Off-line Signature Verification based on Hu's Moment Invariants and Zone Features using Support Vector Machine

Mandeep Kaur Randhawa
Lecturer, Department of Computer Science
Guru Nanak Khalsa College, Karnal,132001, Haryana, India

A.K. Sharma
Senior Scientist, Computer Centre, DES&M Division
National Dairy Research Institute, (Deemed University), Karnal –132001, Haryana, India

R.K Sharma
Professor, School of Mathematics and Computer Applications
Thapar University, Patiala–147004, Punjab, India

Abstract- An off-line signature verification system that incorporates a novel feature extraction technique is proposed in this paper. Moments of images have been used expansively in image analysis applications. In this paper, the fusion of Hu's moment invariants and zone features extracted from signature images is used as input patterns. Support Vector Machine (SVM) technique is used to verify the system. From the experimental results, the new features proved to be more robust than other related features used in the earlier systems. The proposed system has 4% errors in rejecting skilled forgeries (FRR), 2% errors in accepting genuine signatures (FAR) and 4% errors for identifying the particular person signature as some other signature (FIR). In all, the model exhibited 90% acceptance rate in terms of percent acceptance factor.

Keywords: Moments, Off-line Signature Verification, Statistical Features, Support Vector Machine.

I. Introduction

Handwritten or off-line signature verification is the process of confirming the identity of a user based on the handwritten signature of the user as a form of behavioral biometrics. Automatic handwritten signature verification is not a new problem. Many early research attempts were reviewed in the survey papers. The primary advantage that signature verification has over other types of biometric technologies is that handwritten signature is already the most widely accepted biometric for identify verification in daily use. The long history of trust over signature verification means that people are very willing to accept a signature-based biometric authentication system. However, there has not been any major international effort that aims at comparing different signature verification methods systematically. As common benchmark databases and benchmarking rules are often used by researchers in such areas as information retrieval and natural language processing, researchers in biometrics increasingly see the need for such benchmarks for comparative studies. Research is very actively under way in the signature verification domain.

In off-line systems, a signature is digitised using a flat bed scanner and then stored as an image. These images are called statistical or off-line signatures. Off-line data is a 2-D image of the signature. Off-line signature verification is considered as a behavioural characteristic based biometric trait in the field of security and the prevention of fraud [2]. Off-line systems are of interest especially in scenarios where only hard copies of a signature are available, *e.g.,* where a large number of documents need to be authenticated. Research in the field of off-line signature verification is relatively unexplored; this apathy can be attributed to the inherent limitation of available features from a statistical image of signatures. However, several research studies in on-line signature verification have been reported with high success rates. Verification decision is usually based on local features extracted from signature under processing. 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.

Much work pertaining to the development of off-line signature verification and recognition systems with variations in feature extraction and matching has been reported in literature, for instance, Chinese signature verification using statistical features specific to Chinese handwriting involving four main features for the optimisation of the verification of the Chinese signatures, *viz*., the envelop of the signature, cross-count feature, centre of gravity of sub-region and distance between vectors made of centre of gravity, and area of embedded white space [3]; an Artificial Neural Network (ANN) classifier has been used with statistical feature extraction technique [4]; moment-based representations, envelope characteristics, tree structured wavelet features, *etc.*, involving individual feature components being weighted by their pattern characterisation capability using genetic algorithm [5]; different types of pattern recognition schemes, *viz.*, statistical features[6]; off-line Persian signature identification and verification based on image registration DWT and image fusion [7] where DWT was used to access details of signature, then several registered instances of each person's signatures were fused together to generate reference pattern of person's signatures, while in the classification phase, Euclidean distance between the test image and each pattern was used in different sub-bands; a Dynamic Time Warping (DTW) method [8] that works by extracting the vertical projection feature from signature images and by comparing reference and probe feature templates with elastic matching; an off-line Arabic signature verification system based on statistical and local features and multistage classifiers has been reported in [9]; a technique based on thickened templates that can be utilised as an initial face of a signature recognition and verification system was built in order to reject signatures that are completely unmatched [10]; a system was developed based on statistical features in combination with neural network classifier [11];a signature retrieval and identification system was developed based on statistical and topologic features [12], *etc.*

