2009 10th International Conference on Document Analysis and Recognition

# Geometric Centroids and their Relative Distances for Off-line Signature Verification

H.N. Prakash and D.S. Guru
*Department of Studies in Computer Science, University of Mysore, Mysore-570 006, India.*
*prakash_hn@yahoo.com, dsguruji@lycos.com.*

## Abstract

*In this paper, we propose a new approach for symbolic representation of off-line signatures based on relative distances between centroids useful for verification. Distances between centroids of off-line signatures are used to form an interval valued symbolic feature vector for representing signatures. A method of off-line signature verification based on the symbolic representation is presented. We investigate the feasibility of the proposed representation scheme for signature verification on a MCYT_ signature database [1]. We cluster similar signatures in each class and also investigate the cluster based symbolic representation for signature verification. Unlike other signature verification methods, the proposed method is simple and efficient. Several experiments are conducted to demonstrate the efficacy of the proposed scheme.*

**Keywords**: Off-line signature verification; Relative distance between Geometric centroid; Symbolic feature vector; Fuzzy c means clustering.

## 1. Introduction

Signature verification is a difficult discrimination problem as handwritten signature is a result of complex physical and psychological processes. In the last two decades many features are proposed and many methods have been developed for signature verification. Several statistical, geometric and structural features have been proposed for off-line signature verification. However, no feature (set of features) guarantee a good level of performance and hence off-line signature verification is still a challenging task.

Concerning off-line signature features, many parameters are extracted by the geometric analysis of the signature. The most diffuse parameters are signature image area, signature height and weight, length to width ratio, number of characteristic points [2]. In addition, direction based features, slant-based features, orientation based features, contour based features, grid based features, texture based features and spectrum based features [2] are commonly used for signature verification.

In verification, the authenticity of a test signature is evaluated by matching its features against those stored in the knowledgebase. For matching various pattern recognition strategies like Neural Networks [3], Time Warping [4], Hidden Markov Model (HMM) [5] and Support Vector Machine (SVM) [5] have been employed.

Symbolic data [6] appear in the form of continuous ratio, discrete absolute interval and multi-valued, multi-valued with weightage, quantitative, categorical, etc. Recently, a symbolic representation model for 2D shapes has been proposed and it has also shown that symbolic representation model effectively captures shape information [7]. Recently, we have proposed relative centroid orientations for off-line signature verification [8]. In this paper, the distances between geometric centroids are proposed for off-line signature verification and in addition cluster based symbolic representation for signature verification is also presented. The distances between geometric centroids are used to form interval-type of symbolic representation. A method of signature verification based on symbolic representation is presented.

The rest of the paper is structured as follows: In section 2, extraction of geometric centers, method of symbolic representation and verification of off-line signatures are presented. In section 3, the details of the experimentations and results are summarized. Section 4, compares the proposed methodology with other similar works. Finally, the conclusions are drawn in section 5.

## 2. Proposed method

In this section, we present a methodology for the extraction of geometric centers (centroids),

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.67

methodology for symbolic representation of signatures and also cluster based symbolic representation.

## 2.1. Extraction of geometric centroids

The geometric centers represent the pixel distribution of the signature image which in turn depends on handwritten signature pattern. In the proposed system signature image is binarized using the histogram based global threshold [9]. Then, we find the geometric centroid of the image [14] and then we split the signature image vertically at the geometric centroid to get two partitions. In the next step, we find the geometric centroid of each partition to split each of the partitions horizontally at their geometric centroids. This procedure of finding centroids and splitting the partitions at the centroids is continued recursively vertically and horizontally in an alternative way till a desired depth of the splitting is reached. Generally we extract $n = [(2)^{r} - 1]$ centroids, where $r = 1, 2, 3, .., k.$, where $r$ is the depth of the splits. Centroids extracted for each split portions are labeled as 1, 2, 3,…, $n$ in sequence as shown in Fig.1.

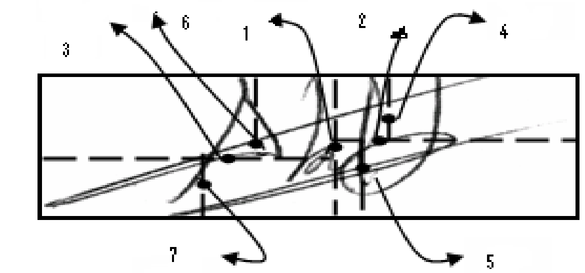


Fig.1. Geometric centroids of a signature image

## 2.2. Symbolic representation of signature

Recently, we proposed on-line signature verification model based on symbolic representation using global features [10] which has shown a good verification performance. In the present work we use relative distances between geometric centroids for symbolic representation of off-line signatures.

