, pp. 95–100 (2009)
12. Tarafdar, A., Mandal, R., Pal, S., Pal, U., Kimura, F.: Shape code based word-image matching for retrieval of Indian multi-lingual documents. In: ICPR, pp. 1989–1992 (2010)