

# Multi-Script Off-line Signature Identification

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

Alaei Alireza  
Computer Science  
Laboratory (LI, RFAI  
group), Francois  
Rabelais University,  
Tours, France  
alireza20alaei@yahoo.com

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

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

**Abstract** - In this paper, we present an empirical contribution towards the understanding of multi-script signature identification. In the proposed signature identification system, the signatures of Bengali (Bangla), Hindi (Devanagari) and English are considered for the identification process. This system will identify whether a claimed signature belongs to the group of Bengali, Hindi or English signatures. Zernike Moment and histogram of gradient are employed as two different feature extraction techniques. In the proposed system, Support Vector Machines (SVMs) are considered as classifiers for signature identification. A database of 2100 Bangla signatures, 2100 Hindi signatures and 2100 English signatures are used for experimentation. Two different results based on two different feature sets are calculated and analysed. The highest accuracy of 92.14% is obtained based on the gradient features using 4200 (1400 Bangla +1400 Hindi + 1400 English) samples for training and 2100 (700 Bangla +700 Hindi +700 English) samples for testing.

**Keywords** - Off-line system; Signature Identification; Biometrics; Verification systems; SVMs.

## I. INTRODUCTION

Signatures are socially and legally well established as a convenient means of writer verification, document authentication and authorisation. Signature verification has been a topic of renewed intensive research during the past several years [1-9] due to its importance role in numerous areas including the financial system.

Numerous techniques for feature extraction and classification have also been put forward in the literature for the purpose of script identification. Hochberg et al. [10] proposed an algorithm for script and language identification from handwritten document images using statistical features based on connected component analysis. Hangarge and Dhandra [11] investigated texture as a tool for determining the script of a handwritten document image, based on the observation that text has a distinct visual texture. Their proposed algorithm achieved average classification accuracy as high as 99.2% for bi-script and 88.6% for tri-script separation at the text line and text block level respectively with a fivefold cross validation test. Roy and Pal [12] presented an automatic scheme for word-wise identification of hand-written Roman and Oriya scripts for Indian postal automation. In their proposed scheme, a piecewise projection method was used for line and word segmentation. Finally, using different features such as water reservoir concept-based features, fractal dimension-based features,

topological features, script characteristics-based features etc.; a Neural Network (NN) classifier was used for word-wise script identification. For experimentation, 2500 words were considered and an overall accuracy of 97.69% was obtained from the proposed identification scheme. To the best of our knowledge there is only one paper for signature identification in the literature [19]. In [19], a bi-script (English and Bangla) signature identification scheme was considered where back ground and fore ground information of the signature images were used. Since signature identification may help the final goal of verification in a multi-script environment, we elaborate its important and justify its need in the following sections.

The remainder of this paper is organised as follows. Section II describes the significance of multi-script signature identification. Properties of Bangla and Devanagari scripts are presented in Section III. Creation of a multi-script signature dataset is described in Section IV. Section V deals with the feature extraction techniques employed. Details of the classifiers used are presented in Section VI. Results and discussion are given in Section VII. Finally, conclusions and future work are discussed in Section VIII.

## II. SIGNIFICANCE OF MULTI-SCRIPT SIGNATURE IDENTIFICATION

Although, many systems involving off-line signature recognition and verification have been developed, all of these systems have solely considered single-script signatures. However, signatures may be written in different languages and there is a need to undertake a systematic study in this area. In the field of signature verification, most of the works reported in the literature have been undertaken primarily for English signatures. Only a few studies have been conducted for Chinese, Japanese, Persian and Arabic signatures.

Moreover, in the field of signature verification, there are only two published works on signatures written in Indian languages [18, 19]. Some countries (e.g. India and Singapore) have two or more scripts that are not only used for handwriting but also for signing purposes. Thus there is a need to work on signature identification considering signatures of two or more scripts. Since most of the states in India use three different languages (Indian state language, Indian National language and international language), signatures of these three categories of languages (Bangla, Hindi and English) which are mainly used in the state of West Bengal are considered for investigation in this research work. Some Bangla, Hindi and English signature samples

used in this research work are shown in Tables 1, 2 and 3, respectively.