Our approach involves extracting geometric centers as explained in previous section 2. Let the first geometric center is labeled as '*1*' and the second as '*2*' and so on and so forth until '$n$', the last geometric point. We illustrate the proposed methodology with $n$ = 5 geometric centers. A graph of edges joining '$n$' geometric centers is envisaged in Fig. 2.

A vector $S$ consisting of the distances of all the edges form the symbolic representation of a signature and is given by

$$s = \{d_{12}, d_{13}, ..., d_{1n}, d_{23}, d_{24}, ..., d_{ij}, ..., d_{n-1n}\} \quad \textbf{(1)}$$

Where $d_{ij}$ is the distance of the edge from node $i$ to node $j$, $1 \le i \le n-1$, $2 \le j \le n$, and $i < j$.

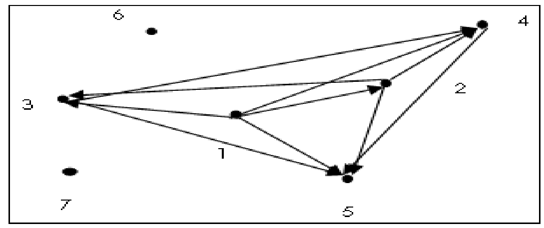


Fig. 2 Geometric centroids with labels as nodes with corresponding edges

For $n$ geometric centers we get $(n(n-1))/2$ distances.

The distances of corresponding edges of signature samples of a class have considerable intra class variations. Thus, we propose to have an effective representation by capturing these variations through their assimilation by the use of interval-valued distances called symbolic feature vector as follows

Let $[S_1, S_2, S_3, ..., S_n]$ be a set of $n$ samples of a signature class say $C_j$ ; $j = 1,2,3,...,N$ ( $N$ denotes number of individuals) and let $F_i$ = $[d_{i1}, d_{i2}, d_{i3}, d_{i4}, ..., d_{im}]$ be the set of $m$ features (distances) of edges of a signature characterizing the signature sample $S_i$ of the class $C_j$. Let $\mu_{jk}$ ; $k = 1, 2, ..., m$ be the mean of the $k^{th}$ features (distances) obtained from all $n$ samples of the class $C_j$. i.e.,

$$\mu_{jk} = \frac{1}{n}\sum_{i=1}^{n} d_{ik} \qquad (2)$$

Similarly, let $\sigma_{jk}$ ; $k = 1, 2, ..., m$ be the standard deviation of the $k^{th}$ feature values obtained from all the $n$ samples of the class $C_j$.

$$\text{i.e.,} \sigma_{jk} = \left[\frac{1}{n}\sum_{i}^{n}\left(d_{ik} - \mu_{jk}\right)^{2}\right]^{\frac{1}{2}} \qquad (3)$$

Now, we recommend to capture intra class variations in each $k^{th}$ distance value of the $j^{th}$ class in the form of an interval valued feature values $[d^{-}_{jk}, d^{+}_{jk}]$, where

$$d^{+}_{jk} = \mu_{jk} - \tau_{jk} \text{ and } d^{+}_{jk} = \mu_{jk} + \tau_{jk} \qquad (4)$$

where, $\tau_{jk}$ is the feature level threshold which is a function of $\sigma_{jk}$, given by $\tau_{jk} = \alpha\sigma_{jk}$ for some scalar $\alpha$. Thus, each interval $[d^{-}_{jk}, d^{+}_{jk}]$ $[d^{-}_{jk}, d^{+}_{jk}]$ representation depends on mean and its standard deviation of respective individual features. The interval $[d^{-}_{jk}, d^{+}_{jk}]$ represents the upper and lower limits of a feature value of a signature class in the knowledgebase. Now, the reference signature of class $C_{j.}$ is formed by representing each distance (*k = 1, 2, 3,…, m*) in the form of an interval and is given by

$$RF_j = \{[d^{-}_{j1}, d^{+}_{j1}], [d^{-}_{j2}, d^{+}_{j2}], ..., [d^{-}_{jm}, d^{+}_{jm}]\} \quad (5)$$

It shall be noted that unlike conventional feature vector, this is a vector of interval-valued features and this symbolic feature vector is stored in the knowledge base as a representative of the $j^{th}$ signature class. Thus, the knowledgebase has *N* number of symbolic vectors each corresponding to a class.

