

Real time verification of Offline handwritten signatures using K-means clustering

Alpana Deka¹, Lipi B. Mahanta^{2*}

¹ Department of Computer Science, NERIM Group of Institutions, Guwahati, Assam, India

² Central Computational and Numerical Sciences Division, Institute of Advanced Study in Science and Technology, Vigyan Path, Paschim Boragaon, Guwahati, P.O. Garchuk, Guwahati, Assam, India.

Abstract

Signature is considered as an authentication symbol for any document. Mostly, a document is considered as valid only when there will be an approved signature. Thus in order to reduce the number of frauds in our society, we are required to verify the signatures for different financial documents, banking cheques etc. Depending on the extracted features as well as the device with which the signature is done, signature verification can be offline or online. If during the signing process, special device like stylus is used then that type of verification will be online. But if normal pen is used then it will be offline where only static features can be extracted. In proposed system, verification is done for English scripts based offline handwritten signatures where total database size is 900. Here, the global and local features are extracted from genuine, skilled and simple forgeries. Finally the K-means clustering technique is applied as classifier where approximately 79% and 84% are obtained as accuracies for genuine-skilled and genuine-simple forgeries respectively.

Keywords: *Handwritten Signature; Offline; Preprocessing; Feature Extraction; Classification; k-means*

1. Introduction

Biometrics helps to identify the personal identity by studying physiological and behavior of a person. The physiological biometrics include shape or measurements of iris, retina, face, fingerprint, DNA etc. On the other hand, behavioral identities include characteristics such as signature, voice, keyboard typing etc. (Sabhanayagam et al.,2018). Compared to the other biometrics, handwritten signature is found to be the most widely accepted

biometric behavior (Maghooli, et al.,2017). The advantages of handwritten signature verification system over other existing biometrics systems are: its social and legal acceptance by the society which increases its market popularity, user-friendly behavior and inexpensive prices of hardware devices applied in the application etc. On the other hand it has also some disadvantages because signatures of same person may not be exactly same since the physical, psychological, environmental and timing factor can affect the signatures of a person which enhances the intrapersonal variation. And hence the professional skilled forger may take the advantages of intrapersonal variation to fraud the signatures, which is the main disadvantage of handwritten signatures system (Mohammed et al. 2015). Thus to reduce such disadvantages, research is going on with different features and different methodologies.





Based on the acquisition method, a signature verification system may be online or offline. In online one, static and dynamic features can be extracted. But in offline mode, since no specialized device like stylus is used, therefore dynamic features cannot be extracted, only static features can be evaluated. Hence offline system is more challenging than the online one. Since more challenging and less expensive, therefore in proposed system we have considered the offline one.

The main aim of any signature verification system is to establish the given signature as either genuine or forgery and hence it can be considered as a binary verification system. Here, the genuine signature represents the original signature signed by the actual/owner of the signature. Again the forgery means the traced or the stolen signature signed by other person. Depending on the variation between original and traced signatures, the forgery can be of three types: random, simple and skilled (**Table 1**). Among all the three types of forgery signatures,

detection of the skilled forgery is most difficult (Majhi et al. 2006).

Since, random forgery can be detected with visual inspection also, therefore in proposed system we have concentrated with genuine, skilled and simple forgeries but not the random one. The remaining part of the paper is organized as follows: section 2 describes the data acquisition, section 3 defines the preprocessing, section 4 introduces the features extraction, section 5 illustrates the results of applied classifier, and section 6 gives performance evaluation, section 7 compares the existing system with other systems and section 8 ends with the conclusion.

Table 1: Pictorial Comparison of Genuine Signature with Three Types of Forgery Signatures

Original/fake signatures	Nature of signature
Original signature	
Random Forgery	
Simple Forgery	
Skilled Forgery	

2. Data Acquisition

In proposed system, signature samples are collected with ink or ball pen on white piece of papers from 60 persons. By trial and error method, the 360 (genuine) and 540 (genuine + forgery) samples are taken as training and testing datasets respectively. For training, 6 genuine signatures are taken against each person and for testing, we have used 9 forged signatures each from genuine, skilled and simple category, for each of the respective persons. The signatures are extracted with a scanner and the final digitalized form is taken as input image for preprocessing step.