Development of a generic multi-script signature verification system, which can verify signatures of all scripts, is very complicated and it is very difficult to develop such system in the Indian scenario. The verification accuracy in such multi-script signature verification system will be much less compared to single script signature verification system. For higher accuracy in multi-script signature verification system and make this system more simplified, it is better to first identify different scripts from a multi-script signature database and then use an individual, single script signature verification system for the identified signature. Based on this observation, in the proposed system the signatures of different scripts are separated to feed into the individual signature verification system. The block diagram of such a system is shown in Figure 1.

TABLE 1. BANGLA SIGNATURE SAMPLES

শ্রীশ্রী কুমার মোহন উদ্ভাসন	ঐক্যমতি কুমার
হুমায়ুন কান	জামিলা বেগম

TABLE 2. DEVANAGARI SIGNATURE SAMPLES

স্নেহা ঐশিক	বিবেক মহেশ্বরী
অর্চনা আনন্দ	পারমিতা বসু

TABLE 3. ENGLISH SIGNATURE SAMPLES

Arghadeep Datta Chaudhury	PJ Darcy
Jess Liew	T. Higgs

### III. PROPERTIES OF BANGLA AND DEVNAGARI SCRIPTS

Most of the Indian scripts including Bangla and Devanagari have originated from ancient Brahmi script through various transformations and evolution [13]. 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. Both scripts have about fifty basic characters including vowels and consonants. Basic characters of Bangla and Devanagari scripts are shown in Figure 2 and Figure 3.

Apart from the vowels and consonants, there are compound (modifier) characters in most Indian scripts including Bangla and Devanagari, which are formed by combining two or more basic characters. 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 Bangla and Devanagari scripts are shown in Figure 4.

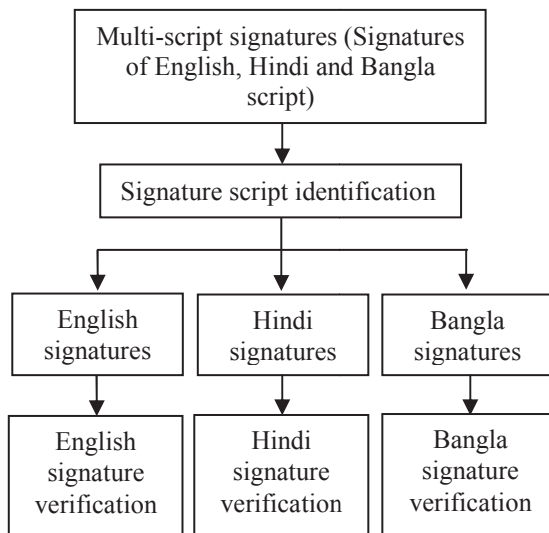


Figure 1. Flow-diagram of multi-script signature identification.

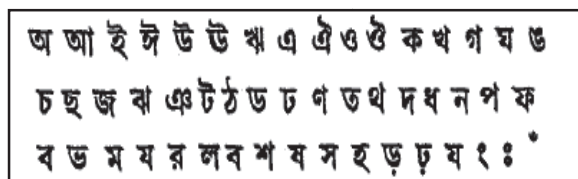


Figure 2. Bangla basic characters

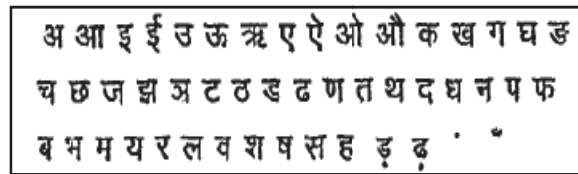


Figure 3. Devnagari basic characters

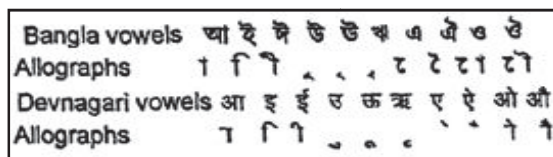


Figure 4. Vowel modifiers of Bangla & Devnagari

### IV. CREATION OF A MULTI-SCRIPT SIGNATURE DATASET

In the field of signature verification, there is a lack of publicly available signature databases. Besides, it is very costly to create a large corpus with different scripts and various types of forgeries, especially skilled forgeries.

