Hindi Off-line Signature Verification

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

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

Umapada Pal
Computer Vision and Pattern
Recognition Unit, Indian Statistical
Institute, Kolkata-700108, India.
Email:
umapada@isical.ac.in

Abstract—**Handwritten Signatures are one of the widely used biometrics for document authentication as well as human authorization. The purpose of this paper is to present an off-line signature verification system involving Hindi signatures. Signature verification is a process by which the questioned signature is examined in detail in order to determine whether it belongs to the claimed person or not. Despite of substantial research in the field of signature verification involving Western signatures, very little attention has been dedicated to non-Western signatures such as Chinese, Japanese, Arabic, Persian etc. In this paper, the performance of an off-line signature verification system involving Hindi signatures, whose style is distinct from Western scripts, has been investigated. The gradient and Zernike moment features were employed and Support Vector Machines (SVMs) were considered for verification. To the best of the authors' knowledge, Hindi signatures have never been used for the task of signature verification and this is the first report of using Hindi signatures in this area. The Hindi signature database employed for experimentation consisted of 840 (35x24) genuine signatures and 1050 (35x30) forgeries. An encouraging accuracy of 7.42% FRR and 4.28% FAR were obtained following experimentation when the gradient features were employed.**

Keywords- Signature verification, Indian script, Hindi signatures, Document security.

I. INTRODUCTION

Handwritten signatures are one of the most widely accepted personal attributes for identity verification. Signature verification has been a topic of renewed intensive research over the past several years [1, 2] due to the important role it plays in numerous areas, including in financial applications.

Automatic signature verification systems can be classified into two categories: on-line and off-line [3]. In an on-line technique, signatures are signed on a digitizer and dynamic information such as speed and pressure is captured in addition to a static image of the signature [4, 5]. In an off-line technique, signatures are signed on a piece of paper and then scanned to digitally store the signature image [6]. Hence, off-line signature verification deals with the verification of signatures, which appear in a static format [7]. Verification decisions are usually based on local or global features extracted from the signature being processed. Excellent verification results can be achieved by comparing the robust features of the test signature with that of the user's signature using an appropriate classifier [20].

Signatures are considered as a complete image with a special distribution of pixels, and a particular writing style. They are not considered as a collection of letters and words [8]. A person's signature may change radically during their lifetime. Great inconsistency can even be observed in signatures according to country, habits, psychological or mental state, physical and practical conditions [9].

There has been substantial work in the area involving off-line verification of Western signatures. Armand et al. [10] presented an effective method to perform off-line signature verification and identification. Unique structural features were extracted from the signature's contour. Using a publicly available database of 2106 signatures containing 936 genuine and 1170 forgeries, the verification rate of 91.12% was obtained. Ramachandra et al. [11] proposed an off-line signature verification system based on a Cross-Validation principle and graph matching. Schafer and Viriri [12] presented an off-line signature verification system based on the combination of feature sets. Some extracted features were: Aspect ratio, centroid feature, four surface features, six surface features, number of edge points, transition features etc. The verification of signatures was accomplished using the Euclidean distance classifier.

Signatures may be written in different languages and there is a need to undertake a systematic study in this area. Many published works are available for Western signatures and only a few studies have been undertaken for signatures written in Chinese, Japanese, Persian, Arabic etc. [19]. To the best of the authors' knowledge there is no published work on Hindi signature verification and this paper deals with Hindi signature verification. The present work of Hindi signature verification would be considered as a novel

contribution to the field of signature verification. Some signature samples of Hindi script are shown in Figure 1.

The remainder of this paper is organized as follows. The different types of forgeries are described in Section II. The Hindi signature database developed for the current research is described in Section III. Some notable properties of Devnagari script are introduced in Section IV. Section V briefly describes the feature extraction techniques employed in the work. Details of the classifiers used are presented in Section VI. The experimental settings are presented in Section VII, and results and discussion are given in Section VIII. Error analyses are described in Section IX. Finally, conclusions and future work are discussed in Section X.
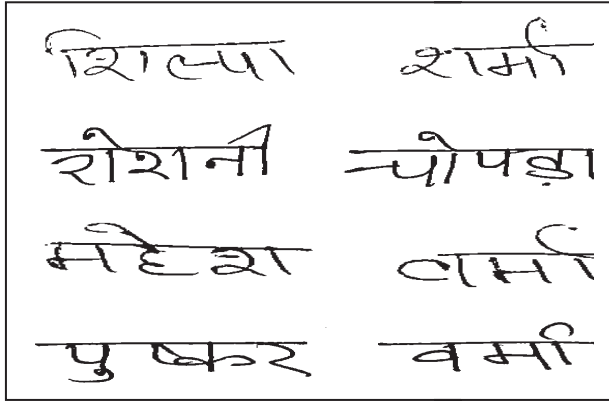

