Hindi and English Off-line Signature Identification and Verification

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

[1] School of Information and Communication Technology,
Griffith University, Gold Coast, Australia
srikanta.pal@griffithuni.edu.au
m.blumenstein@griffith.edu.au
[2] Computer Vision and Pattern Recognition Unit,
Indian Statistical Institute, Kolkata, India
umapada@isical.ac.in

Abstract. Biometric systems play a significant role in the field of information security as they are extremely required for user authentication. Signature identification and verification have a great importance for authentication intention. The purpose of this paper is to present an empirical contribution towards the understanding of multi-script (Hindi and English) signature verification. This system will identify whether a claimed signature belongs to the group of English signatures or Hindi signatures from a combined Hindi and English signature datasets and then it will verify signatures using these two resultant signature datasets (Hindi script signature and English script signatures) separately. The modified gradient feature and SVM classifier were employed for identification and verification purposes. To the best of authors' knowledge, the multi-script signature identification and verification has never been used for the task of signature verification and this is the first report of using Hindi and English signatures in this area. Two different results for identification and verification are calculated and analysed. The accuracy of 98.05% is obtained for the identification of signature script using 2160 (1080 Hindi + 1080 English) samples for training and 1080 (540 Hindi + 540 English) samples for testing. The resultant data sets obtained in script identification of signatures were used for verification purpose. The FRR, FAR for Hindi and English was obtained 8.0%, 4.0% and 12.0%, 10.0% respectively.

Keywords: Signature verification, biometrics, SVMs, Gradient Feature.

1 Introduction

Signature verification has been a topic of intensive research during the past several years [1-8] due to an important role it plays in numerous areas including the financial system. The verification of human signatures is particularly concerned with the improvement of the interface between human beings and computers [2]. A signature verification system and the associated techniques used to solve the inherent problems of authentication can be divided into two classes [3]: (a) on-line methods [4] to measure temporal and sequential data by utilizing intelligent algorithms [5] and (b)

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 905–910.
springerlink.com

off-line methods [6] that use an optical scanner to obtain handwriting data written on paper. Off-line signature verification deals with the verification of signatures, which appear in a static format [7]. On-line signature verification has been shown to achieve much higher verification rates than off-line verification [6], since a considerable amount of dynamic information is lost in the off-line mode. However off-line systems have a significant advantage as they do not require access to special processing devices when the signatures are produced. Moreover, the off-line group has many more practical application areas than that of its on-line counterpart.

2 Database Preparation and Pre-processing

A database of 1620 Hindi signatures and 1620 English signatures are used for identification purpose. English signatures from GPDS were used in our experimentation. Each Hindi and English signature set consists of 24 genuine signatures and 30 skilled forgeries. A total number of 720 genuine Hindi signatures from 30 individuals were collected. For each contributor, all genuine specimens were collected in a single day's writing session. In order to produce the forgeries, the imitators were allowed to practice their forgeries as long as they wished with static images of genuine specimens. A total number of 900 Hindi skilled forgeries were collected from the writers.

3 Modified Gradient Feature

The gray-scale local-orientation histogram of the component is used for 576 dimensional feature extractions. To obtain 576-dimensional gradient-based feature vector, the following steps are executed.

Step 1: A 2 x 2 mean filtering is applied 5 times on the input image.

Step 2: The gray-scale image obtained in Step 1 is normalized so that the mean gray scale becomes zero with maximum value 1.

Step 3: The normalized image is then segmented into 17x7 blocks. Compromising trade-off between accuracy and complexity, this block size is decided experimentally. To get the bounding box of the grey-scale image, the image is converted into two-tone using Otsu's thresholding algorithm [9]. This will exclude unnecessary background information from the image.

Step 4: A Roberts filter is then applied on the image to obtain the gradient image. The arc tangent of the gradient (direction of gradient) is quantized into 32 directions and the strength of the gradient is accumulated with each of the quantized direction. The strength of the Gradient $f(x, y)$ is defined as follows:

$f(x, y) = \sqrt{(\Delta u)^2 + (\Delta v)^2}$ and the direction of gradient $(\theta(x, y))$ is: $\theta(x, y) = \tan^{-1} \frac{\Delta v}{\Delta u}$ Where $\Delta u = g(x+1, y+1) - g(x, y)$ and $\Delta v = g(x+1, y) - g(x, y+1)$ and $g(x, y)$ is the gray level of (x, y) point.