As no public signature dataset was available for Bangla and Hindi scripts and for the purpose of experiments in this research work, it was necessary to create a custom database, which included Bangla, Hindi as well as English signatures.

The Bangla and Hindi signatures were collected from 55 individuals from different parts of West Bengal, India. In order to collect the signatures, a collection form was designed. The form contained 24 boxes where the signatures could be written. Few of the writers signed 24 signatures and few of them signed 48 signatures. The signatures in the data set were not all from unique individuals. Another, 2100 English signatures were collected from 65 individuals from Australia and India.

The collection forms were then scanned using a flatbed scanner in 256 grey-scales at 300 dpi and stored in TIFF format (Tagged Image File Format). A grey-scale scanned signature collection form with signatures written in English is shown in Figure 5. The signatures were then extracted from the filled in signature collection forms by a simple segmentation technique. A histogram-based threshold technique was applied for the binarisation of the signatures extracted from the filled in signature form. Each binary signature image was also stored in TIFF format. A typical segmented signature and its binarised equivalent are shown in Figure 6 and Figure 7, respectively.



Figure 5. Signature collection form with signatures

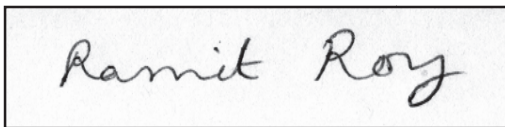


Figure 6. Scanned image

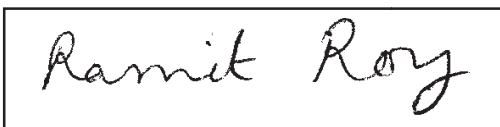


Figure 7. Binary image

## V. FEATURE EXTRACTION TECHNIQUES

The choice of a powerful set of features is a crucial step in signature identification and verification systems. The gradient feature and Zernike Moment feature are two different powerful features used in many pattern recognition systems [16, 17]. These two feature extraction techniques are used in this research work and their description are given in the following subsections.

### A. 400-Dimensional Gradient Feature Extraction

The grey-scale local-orientation histogram of a component is used for 400-dimensional feature extraction. These features help us to compute segmentation confidence. To obtain 400-dimensional features the following steps are employed.

**Step 1:** At first, size normalisation of the input binary image is performed. Here we normalise the image into  $126 \times 126$  pixels. This size is decided based on the experiment.

**Step 2:** The input binary image is then converted into a grey-scale image applying  $2 \times 2$  mean filtering 5 times.

**Step 3:** The grey-scale image is normalised so that the mean grey scale becomes zero with a maximum value of 1.

**Step 4:** The normalised image is then segmented into  $9 \times 9$  blocks.

**Step 5:** A Roberts filter is then applied on the image to obtain the gradient image. The arc tangent of the gradient (strength of the gradient) is quantized into 16 directions (an interval of  $22.5^\circ$ ) and the strength of the gradient is accumulated with each of the quantized directions. By the strength of Gradient ( $f(x, y)$ ) we mean

$$f(x, y) = \sqrt{(\Delta u)^2 + (\Delta v)^2}$$

and by direction of gradient ( $\theta(x, y)$ ) we mean  $\theta(x, y) = \tan^{-1} \frac{\Delta v}{\Delta u}$ , here

$$\Delta u = g(x + 1, y + 1) - g(x, y),$$

$$\Delta v = g(x + 1, y) - g(x, y + 1)$$

and  $g(x, y)$  is a grey scale value at an  $(x, y)$  point.

**Step 6:** Histograms of the values of 16 quantized directions are computed in each of the  $9 \times 9$  blocks.

**Step 7:** The  $9 \times 9$  blocks are down-sampled to  $5 \times 5$  by a Gaussian filter. Thus, we get  $5 \times 5 \times 16 =$  a 400-dimensional feature. For details about this feature extraction see [17].