In past few years, research on use of moments for pattern/object classification in both invariant and non-invariant tasks has received extensive attention. A diverse form of moment descriptors have been largely engaged as pattern features in image object registration and matching. Hu [13] proposed a set of invariant moments that have the desirable characterstics of being invariant under image translation, scaling and rotation. For using the comprehensive properties of the image, moments are used. In this paper, the fusion of Hu's moments and zone features is used as a features set and verification is done with Support Vector Machine (SVM) method. The major areas of application of off-line signature verification systems are authentication of bank cheques, attendance register monitoring, document authentication and visa applications. One of the main reasons for developing an off-line system is in banks, Forensic Science Labs (FSL's) for detecting the forged and genuine signatures. The proposed system is based on strong feature set in combination of SVM.The rest of the paper is organized as follows. Making of database is explained in section II. Different pre-processing steps are explained in section III. Feature extraction stage is explained in section IV. SVM as a signature verifier and experimental results are presented in section V. Concluding remarks are given in section VI.

II. SIGNATURE DATABASE

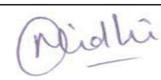
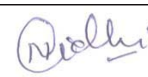
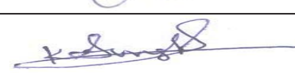
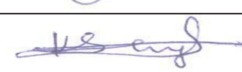
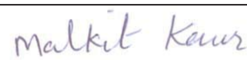
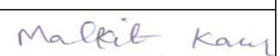
The signature database contains both genuine and skilled forgeries of ten different people. Genuine signature means the original/ authentic signature of the owner/signer whereas a skilled forgery is done by closely imitating the genuine signature of the signer by practising through various time sessions. For correctly categorising the signature class, testing is done with both genuine and forged signature. Three hundred genuine signatures from ten different signers (given identity codes as) have been collected in this study. These signers are of different age groups and from different fields. The signatures are collected during ten days so as to account for the variations in the signatures with interval of time. The A4 size sheet of paper is given to the signers and each signer was requested to give their signature (30 signature images from each signer). Therefore, three-hundred genuine signatures were collected. Also, two hundred forged signatures were collected in this work. The genuine signatures were shown to a new signer and were requested to repeat that genuine signature ten times after going through a practice session, for collecting the forged signatures. Overall, database of 500 signatures has been created. This database consists of 300 signatures (30 signatures each of the 10 signers) and 200 forged signatures (20 signatures each of the 10 users). This database of signatures has further been divided randomly into training database and testing database as shown in Table-1.

Table 1. Training and Testing Databases.

	Genuine Signatures	Forged Signatures	Total
Training Set	**240**	**160**	**400**
Testing Set	**60**	**40**	**100**
Total	**300**	**200**	**500**

These signatures are scanned using HP-scan jet 5400c at 300dpi, which are usually a good quality and a low noise images. The digitised images are stored as JPEG format. A sample of the signatures from the database is shown in Table-2. The block diagram of the proposed recognition system is shown in Fig. 1.

Table 2.Sample Signatures.

Signature Identity Code	Genuine Signature	Skilled Forgery
Id_1		
Id_2		
Id_{10}		

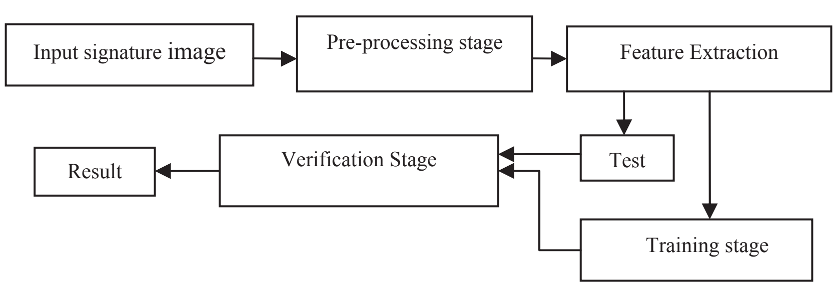


Fig. 1. Block diagram of proposed signature verification system.

In the first stage, both digitized genuine and forged signatures are input which is unprocessed/unrefined images. These dataset are then pre-processed by normalizing, changing to grayscale, skeletonising and by binarising the signature images. At the third stage features including, Hu's moment invariants and zone features are extracted.