3. Preprocessing

Preprocessing is just the previous step of features extraction which helps us to obtain a more detailed image of a signature. Preprocessing is required to improve the quality of the images such that removal of unwanted spaces, noises, involuntary scratches etc. can be done (Sharif et al. 2018). The pre-processing steps included in our methodology is shown in **Figure 1**.

3. Feature Extraction

In proposed system, we have taken global as well as local features to improve the system performance. These features are illustrated below (Sashi Kumar et al., 2010; Ahmed et al., 2012; Rathi et al., 2012; Azzopardi et al., 2006; Kisku et al., 2010; Biswas et al., 2010; Patil et al., 2013). Here, the first five features such as IA, HWR, NA, MHP, MVP are global features and the last one is local feature.

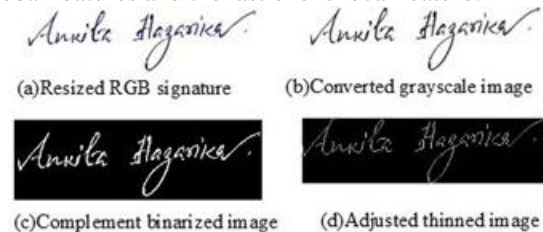


Figure 1: Some of the pre-processing steps of proposed system

- Image Area (IA): IA is calculated by taking the sum of on pixels of the signature.
- Height to Width Ratio (HWR): HWR is evaluated just by taking the ratio of height and width of the signature sample.
- Normalized Area of the signature (NA): NA is as of the following:

$$\text{Normalised area} = \frac{\text{Area of the signature}}{\text{height} \times \text{width}}$$
- Maximum Horizontal Projection (MHP): MHP is the row with maximum number of white pixels along horizontal direction.
- Maximum Vertical Projection (MVP): MVP is the column which gets maximum number of white pixels along vertical direction.
- Sum of Local Normalized Areas of signature (SLNA): Here, the whole image is vertically partitioned into four equal divisions. Now, the normalized area is calculated from each of the part. Then finally, SLNA is obtained by adding these four local normalized areas.

Table 2: Values of extracted features of training sample

IA	HWR	NA	MVP	MHP	SLNA
827	0.3185	0.0305	18	34	0.1218
797	0.2755	0.0335	20	32	0.1336
808	0.2963	0.0309	17	32	0.1237
817	0.2458	0.0377	14	38	0.1507
757	0.2752	0.0310	11	35	0.1238
747	0.2635	0.0324	17	31	0.1294

From **Table 2**, it is clear that although all the six genuine signature samples are collected from the same original signer, but still the measurements vary among themselves due to the psychological, environmental factors etc. By considering Table 2, the boxplots for each of the features are displayed in **Figure 2**, from where median, minimum and

maximum values for each of the features are easily understandable.

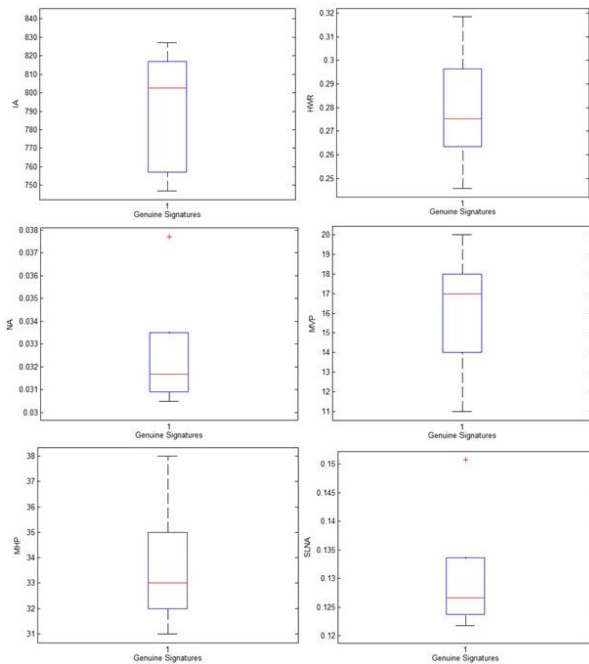


Figure 2: Boxplots for each of the features

As an example, the calculated values of features of Figure 1(d) are given below in Table 2:

This section must contain specific details about the materials studied, instruments used, specialized chemicals source and related experimental details which allow other research worker to reproduce the results. The journal will not be held responsible if any kind of plagiarism followed and the editor's decision would be final if any litigation arises during processing or after publishing.