Step 5: Histograms of the values of 32 quantized directions are computed for each of the 17 x 7 blocks.

Step 6: The directional histogram of the 17 x 7 blocks is down sampled into 9 x 4 blocks and 16 directions using Gaussian filters. Finally, a 9 x 4 x 16 = 576 dimensional feature vector is obtained.

4 Classifier and Experimental Settings

In our experiments, we have used Support Vector Machines (SVM) as classifiers. SVMs have been originally defined for two-class problems and they look for the optimal hyper plane, which maximizes the distance and the margin between the nearest examples of both classes, namely 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 [7]. The linear SVM can be extended to various non-linear variants; details can be found in [7, 8]. In our experiments, the RBF kernel SVM outperformed other non-linear SVM kernels, hence we are reporting our recognition results based on the RBF kernel only. The experimental settings we used are described below.

4.1 Settings for Script Identification

For the experiments in the proposed research, our developed Hindi signature database described in section 4 was used. A numbers of 60 set of signatures (30 Hindi dataset and 30 English dataset) were used for identification of signature script. A signature samples of 1080(20x54) Hindi and 1080(20x54) English were used for training phase whereas 540 (10x54) Hindi and 540(10x54) English signature samples were used for testing purpose for identification of signature script. The number of samples for training and testing for experimentation of identification are shown in Table 1.

4.2 Settings for Signature Verification

The accuracy of 98.05% is obtained for the identification of signature script using 1080 (540 Hindi + 540 English) samples for testing. SVMs classifier misidentified 21signatures, i.e 1.95% (100.00-98.05) of 1080 samples. The number of errors occurred in testing dataset for identification is shown in Table 2. The signature verification was done using 1059(1080-21 samples) correctly identified script of signatures. For verification, the database was split in two parts, to perform the training and testing components. The signature samples of 466 genuine signatures (226 Hindi+ 240 English) and 595 skilled forgeries (299 Hindi + 296 English) were used for verification purpose from 10 set of Hindi signatures and 10 set of English signatures, respectively. For each signature set, an SVM was trained with 14 randomly chosen genuine signatures. The negative samples for training were the 20 skilled forgeries of signatures. For testing, the remaining genuine signatures and remaining skilled forgeries were used. The Hindi and English signature samples used for verification with each signature set are shown in Table 3 and Table 4.

Table 1. Number of Signature Samples Used for Identification of Signature Script

	Hindi Signature		English Signature	
	Genuine	**Forged**	**Genuine**	**Forged**
Training	480	600	480	600
Testing	240	300	240	300

Table 2. Number of Signature Script Identification Errors Occurred in Different Datasets

No. of Errors in Test Datasets Obtained in Identification Part				
	Hindi Test Samples		**English Test Samples**	
Datasets	**Genuine Signatures**	**Forged signatures**	**Genuine Signatures**	**Forged signatures**
Set-1	0	0	0	0
Set-2	2	0	0	0
Set-3	1	0	0	0
Set-4	0	0	0	0
Set-5	0	0	0	0
Set-6	4	1	0	4
Set-7	7	0	0	1
Set-8	0	0	0	1
Set-9	0	0	0	0
Set-10	0	0	0	0
Total Errors	14	1	0	6

5 Results and Discussion

As mentioned earlier, the accuracy of 98.05% is obtained for the identification of signature script. Using the Gradient feature, an FAR (False Acceptance Rate) and FRR (False Rejection Rate) was computed. At this operational point, the FRR, FAR for Hindi were 8.0%, 4.0% and FRR, FAR for English were 12.0 %, 10.0% respectively. The FRR, FAR and AER (Average Error Rate) obtained from our experiments are shown in Table 5. The AER obtained in this research is 6.0% for Hindi and 11.0 % for English.