Figure 1. Hindi signature samples

II.   TYPES OF FORGERIES

In general, off-line/on-line signature verification can be considered as a two-class classification problem. Here the first class represents the genuine signature set, and the second class represents the forged signature set. Usually two types of errors are considered in a signature verification system: The False Rejection or Type-I error and the False Acceptance or Type-II error. These error types are associated with two common types of error rates: the False Rejection Rate (FRR) which is the percentage of genuine signatures misclassified as forgeries, and False Acceptance Rate (FAR) which is the percentage of forged signatures misclassified as genuine. According to Coetzer et al. [13], three basic types of forged signatures, which are often taken into account, are:

1. Random forgery. The forger has no access to the genuine signature (not even the author's name) and reproduces a random one. In many cases, the forgeries are the forger's own genuine signature.

2. Simple forgery. The forger knows the author's name and the script, but has no access to a sample of the signature.

3. Skilled forgery. The forger has access to one or more samples of the genuine signature and is able to reproduce it.

III.   HINDI SIGNATURE DATABASE

Although automatic signature verification has been an active research area for several decades, there has been no publicly available signature database for Hindi, the most popular official Indian script. Therefore, a Hindi signature database was created for the purpose of this work. So, the research in automatic signature verification has long been constrained by the unavailability of a standard database.

TABLE 1. GENUINE AND FORGED SIGNATURES

Hindi Signatures	
Genuine Signatures	*Forged Signatures*

A.   Data collection and database preparation

The signatures of Hindi script were considered for this signature verification approach. As there has been no public signature corpus available for Hindi script, it was necessary to create a database of Hindi signatures. The signatures were collected from West Bengal, India. The majority of the signatures were contributed by students. This Hindi signature database consists of 35 sets whereby the writer number ranges from H-S-Set-001 to H-S-Set-035 (Hindi - Signature-Set). In order to collect the genuine signatures corresponding to each individual, a collection form was designed. The form contained 24 boxes where the signatures could be written. From each individual, 24 genuine signatures were collected. A total number of 840 genuine signatures from 35 individuals were collected. For each contributor, all genuine specimens were collected in a single day's writing session. In addition, only skilled forged signatures were collected for this proposed work. In order to produce the forgeries, the imitators were allowed to practice their forgeries for as long as they wished with static images of genuine specimens. A total number of 1050 forged

signatures were collected from the writers. Some genuine signature samples with their corresponding forgeries are displayed in Table 1.

B. Pre-processing

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 on 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. In this step, the digitized grey-level image is converted into a two-tone image. Then the signature images were extracted from the signature-collection document forms. The signature collection form containing 24 genuine signatures in grey level are shown in Figure 4. The extracted binary signature images were stored in TIFF format. A typical scanned signature and its corresponding binary image are shown in Figure 2 and Figure 3, respectively.
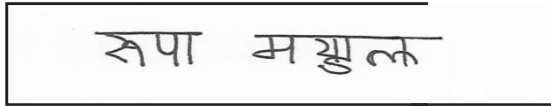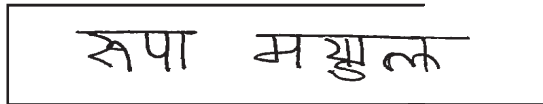

Figure 2. Scanned signature image

Figure 3. Binary signature image

Figure 4. Signature-collection form with genuine signatures
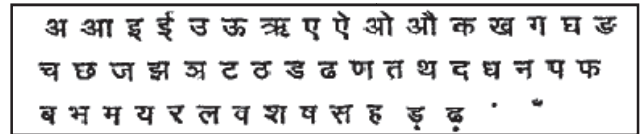

Figure 5. Basic characters of Devnagari script

A text line in such scripts can be partitioned into three zones (upper, middle and lower). The upper zone denotes the portion above the headline, the middle zone denotes the portion between the headline and baseline and the lower zone is the portion below the baseline. The imaginary line separating the middle and lower zones is called the baseline.