We call the above method of symbolic representation as conventional method of symbolic representation. In the above method we use all signatures of each class to obtain a symbolic feature vector instead we can group signatures of each class into several clusters and each cluster is represented by a symbolic vector. We call this method as cluster based symbolic representation**.** These symbolic feature vectors are stored in the knowledgebase as a representative of a class. Thus if signatures in each class are grouped into *M* clusters then, the knowledgebase will have *M* × *N* symbolic vectors because of N classes.

### 2.3. Proposed signature verification model

The signature verification technique proposed in this work considers a test signature, which is described by a set of *m* distances of type crisp and compares it with the corresponding interval type features (distances) of the respective symbolic reference signature $RF_j$ stored in the knowledgebase to ascertain the authenticity.

Let $F_t$ = $[d_{t1}, d_{t2}, d_{t3}, d_{t4}, ..., d_{tm}]$ be an *m* dimensional vector (of distances between geometric centroids) describing a test signature. During signature verification process each $k^{th}$ distance (feature) value of the test signature is compared with the corresponding interval in $RF_j$ to examine whether the test signature feature value lies within the corresponding interval. The number of features of a test signature, which fall inside the corresponding interval of the respective reference signature, is defined to be the degree of authenticity. We define the degree of authenticity by an acceptance count $A_c$ to decide if signature is authentic is as follows [10]

$$A_c = \sum_{k=1}^{m} C(d_{tk}, [d^{-}_{jk}, d^{+}_{jk}]) \quad (6)$$

where,

$$C(d_{tk}, [d^{-}_{jk}, d^{+}_{jk}]) = \begin{cases} 1 & if\,(d_{tk} \geq d^{-}_{jk} and\, d_{tk} \leq d^{+}_{jk}) \\ 0 & otherwise \end{cases} \quad (7)$$

All the features of a test signature falling into respective feature interval of reference signature contribute a value 1 towards acceptance count and there will be no contribution from other features which fall outside the interval. If the acceptance count $A_c$ is greater than a predefined threshold *(T),* then the test signature is considered as authentic otherwise as a forgery. In the case of cluster based representation we compare the test signature with reference signatures which represent the clusters in similar fashion.

## 3. Experimentation and results

*The dataset:* The *MCYT*-75 off-line signature corpus [1] consists of 2250 signatures from 75 individuals. Each individual class consists 30 signatures, out of which 15 are genuine and remaining 15 are skilled forgeries. Totally it forms a signature database of 1125 (i.e. 75×15) genuine and 1125 (i.e. 75×15) forged online signatures.(see Fig. 3)

**Fig.3 Sample off-line signatures from *MCYT_* signature corpus**

*Experimental setup:* The *MCYT_*signature subcorpus is split into training and testing sets. We trained the system with training set of 5, 7 and 9 genuine signatures of each individual. The test set consists of the remaining samples of genuine signatures and all the forgery signatures. Our procedure is similar to the international signature verification competition SVC 2004.

We have used normalized distances between centroids for our experimentations to make the system scale independent. For evaluation of the proposed method for verification performance, in this work we adopt AER (Average Error Rate), which is average of FAR (False Acceptance Rate) and FRR (False Rejection Rate). How many features (distances between centroids) of test signatures should match with that of the reference features for optimal AER decides the value for threshold. We have empirically fixed up the threshold *(T) = m/2* for our experiments so that AER is minimum. The feature level threshold $\alpha$, allows variable width representation for a feature value is set ($\alpha = 1$) so that the AER is minimum. Thus the operating point for our experimentation is set by empirically fixing up the values for *T* and $\alpha$.

### 3.1. Results of conventional symbolic representations

The variations of FAR and FRR for various training samples and various number of geometric centers (centroids) are given in Tables 1-3. We measure the verification performance in terms of commonly used average error rate (AER).

Table 1. Verification performances for 31 centroids, Threshold = 233

| Training Samples per class | FRR | FAR | AER |
|---|---|---|---|
| 5 | 42.53 | 19.82 | 30.50 |
| 7 | 32.83 | 24.04 | 28.31 |
| 9 | 27.77 | 26.11 | 26.90 |

Table 2. Verification performances for 63 centroids, Threshold = 977

| Training Samples per class | FRR | FAR | AER |
|---|---|---|---|
| 5 | 37.20 | 20.26 | 28.73 |
| 7 | 26.16 | 26.13 | 26.10 |
| 9 | 20.22 | 29.51 | 24.86 |

Table 3. Verification performances for 127 centroids, Threshold = 4001

| Training Samples per class | FRR | FAR | AER |
|---|---|---|---|
| 5 | 37.20 | 21.06 | 28.23 |
| 7 | 22.83 | 26.57 | 24.12 |
| 9 | 19.11 | 24.11 | **21.61** |

### 3.2. Results of cluster based symbolic representations

The training signatures are clustered using Fuzzy C-means (FCM) [11] method. The cluster representation and verification are carried out as proposed in section 2. The results are summarized in Tables 4-6 for varying number of centroids and clusters. We use 9 genuine signatures from each class to form different number of clusters.
The comparison between best results (for different number of centroids) of symbolic and cluster based representations is given in Table 7. It can be observed that the best result (minimum AER) is obtained for 127 centroids (for conventional symbolic representation minimum AER = 21.61 and cluster based representation minimum AER = 22.14).