III. PRE-PROCESSING OF SIGNATURE IMAGE

The procedure used before processing for smoothing, enhancing, filtering, cleaning-up a digital image so that subsequent algorithms along the way to final classification can be made simple and more accurate [14]. All the pre-processing operations are performed by using Image Processing Toolbox (IPT) under MATLAB 6.0. Following is the four-stage process, used in off-line signature image pre-processing:

A. Normalisation –

To make constant dimensions, the range of pixel intensity values changes in an image, this process is called normalisation. By using resize algorithm with nearest neighbour interpolation, the signature image can be scaled to the standard normalise size. It preserves the aspect ratio of the signature image. In this paper, the digitised signature image is resized to pixels. Fig.2 shows the normalised image of a sample genuine signature ( ), which is used in all consequent illustrations unless specified.

Fig. 2. Normalised image of a sample signature.

B. Greyscales –

A compressed file was made by collecting all the digitised signature images having JPEG format. The compressed file causes loss of picture quality. Whereas, BMP files may be easily produced from existing pixel data stored in a group in memory and retrieving pixel data stored in a bitmap file may often be accomplished by using a set of coordinates. Hence, the JPEG format is converted to BMP format in the proposed system.

C. Skeltonisation (Thinning) –

Reducing the width of signature image to a single black pixel or reducing the connected region in the image to a smaller size and minimum cross-sectional width character is called thinning. By thinning we can remove the thickness differences of pen, making image one pixel thick. But the basis structure and the connectedness of signature is preserved while doing this process. In the proposed system, Zhang [15] algorithm is used. The signature image after thinning is shown in Fig.3. Many features can be extracted from the image having unit thickness.

Fig. 3. Thinned image of the sample signature shown in Fig.2.

D. Binarisation –

Binarisation (changed into) of thinned image is done by using Otsu's binarisation technique [16].

IV. FEATURE EXTRACTION

Many procedures were applied on the pre-processed image of the signatures for measuring the relevant shape information so that the job of classifying the signature image can be made easily. The extraction and selection of the representative stable features is the main task of the pattern recognisation system. In the proposed system, novel feature extraction technique based on fusion of Hu Moments and zone features is used by using Image Processing Toolbox (IPT) of MatLab 6.0. Functions of moments can be employed as the invariant statistical features of an image in pattern recognition, image classification, target identification and scene analysis. Therefore, Hu's seven moment invariants and zone features are collaboratively used as strong features.

A. Hu's Moment Invariant –

In the recent years, research on the utilisation of moments for image characterisation in both invariant and non-invariant jobs has received extensive attention. Historically, Hu [17] published the first significant paper on the utilisation of moment invariants for image analysis and object representation way back in 1961. Hu's approach was based on the work of the nineteenth century mathematicians Boole Cayley and Sylvester, on the theory of algebraic forms. Hu's Uniqueness Theorem states that if is piecewise continuous and has nonzero values only in the finite part of the plane, then statistical moments of all orders exist. Moment set can be computed and used to exceptionally explain the information contained in the image segment.

Hu introduced seven moment invariants. These moments having the desirable properties of being invariant under image scaling, translation, rotation, and shear, this can be defined by following equations (1) through (7):

$$M_1 = (\mu_{20} - \mu_{02}) \quad \cdots (1)$$

$$M_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \quad \cdots (2)$$

$$M_3 = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \quad \cdots (3)$$

$$M_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \quad \cdots (4)$$

$$M_5 = (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 \mathrm{I}\, 3(\mu_{21} + \mu_{03})^2] \quad \cdots (5)$$

$$M_6 = (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 \mathrm{I}\, (\mu_{21} + \mu_{03})^2] + [(3\mu]_{12} - \mu_{03})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \quad \ldots (6)$$

$$M_7 = (3\mu_{21} - \mu_{03})[(\mu]_{30} - \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] \quad \cdots (7)$$

Following is the summary of extracted features from Hu's moment:

- **Centroid**

  It is thr ratio of sum of all x and y locations of black pixels and the total number of pixels. The resulting two numbers (one for x and other for y) is the centroid location.

- **Maximum Horizontal Projection**

  Finding the maximum among the sum of number of black pixels of all the rows in binary and thinned image [6] is called Maximum Horizontal Projection (MHP).

- **Maximum Vertical Projection**

  Finding the maximum from the sum of black pixels column wise in binary and thinned image [17] is called Maximum Vertical Projection (MVP).

- **Mass**

  It is the sum of $1's$ in the thinned binary image.

- **Baseline/Orientation**

The orientation of an image describes how the image lies in the field of view or the directions of the principal axes.