Figure 6. Vowel modifiers of Devnagari

IV. PROPERTIES OF DEVNAGARI SCRIPTS

Devnagari is an oriental script descended from Brahmi script [14]. It is the most popular official script and national language of India. In Hindi script, the writing direction is from left to right and there is no concept of upper/lower case. Hindi script has about fifty basic characters. These characters are presented in Figure 5.

Vowels in this script generally take a modified shape in most words and are called modifiers or allographs. Modifiers generally do not disturb the shape of basic characters in the middle zone of a line. If the shape is disturbed in the middle zone, we call the resultant shape a compound character. Vowel modifiers of Devnagari scripts are shown in Figure 6.
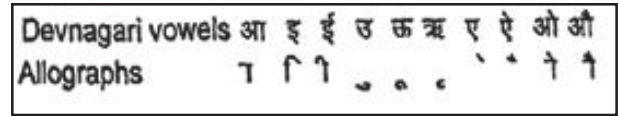
V. FEATURE EXTRACTION

Feature extraction is a crucial step in any pattern recognition system. The Zernike feature and the gradient feature extraction technique are described below.

A. Zernike moments feature

Zernike polynomials are an orthogonal set of complex-valued polynomials:

$$V_{nm}(x,y) = R_{nm}(x,y).\exp\left(\left(jm\,\tan^{-1}(\frac{y}{x})\right)\right)$$

where

$$x^2 + y^2 \leq 1, \quad j = \sqrt{-1}, \qquad n \geq 0,$$
$$|m| \leq n$$

and $\quad n - |m| \quad$ is even and Radial polynomials $\{R_{nm}\}$
are defined as:

$$R_{nm}(x, y) = \sum_{s=0}^{\frac{n-|m|}{2}} B_{n|m|s}(x^2 + y^2)^{\frac{n}{2}-s}$$

where

$$B_{n|m|s} = \frac{(-1)^s (S-1)!}{S! \left(\dfrac{n+|m|}{2} - S\right)! \left(\dfrac{n-|m|}{2} - S\right)!}$$

The complex Zernike moments of order n and repetition m are given by:

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x, y) V_{nm}^*(x, y)$$

Where $x^2 + y^2 \leq 1$ and the symbol $*$ denotes the complex conjugate operator [15].
The Zernike moments can be computed by the scale invariant Central moments as follows:

$$A_{nm} = \frac{n+1}{\pi} \sum_{\substack{k-|m| \\ n-k=even}}^{n} \sum_{a=0}^{b} \sum_{d=0}^{|m|} (-j)^d$$

$$\cdot \binom{|m|}{d} \binom{b}{a} B_{n|m|s} G_{k-2a-d, 2a+d}$$

where

$$b = \frac{n-|m|}{2} - S \quad and \quad j = \sqrt{-1} .$$

B.   Computation of 576-dimensional gradient features

The grey-scale local-orientation histogram of the handwriting component is used for 576-dimensional feature extraction.
To obtain a 576-dimensional gradient-based feature vector, the following steps were executed.
Step 1: 2 x 2 mean filtering is applied 5 times on the input image.
Step 2: The grey-scale image obtained in Step 1 is normalized so that the mean grey scale becomes zero with a maximum value of 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, it is converted into a two-tone image using Otsu's thresholding algorithm [16]. 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 directions. The strength of the Gradient ($f(X, Y)$) is defined as follows:

$$f(X, Y) = \sqrt{(\Delta u)^2 + (\Delta v)^2} \qquad (1)$$

and the direction of gradient $\quad (\theta(x, y)) \quad$ is:

$$\theta(x, y) = \tan^{-1} \frac{\Delta v}{\Delta u} \qquad (2)$$

where $\quad \Delta u = g(x+1, y+1) - g(x, y) \qquad (3)$

and $\quad \Delta v = g(x+1, y) - g(x, y+1) \qquad (4)$

and $\quad g(x, y) \quad$ is the grey level of point (x, y).