5. Results

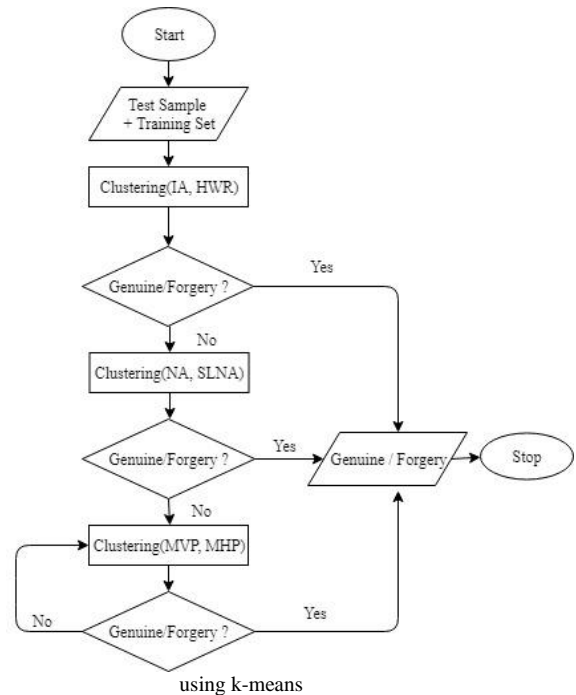
In proposed system, we have decided to apply the K-means clustering technique since, as far as our knowledge, it is not so much used as a classifier to recognize a given signature as genuine or forgery. Here, by trial and error method, number of clusters is taken as 2.

The main steps of verification stage are mentioned below:

1. For each of the iterations, two features are taken at a time as displayed in **Figure 3**.
2. Here, every time two clusters will be generated where we will count the number of data items of the cluster where the testing data presents.
3. Now, if the number of data items including the testing data is 2 i.e. if the testing data creates a cluster with any one of the training data then that signature will be considered as genuine. On the other hand, if the testing data does not create a cluster with any other training data then this will

be recognized as forgery signature. That means the clustering process will stop when the number of data items (including testing data) becomes either 2 or 1. Here, following figure (Figure III) illustrates the method of applying k-means in our proposed system.

Figure 3: Flowchart for detection of genuine/forgery signature



Verification of each type of signature is illustrated below separately.

- i) Verification with **genuine** signature: For illustration, we have considered the signature as given in Figure 1(d), as a testing signature. For that the testing set is given in **Table 3**.

Table 3: Values of extracted features of test (genuine) sample

IA	HWR	NA	MVP	MHP	SLNA
769	0.2705	0.0333	14	28	0.1333

training and testing sets respectively, the result of K-means is given in **Table 4**.

Table 4: Results of k-means clustering technique

Number of iterations	Number of data items including testing data in testing cluster
1	3
2	2

Since, from the above table, number of data item in last iteration is found to be 2, it indicates that the testing data is making a group with genuine data

of training set. Hence the tested signature is recognized as genuine.

ii) Verification with **skilled** forgery signature:

Suppose a signature is given to us as in **Figure 4**.



Figure 4: Before and after preprocessing of the skilled forgery signature

The training dataset for the above signature will be same as in Table 2. Now testing dataset is given below in **Table 5**.

Table 5: Values of extracted features of test (skilled) sample

IA	HWR	NA	MVP	MHP	SLNA
722	0.3287	0.0269	26	27	0.1073

The verification result is given below (**Table 6**):

Number of iterations	Number of data items including testing data in testing cluster
1	3
2	1

Table 6: Results of k-means clustering technique

Again, Table 6 (since in last iteration, number of data item is obtained as 1) indicates that the testing data is not making a group with any other genuine data of training set. Hence we can conclude that the tested signature is forgery.

iii) Verification with **simple** forgery signature:

Suppose we are given signature to verify as in below (**Figure 5**):



Figure 5: Before and after preprocessing of the simple forgery signature

Similarly, as in the skilled forgery, the training set will be same. The values of testing set are given in **Table 7**.