- **Aspect Ratio**
  It is ratio of width to height of the signature image. As the signature image is resized to fix size, so the ratio of maximum number vertical projection to maximum horizontal projection is taken in it.
- **Spreadness**
  It is the sum of 1's in the binary image of an image (before thinning). Table 3 below shows the seven Hu moments of the sampe image shown in Fig.3.

Table 3. Some of the Extracted Hu's Features from Sample Image Shown in Fig. 3.

Feature Identity	Features	Extracted Values of Features
F1	Centroid	61.02,55.27
F2	MHP	49
F3	MVP	43
F4	Mass	3168
F5	Baseline/Orientation	14.57
F6	Aspect Ratio	0.72
F7	Spreadness	4788

B. Zone Features–

Dividing the digitised pre-processed image to the three equal size horizontal rows is called zones. Bysequentially repeating the procedure of extracting Hu moments for all the three zones, we can extract the zone features (7*3=21 features). Fig.4. depicts the the three zones of signature image shown in Fig.3.

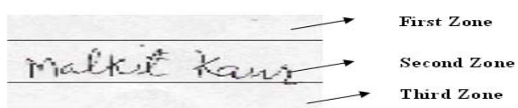


Fig.4.The three zones of signature image shown in Fig.3.

Thus, feature vector is made by the combining the Hu's moments and zone features. These features are then fed to the classifier for successfully recognising the genuine signature.

Overall, twenty- eight features are extracted from the whole signature image, consisting of seven Hu's moment invariant features and twenty-one zone features, which is used as input to the recognition system [7+(7*3)=28].

V. SVM AS A SIGNATURE VERIFIER

SVM is a new learning method introduced by V. Vapnik *et al.* [18, 19]. Between two classes SVM finds hyperplane, which maximises the distance from either class to hyperplane and distinguish the largest possible number of points belonging to the same class on same side, which reduces the misclassification error of both training and testing set. Fig.5 shows the classification between two-classes. This multi- class problem can be solved by SVM. The LIBSVM 3.0 (a library for Support Vector Machine) software has been used to train the SVM model. LIBSVM is simple, easy-to-use, and efficient software for SVM classification and regression.

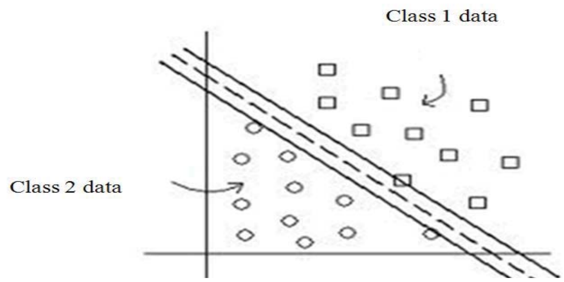


Fig.5. Classification between two classes using SVM.

A. Experimental Results and Discussion –

In all, 500 signatures were collected, out of which 400 signatures were used for training and 100 signatures for the testing purpose. The feature set comprising 28 features extracted by means of the Hu's moment invariants and zone features was used for training SVM. Also, the proposed system was tested based on the same 28 features. The probable kernel options for SVM are linear, polynomial, radial basis function and sigmoid. Empirically, radial basis function (RBF) gave the best results as compare to linear and polynomial kernels. While recognising genuine signatures, 55 genuine signers were rightly recognised, *i.e.*, genuine as genuine; and 4 genuine signers were wrongly recognised, *i.e.*, genuine as forged whereas for recognising forged signatures 35 signatures were correctly recognised, *i.e.*, forged as forged and 2 were wrongly recognised, *i.e.*, forged as genuine. However, 4 signatures were wrongly identified while recognising genuine signatures and signature as forged signature, *i.e.*, signer was identified wrongly, *i.e.*, giving wrong Id while recognising forged and genuine signatures. The Table-4 depicts the confusion matrix for analysing proposed system's accuracy. The Table-5 shows FRR, FAR and False Identification Rate (FIR-the rate of error, when the system / model wrongly identifies the signer, *i.e.*, signer with Id_1 as signer with Id_2) to assess the proposed system's performance. The overall accuracy of the system is 90%.

Table 4: Confusion Matrix to Analyse Accuracy of the Developed Model.