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.

VI.   CLASSIFIER DETAILS

In our experiments, we have used Support Vector Machines (SVMs) 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 [17]. The linear SVM can be extended to various non-linear variants; details can be found in [17, 18]. In our experiments, the RBF kernel SVM outperformed other non-linear SVM kernels, hence we are reporting our verification results based on the RBF kernel only. Different parameters of the kernel are chosen experimentally.

VII. EXPERIMENTAL SETTINGS

In the field of signature verification, there is lack of publicly available signature databases. The quality of available databases also varies, as there has been no standard collection protocol. Besides, it is very costly to create a large corpus with different types of forgeries, especially skilled forgeries.

For the experiments in the proposed research, our own database described in Section 4 was used. The database was split into two parts to perform the training and testing phases. 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 (10 signatures) and the remaining skilled forgeries (10 signatures) of the signatures being considered were employed. The number of samples for training and testing for the purpose of experimentation with each signature set, are shown in Table 2

TABLE 2. NUMBER OF SIGNATURE SAMPLES USED PER SET FOR EXPERIMENTATION

	Genuine Signature	Skilled Forgeries
Training	14	20
Testing	10	10

VIII. RESULTS AND DISCUSSION

For training and testing of the system, 1890 Hindi signatures were employed. Among these signatures, 1190 (14x35+20x35) samples were considered in the training phase and 700 (10x35+10x35) samples were considered in the testing phase. Using the gradient feature and Zernike moment feature, an FRR (False Rejection Rate) and FAR (False Acceptance Rate) were obtained. At this operational point, the Support Vector Machines produced an FRR of 7.42% and an FAR of 4.28% using the gradient feature whereas an FRR of 8.57% and an FAR of 5.14% were obtained using the Zernike moment Feature. We obtained better results compared to our previous Bangla signature verification work [21].

The FRR and the FAR rates changed rapidly with the value of gamma used in the SVM classifiers. The error rates obtained from these experiments are shown in Table 3 to Table 6. These encouraging results demonstrate the robustness of the gradient feature and Zernike moment feature extraction techniques for signature verification.

From the experimental results obtained, it is also observed that the performance of the gradient feature is very encouraging, compared to the results obtained using the Zernike moment feature. The results obtained from this experiment using our Hindi dataset are comparatively better than the state-of-the-art methodologies proposed [9].

TABLE 3. VALUE OF FRR USING THE GRADIENT FEATURE

Gamma	FRR (%)
0.000005	8.57
0.000003	8.00
0.000001	7.42

TABLE 4. VALUE OF FAR USING THE GRADIENT FEATURE

Gamma	FAR (%)
0.000005	5.14
0.000003	4.87
0.000001	4.28

TABLE 5. VALUE OF FAR USING ZERNIKE MOMENTS

Gamma	FRR (%)
0.000003	28.57
0.000002	17.14
0.000001	8.57

TABLE 6. VALUE OF FRR USING ZERNIKE MOMENTS

Gamma	FAR (%)
0.000003	13.71
0.000002	8.28
0.000001	5.14

IX. ERROR ANALYSIS

Confusing samples obtained using the SVM classifier based on the gradient feature and the Zernike moment feature are shown in Figure 7, Figure 8 and Figure 9, Figure 10, respectively. Two categories of confusing samples were obtained by the classifier. The first category illustrates the genuine signature samples treated as forged signature samples. The second one illustrates the forged signature samples treated as genuine signature samples.

Figure 7. Genuine signature sample treated as a forged signature (based on the gradient feature).

Figure 8. Forged signature sample treated as a genuine signature (based on the gradient feature).
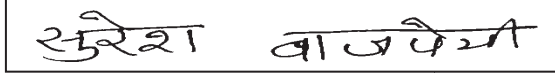

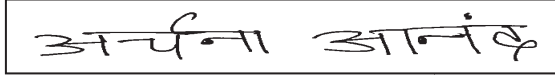Figure 9. Genuine signature treated as a forged signature (based on the Zernike moment feature).