Confusion matrix of signature identification obtained from SVM classifiers and gradient features are shown in Table 6. It is noted that only 6 English signatures were misidentified as Hindi signatures and 15 Hindi signatures were misidentified as English. Two samples of signature script identification errors (English and Hindi signature treated as Hindi and English respectively) are shown in Figure 1 and figure 2.

Some verification errors (Hindi and English genuine signature treated as Hindi and English forged signature and Hindi and English forged signature treated as Hindi and English genuine signature) are shown in Figure 3, Figure 4 and Figure 5, Figure 6, respectively.

Table 3. Hindi Samples Used for Verification

Hindi datasets used for verification		
Hindi Datasets	**Genuine Signatures**	**Forged signatures**
Set-1	24	30
Set-2	22	30
Set-3	23	30
Set-4	24	30
Set-5	24	30
Set-6	20	29
Set-7	17	30
Set-8	24	30
Set-9	24	30
Set-10	24	30
Total samples	**226**	**299**

Table 4. English Samples Used for Verification

English datasets used for verification		
English Datasets	**Genuine Signatures**	**Forged signatures**
Set-1	24	30
Set-2	24	30
Set-3	24	30
Set-4	24	30
Set-5	24	30
Set-6	24	26
Set-7	24	29
Set-8	24	29
Set-9	24	30
Set-10	24	30
Total samples	**240**	**294**

Table 5. Results of FRR, FAR and EER

	FRR	**FAR**	**AER**
Hindi	8.0 %	4.0 %	6.0 %
English	12.0%	10.0%	11.0 %

Table 6. Confusion Metric for Identification of Signature Script

	Hindi	**English**
Hindi	525	15
English	6	534

Fig. 1. English Signature Sample Treated as Hindi Signature

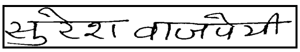

Fig. 2. Hindi Signature Sample Treated as English Signature

Fig. 3. Hindi Genuine Signature Sample Treated as Hindi Forged Signature

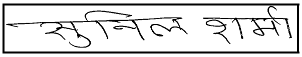

Fig. 4. Hindi Forged Signature Treated as Hindi Genuine Signature

Fig. 5. English Genuine Signature Sample Treated as English Forged Signature

Fig. 6. English Forged Signature Sample Treated as English Genuine Signature

6 Conclusions and Future Work

This paper presents a signature identification and verification scheme of bi-script off-line signatures. To the best of our knowledge, bi-script signatures have never been used for the task of signature verification and this is the first report in this area. This scheme of bi-script off-line signature identification is a novel contribution to the field of signature verification. In near future, we plan to extend our work for multi-script off-line signature identification and verification.

References

[1] Chen, S., Srihari, S.: Use of Exterior Contour and Shape Features in Off-line Signature Verification. In: 8th ICDAR, pp. 1280–1284 (2005)

[2] Ferrer, M.A., Alonso, J.B., Travieso, C.M.: Off-line Geometric Parameters for Automatic SignatureVerification Using Fixed-Point Arithmetic. IEEE PAMI 27(6), 993–997 (2005)

[3] 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)

[4] Emerich, S., Lupu, E., Rusu, C.: On-line Signature Recognition Approach Based on Wavelets and Support Vector Machines. In: Intl Conf. on Automation Quality and Testing Robotics, pp. 1–4 (2010)

[5] Kholmatov, A., Yanikoglu, B.: Identity Authentication using improved online signature verification method. PRL 26, 2400–2408 (2005)

[6] Kalera, M., Srihari, S., Xu, A.: Offline signature verification and identification using distance statistics. In: IJPRAI, pp. 1339–1360 (2004)

[7] Vapnik, V.: The Nature of Statistical Learning Theory. Springer (1995)

[8] Burges, C.: A Tutorial on support Vector machines for pattern recognition. In: Data Mining and Knowledge Discovery, pp. 1–43 (1998)

[9] Otsu, N.: A threshold selection method from gray-level histogram. IEEE Trans. on SMC 9, 62–66 (1979)