Result of the System	**Actual Condition Truth**		
	(+ve) Genuine	**(-ve) Forged**	**FIR**
(+ve) Genuine	**55**	**4**	**1**
(-ve) Forged	**2**	**35**	**3**

Table 5: Comparative Analysis for Accuracy of the Developed Signature Verification Model.

FAR	**FRR**	**FIR**
2%	**4%**	**4%**

IV. CONCLUSION

A novel signature verification system has been developed that is based on statistical features and Hu's moment invariants. A database comprising of genuine and forged signatures of several people of different age groups, collected at different intervals of time was also developed. No feature reduction practice was used. The signers were asked to use as much variation in their signatures as they should ever use under real circumstances. The signature database was utilised for training the SVM model. The signature verification accuracy of the model has been evaluated in terms of False Acceptance Rate, False Rejection Rate and False Identification Rate. Accordingly, the model described in this paper successfully verifies the off-line signature with 90% accuracy.

V. References

[1] R. Plamondon, J.J Brault," A complexity measure of handwritten curves modeling of dynamic signature forgery", IEEE Transactions on Systems, Man and Cybernetics, vol. 23, pp. 400–413 , 1995.
[2] E.J.R. Justino,F. Bortolozzi, R. Sabourin," Off-line signature verification using HMM for random, simple and skilled forgeries", In: Proceedings of the International Conference on Document Analysis and Recognition, Seattle, Washington, USA, pp. 105 – 110 ,2001.
[3] Q., Qianghua, S. Yaiqian, P. Jingui, " Off-line signature verification using statistical features specific to Chinese handwriting", In: Proceedings of the 24th International Conference on Information Technology Interfaces, pp. 229–235 ,2002.
[4] K. Huang, H. Yan," Off-line signature verification based on statistical feature extraction and neural network classification", Pattern Recognition, vol. 30, pp. 9–17, 1997
[5] V.E Ramesh, M.N. Murty,"Off-line signature verification using genetically optimized weighted feature", Pattern Recognition, vol. 32,pp. 217–233, 1999.
[6] S. Gupta," Face Recognition System using Neural Network", Master of Engineering Thesis submitted in the Department of Electrical and Instrumentation Engineering, Thapar University, Patiala-147004 India, 2008.
[7] S. Ghandali, M. Moghaddam," A method for off-line Persian signature identification and verification using DWT and image fusion", Journal of Multimedia, vol. 4, pp. 137–134, 2009.
[8] A.P Shanker, A.N Rajagopalan, "Off-line signature verification using DTW". Pattern Recognition Letters, vol. 28, pp. 407–414, 2007.
[9] M.A Ismail, S. Gad," Off-line Arabic signature recognition and verification", Pattern Recognition , vol. 33, pp. 1727–1740 , 2002.
[10] O. Yedekcoglu,M.B. Akban,Y.H Lim, " Off-line signature verification with thickened templates", In: Proceedings of the 5th International Conference on Advances in Communication and Control, Crete, Greece, pp.131–142, 1995.
[11] K. Huang, H. Yan, " Off-line signature verification based on statistical feature extraction and neural network classification", Pattern Recognition, vol. 30,pp. 9–17 ,1997.
[12] K. Han, I.K Sethi," Handwritten signature retrieval and identification", Pattern Recognition Letters , vol. 17,pp. 83–90 ,1996.
[13] M.K Hu,"Pattern recognition by moment invariants". In: Proceedings of the IRE, vol.49, pp.1428–1432 ,1961.
[14] A. Sharma, R.K Sharma, R. Kumar,"Online handwritten Gurmukhi strokes pre-processing", International Journal of Machine Graphics and Vision, vol. 18, pp.105–120 ,2009.
[15] T.Y Zhang, C.Y, Suenc, "A fast parallel algorithm for thinning digital patterns", Communications of ACM , vol. 27, pp. 236–239 ,1994.
[16] N. Otsu," A threshold selection method from gray-level histograms", IEEE Transactions on Systems, Man, and Cybernetics, vol. pp.9, 62–66 , 1979.
[17] M.K Hu, " Visual problem recognition by moment invariant", IRE Transactions on Information Theory, vol 8,pp. 179–187 , 1962.
[18] C. Cortes, V. Vapnik, " Support-vector networks", Machine Learning, vol. 20,pp 273-297 , 1995.
[19] V.N Vapnik, "The Nature of Statistical Learning Theory". Springer-Verlag ,1995.