Figure 10. Forged signature treated as a genuine signature (based on the Zernike moment feature).

X. CONCLUSIONS AND FUTURE WORK

This paper presents an investigation of the performance of a signature verification system involving Hindi off-line signatures. The gradient feature, Zernike moment feature and SVM classifiers were employed, and encouraging results were obtained. To the best of our knowledge, a Hindi signature database has not previously been used for the task of signature verification and this is the first report of using Hindi signatures in this area. The proposed off-line verification scheme is the first investigation for Hindi signatures in the field of off-line signature verification. In the near future, we plan to extend our work considering more samples of Hindi signatures with the exploration of different feature extraction techniques and classifiers. We also plan to prepare a Hindi signature database and make it publicly available.

REFERENCES

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

[2] M.A. Ferrer, J.B. Alonso and C. M. Travieso, "Off-line Geometric Parameters for Automatic SignatureVerification Using Fixed-Point Arithmetic" IEEE PAMI, Vol.27, no.6, pp. 993-997, 2005.

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

[4] R. Plamondon and G. Lorette, "Automatic signature verification and writer identification - the state of the art", Pattern Recognition, 22, pp.107–131, 1989.

[5] A. Kholmatov, and B. Yanikoglu, "Identity Authentication using improved online signature verification method", PRL, pp. 2400-2408, 2005.

[6] M. Kalera, S. Srihari, and A. Xu. "Offline signature verification and identification using distance statistics", IJPRAI, pp.1339-1360, 2004.

[7] I. Pottier and G. Burel, ''Identification and Authentication of Handwritten Signatures with a Connectionist Approach", In Proc. IEEE Conf. On Neural Networks, pp. 2948-2951, 1994.

[8] B. Fang, C.H. Leung, Y.Y. Tang, K.W. Tse, P.C.K. Kwok and Y.K. Wong, "Off-line signature verification by the tracking of feature and stroke positions", Pattern Recognition, pp. 91-101, 2003.

[9] K. Franke, "Analysis of Authentic Signatures and Forgeries" In Proc. IWCF, pp 150-164, 2009.

[10] S.Armand, M. Blumenstein and V. Muthukkumarasamy, "Off-line Signature Verification based on the Modified Direction Feature," Proceedings of 18th ICPR, pp. 509-512, 2006.

[11] A. C. Ramachandra, K. Pavithra, K. Yashasvini, K. B. Raja, K. R. Venugopal and L. M. Patnaik, "Off-line Signature Verification based on Cross-Validation for Graph Matching," IEEE International Conference on Electrical and Electronics (INDICON-2008), pp. 17-22, 2008.

[12] B.Schafer and S.Viriri, "An Off-Line Signature Verification System", ICSIPA, pp.95-100, 2009.

[13] J. Coetzer, B. Herbst and J. D. Preez, "Off-line signature verification using the discrete radon transform and a hidden markov model", EURASIP Journal on Applied Signal Processing, vol 4, pp.559–571, 2004.

[14] B. B. Chaudhuri and U. Pal, "An OCR system to read two Indian language scripts: Bangla and Devnagari (Hindi)", Proceedings of 4th ICDAR, pp. 1011–1015, 1997.

[15] C. H. Teh and R. T. Chin, "On Image Analysis By The Methods Of Moments," IEEE Trans. Patt. Anal. Mach. Intell., Vol. 10, No. 4, pp. 496-513, 1988.

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

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

[18] C. Burges, "A Tutorial on support Vector machines for pattern recognition", Data mining and knowledge discovery, pp.1-43, 1998.

[19] H. Lv, W. Wang, C. Wang and Q. Zhuo, "Off-line chinese signature verification based on support vector machines", Pattern Recognition Letters, pp. 2390–2399, 2005.

[20] R. Plamondon and S.N. Srihari, "On-line and off-line Handwriting Recognition: A comprehensive Survey", IEE tran. On Pattern Analysis and Machine Intelligence. Vol. 22 no.1, pp. 63-84, 2000.

[21] S. Pal, V. Nguyen, M. Blumenstein and U. Pal "Off-line Bangla Signature Verification", 10th DAS, Gold Coast, Queensland, Australia, 27-29 March, pp. 282-286, 2012.