Table 7: Values of extracted features of test (simple) sample

IA	HWR	NA	MVP	MHP	SLNA
552	0.1468	0.0438	19	34	0.1751

Table 8: Values of extracted features of test (simple) sample

Number of iterations	Number of data items including testing data in testing cluster
1	1

From the above **Table 8**, since the data item in last iteration does not make group with any training data item, hence the signature will be forgery signature.

6. Performance Evaluation

We have calculated efficiency of the proposed system in terms accuracy, FRR, FAR for genuine-skilled and genuine-simple signatures as given below.

$$\text{Accuracy} = \frac{(A+B)}{(A+\bar{A}+B+\bar{B})} \times 100$$

Here,

A: A genuine signature recognized as genuine,

B: A forgery signature recognized as forgery,

\bar{A} : A genuine signature recognized as forgery and

\bar{B} : A forgery signature recognized as genuine.

$$\text{FRR} = \frac{\text{Number of original signatures rejected}}{\text{Number of original signatures tested}} \times 100$$

$$\text{FAR} = \frac{\text{Number of forgery signatures rejected}}{\text{Number of forgery signatures tested}} \times 100$$

Table 9: Results in terms of performance measurements

Performance measurement	Result
Accuracy (genuine-skilled)	79% (approx.)
Accuracy (genuine-simple)	84% (approx.)
FRR	20% (approx.)
FAR (genuine-skilled)	23% (approx.)
FAR (genuine-simple)	12% (approx.)

Figure 6 depicts the performance comparison of the proposed methods.

6. Comparison of the proposed system with existing systems

Here, we have compared the efficiency of the proposed system with other existing systems which are based on clustering technique as given in **Table 10** (Sikha et al., 2013; Suryani et al., 2017).

7. Conclusion

Development of a real time software for verification of handwritten signatures is a very important application which would reduce the burden of experts in this area and solve many unnecessary forgery cases. The proposed system is free from the limitations such as the persons can sign with any pen either ink or ball pen, any colors of the ink, the signer can sign with short or long signatures. The proposed database size was 900. The

verification result with K-means is found to be better than the other existing clustering techniques.

References

- [1] Ahmed, H. and Shukla, S. "Comparative Analysis of Global Feature Extraction Methods for Off-line Signature Recognition", *International Journal of Computer Applications*, 48(23), (2012).
- [2] Azzopardi, G. "How Effective are Radial Basis Function Neural Networks for Offline Handwritten Signature Verification?", A project report for the Degree of B.Sc. in Computing and Information Systems, University of London, (2006).
- [3] Biswas, S., Kim, T. H. and Bhattacharyya, D. "Features Extraction and Verification of Signature Image using Clustering Technique", *International Journal of Smart Home*, 4(3), (2010).
- [4] D. R. S. K., Raja, K. B., Chhotaray, R. K. and Pattanaik, S. "Off-line Signature Verification Based on Fusion of Grid and Global Features Using Neural Networks", *International Journal of Engineering Science and Technology*, 2(12), (2010).
- [5] Kisku, D. R., Gupta, P. and Sing, J. K. "Offline Signature Identification by Fusion of Multiple Classifiers using Statistical Learning Theory", *International Journal of Security and Its Applications*, 4(3), (2010).
- [6] Maghooli, Y. G. K., Nasrabadi, A. M. and Moein, M. S. "A new method for signature verification based on physiological characteristics of hand muscles and tendons", *Biomedical Engineering: Applications, Basis and Communications*, 29(1), (2017).
- [7] Majhi, B., Reddy, Y. S. and Babu, D. P., 2006: "Novel Features for Off-line Signature Verification", *International Journal of Computers, Communications & Control*, 1(1).
- [8] Mohammed, R. A., Nabi, R. M., Mahmood, Sardasht M. R. and Nabi R. M. "State-of-the-Art-in Handwritten Signature Verification System" *International Conference on Computational Science and Computational Intelligence*, (2015).
- [9] Patil, P. and Patil, A. "Offline Signature Recognition Using Global Features", *International Journal of Emerging Technology and Advanced Engineering*, 3(1), (2013).
- [10] Rathi, A., Rathi, D. and Astya, P. "Offline Handwritten Signature Verification by Using Pixel based Method", *International Journal of Engineering Research & Technology*, 1(7), (2012).
- [11] Sabhanayagam, T., Venkatesan, V. P. and Senthamaraikannan, K. "A Comprehensive survey on various biometric systems", *International Journal of Applied Engineering Research*, 13(5), (2018).
- [12] Sharif, M., Khan, M. A., Faisal, M., Yasmin, M. and Fernandes, S. L. "A framework for offline signature verification system: Best features selection approach", *Pattern Recognition Letters*, (2018).
- [13] Shikha, P. and Shailja, S. "Neural Network Based Offline Signature Recognition and Verification System", *Research Journal of Engineering Sciences*, 2 (2), (2013).
- [14] Suryani, D., Irwansyah, E. and Chindra, R. "Offline signature recognition and verification system using efficient fuzzy kohonen clustering network (EFKCN) algorithm", *Procedia Computer Science*, 116, (2017).