### B. Zernike Moment Feature

Zernike polynomials are an orthogonal set of complex-valued polynomials [15]:

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

where

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

and  $n - |m|$  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

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

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 symbol  $*$  denotes the complex conjugate operator [15].

The Zernike moments can be computed by the scale invariant Central moments as follows:

$$A_{nm} = \frac{n+1}{\pi} \sum_{\substack{k=|m| \\ n-k=\text{even}}}^n \sum_{a=0}^b \sum_{d=0}^{|m|} (-j)^d \cdot \binom{m}{d} \binom{b}{a} B_{n|m|s} G_{k-2a-d, 2a+d}$$

where

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

## VI. CLASSIFIER DETAILS

In our experiments, Support Vector Machines (SVMs) are used as classifiers. SVMs were originally defined for two-class problems and they look for the optimal hyper plane which maximizes the distance (the margin) between the nearest examples of both classes, named support vectors (SVs). Given a training database of  $M$  data:  $\{x_m\}_{m=1, \dots, M}$ , 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  $\alpha_j$  and  $b$  have been determined by solving a quadratic problem [14]. The linear SVM can be extended to various non-linear variants; details can be found in [14]. In our experiments, the Gaussian kernel SVM outperformed other non-linear SVM kernels; hence we are reporting our identification results based on the Gaussian kernel only. Different parameters of the kernel were chosen experimentally.

## VII. RESULTS AND DISCUSSION

The dataset of 2100 Bangla signatures, 2100 Hindi signatures and 2100 English signatures were considered for the experiments. A number of 4200 signature samples (1400 Bangla + 1400 Hindi + 1400 English) were considered for training and a number of 2100 (700 Bangla + 700 Hindi + 700 English) signature samples were considered for testing.

### A. Experimental Results

Table 4 shows the results for two different features obtained from the proposed signature identification technique in this research work. From Table 4 it can be noted that the best result of 92.14% was obtained using the 400-dimension gradient feature. The overall result of 91.23% was obtained when the Zernike Moment feature was used.

Performance comparison is important in scientific research. However, in the area of off-line signature identification/verification, an accurate comparison is difficult due to the unavailability of a standard signature corpus. No such comparison is made here for performance analysis with other previous works as there is not a single published work of multi-script signature identification.

TABLE 4. RESULTS BASED ON TWO DIFFERENT FEATURES USING SVM CLASSIFIER

Feature Sets	Identification rates (%)
Gradient Feature	92.14%
Zernike Moment Feature	91.23%

### B. Error analysis

Confusion matrices obtained using SVM classifiers, and the two different features investigated are shown in Tables 5 and 6. Table 5 shows the confusion matrix of the results obtained using 400-dimension gradient features and SVMs. It is noted that only 1 Bangla signature was misidentified as a Hindi signature and 35 Bangla signatures were misidentified as English. A sample of such errors (Bangla signature treated as English) is shown in Figure 8.

Other signature samples of such errors (Hindi signature treated as Bangla, and English signature treated as Bangla signature) are shown in Figures 9 and 10 respectively. Table 6 shows the confusion matrix obtained using Zernike Moment features and SVMs. In handwritten cases, the structural pattern of some characters of Bangla and Hindi script are nearly similar. Therefore it is little bit difficult to distinguish these two groups of signatures. Our experiments demonstrate that most of the errors occurred because of misclassification of Hindi signatures as Bangla signatures using SVM classifiers and the gradient feature (Table 5).