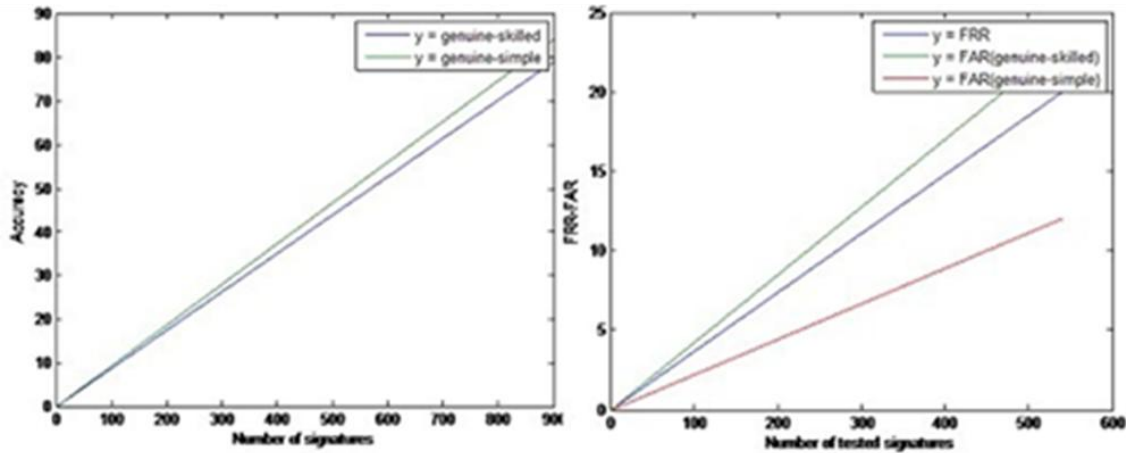


Figure 6: Pictorial representation of performance measurement of proposed system

Table 10: Comparison of the proposed system with other systems based on clustering technique

Author	Database size	Applied classifier	Performance measurement	Comment
Shikha et al. (2013)	60 (genuine, skilled, random)	Self-Organizing Map groups (SOM) neural network	FRR=10% FAR=12.5% AER=11.25%	i) Very small database size as compared to us ii) Not clearly mentioned for which type of forgery, the FAR is calculated.
Dewi et al. (2017)	80	Efficient Fuzzy Kohonen Clustering Network (EFKCN)	Accuracy = 70%	i) Very small database size as compared to us ii) Not clearly mentioned for which type of forgery, the FAR is calculated.
Proposed system	900 (genuine, skilled, simple)	K-means	Genuine Vs Skilled Accuracy=79% (approx.) FRR=20% (approx.) FAR=23% (approx.) Genuine Vs Simple Accuracy=84% (approx.) FRR=20% (approx.) FAR=12% (approx.)	i) Database size of the proposed system is larger than above mentioned systems. ii) We have evaluated the performance for genuine Vs skilled and simple forgeries respectively separately.