TABLE 5. CONFUSION MATRIX OBTAINED USING THE GRADIENT FEATURE AND SVM CLASSIFIER

	Bangla	Hindi	English
Bangla	664	1	35
Hindi	66	625	9
English	54	0	646

TABLE 6. CONFUSION MATRIX OBTAINED USING THE ZERNIKE MOMENT FEATURE AND SVM CLASSIFIER

	Bangla	Hindi	English
Bangla	625	56	19
Hindi	40	636	24
English	10	35	655

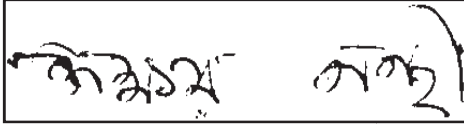


Figure 8. Bangla signature treated as English signature

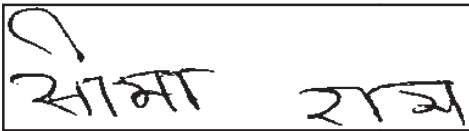


Figure 9. Devanagari signature treated as Bangla signature



Figure 10. English signature treated as Bangla signature

### VIII. CONCLUSIONS AND FUTURE WORK

This paper presents a signature identification scheme for multi-script off-line signatures. Signature script identification taking into consideration of three different scripts (Bangla, Hindi and English) is the first report in the field of signature identification. Different feature sets for off-line signature identification were utilised, and encouraging results (92.14% and 91.23%) were obtained. This scheme of multi-script off-line signature identification is also a novel contribution to progress the field of signature verification. We plan to extend our work considering more scripts for off-line multi-script signature identification. In the near future, we also plan to develop a multi-script off-line signature identification and verification method based on a combined signature dataset consisting of English and various other Indian scripts.

### REFERENCES

- [1] S. Chen, and S. Srihari, "Use of Exterior Contour and Shape Features in Off-line Signature Verification", 8th ICDAR, pp. 1280-1284, 2005.
- [2] M.A. Ferrer, J.B. Alonso and C. M. Travieso, "Off-line Geometric Parameters for Automatic Signature Verification Using Fixed-Point Arithmetic" IEEE PAMI, 27(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] S. Emerich, E. Lupu and C. Rusu. "On-line Signature Recognition Approach Based on Wavelets and Support Vector Machines", Intl Conf. on Automation Quality and Testing Robotics, pp.1-4, 2010.
- [5] A. Kholmatov, and B. Yanikoglu, "Identity Authentication using improved online signature verification method", Pattern Recognition Letters, 26, 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, 36, pp. 91-101, 2003.
- [9] K. Franke, "Analysis of Authentic Signatures and Forgeries" In Proc. IWCF, pp 150-164, 2009.
- [10] J. Hochberg, K. Bowers, Michael Cannon and Patrick Keely, "Script and language identification for handwritten document images," IJDAR, 2, pp. 45-52, 1999.
- [11] M. Hangarge and B. V. Dhandra, "Offline Handwritten Script Identification in Document Images", International Journal of Computer Applications, 4(6), pp. 6-10, 2010.
- [12] K. Roy and U. Pal, "Word-wise Hand-written Script Separation for Indian Postal automation", In Proc. 10<sup>th</sup> IWFHR, pp. 521-526, 2006.
- [13] B. B. Chaudhuri and U. Pal, "A complete printed Bangla OCR system", Pattern. Recognition, 31, pp. 531-549, 1997.
- [14] V.Vapnik, "The Nature of Statistical Learning Theory", Springer Verlag, 1995.
- [15] C. H. Teh and R. T. Chin, "On Image Analysis by the Methods of Moments," IEEE PAMI, 10(4), PP. 496-513, 1988.
- [16] M. Dehghan and K. Faez, "Farsi Handwritten Character Recognition With Moment Invariants", 13th International Conference on Digital Signal Processing Proceedings, (DSP) vol.2, pp.507 - 510, 1997.
- [17] U. Pal, T. Wakabayashi and F. Kimura, "Handwritten Bangla Compound Character Recognition using Gradient Feature", In Proc. 10th International Conf. on Info. Tech., pp. 208-213, 2007
- [18] S. Pal, V. Nguyen, M. Blumenstein, and U. Pal, "Off-line Bangla Signature Verification", In Proc. 10th IAPR International Workshop on Document Analysis Systems, (DAS), pp. 282-286, 2012.
- [19] S. Pal, A. Alaei, U. Pal, and M. Blumenstein, "Off-line Signature Verification based on Foreground and Background information". In Proc. of International Conference on Digital Image Computing: Techniques and Applications, DICTA, pp.672-677, 2011.