Table 4. Verification performances for 31 centroids, Threshold = 233

| No. of Clusters | FRR | FAR | AER |
|---|---|---|---|
| 2 | 26.04 | 24.88 | 25.46 |
| 3 | 25.06 | 25.00 | 25.08 |
| 4 | 21.15 | 28.88 | 25.01 |
| 5 | 18.84 | 34.44 | 26.86 |

Table 5. Verification performances for 63 centroids, Threshold = 977

| No. of Clusters | FRR | FAR | AER |
|---|---|---|---|
| 2 | 22.04 | 22.55 | 22.29 |
| 3 | 21.55 | 27.91 | 24.74 |
| 4 | 28.62 | 20.44 | 24.53 |
| 5 | 20.53 | 29.77 | 25.15 |

Table 6. Verification performances for 127 centroids, Threshold = 4001

| No. of Clusters | FRR | FAR | AER |
|---|---|---|---|
| 2 | 24.06 | 20.22 | **22.14** |
| 3 | 30.33 | 16.88 | 23.60 |
| 4 | 18.44 | 30.39 | 24.68 |
| 5 | 42.22 | 11.46 | 26.84 |

Table 7. Comparison of verification performances between symbolic representations and cluster based representations

| Method | 31 centroids | 63 centroids | 127 centroids |
|---|---|---|---|
| Symbolic representation | 26.90 | 24.86 | **21.61** |
| Cluster based representation | 25.01 | 22.29 | 22.14 |

## 4. Comparison with other methods

It is veryF difficult to compare the performances of different signature verification systems because different systems use different signature databases. Hence here we list the performances of different systems and our system with respect to size of database and the number of writers. From the comparison (see Table 8) it is clear with the large database size the proposed system yields lower AER and hence the performance of the system is encouraging. In literature, an other model which makes use of centroids as features is reported in [14]. However, it employs directly the Euclidean distance between the centroids of a test signature and that of the stored signature and hence it is not invariant to scaling. Thus, the performance is reported only on a small database of their own. So, we feel it is not required to consider for comparative study.

Table 8. Comparison with other methods

| Similar works | No. of Writers | Data base Size | AER (%) |
|---|---|---|---|
| (1) Proposed method | **75** | **2250** | **21. 6** |
| 2) Shankar A. P. and Rajagopalan [4] | 100 | 1431 | 35.0 |
| 3) Srihari et., al [12]<br>a) Distance Threshold (GSC)<br>b) Distance statistics<br>c) Naïve Bayes<br>d) One Class- SVM | 55 | 1320 | <br>21.5<br><br>22.4<br>25.0<br>46.0 |
| 4) Fang B. et. al[13]<br>(a) 2D elastic matching<br>(b) Horizontal and vertical projections<br>(c) Global shape features | 55 | 1320 | <br>23.4<br><br>22.3<br><br>22.8 |

## 5. Conclusions

In this paper, we have proposed a new method for symbolic representation of off-line signatures based on novel relative distances between centroids as features useful for verification. The main finding of this work is that off-line signature verification using symbolic representation approach achieves reduction in AER. We have made a successful attempt to explore the applicability of symbolic data concepts for off-line signature verification using the relative distances between centroids as features. The results obtained by the proposed method as a stand-alone approach is very impressive. The proposed approach shows the lower AER (AER =21.6) than many other existing stand-alone approaches of verification found in the literature. As this method of signature verification is simple and takes less processing time compared to DTW, SVM, HMM based systems it could be used for industrial biometric applications

## References


[1] J. Ortega-Garcia, J. Fierrez-Aguliar and D. Simon, 2003. "MCYT corpus: A bimodal database". IEE Proceedings Vision, Image and Signal Processing, pp.395-401.

[2] G. Dimauro, S. Impedovo, M.G. Lucchese, R. Modugno and G. Pirlo, 2004. "Recent advancement in automatic signature verification". Proceedings of 9th International Workshop on Frontiers in Handwriting Recognition (IWFHR), pp. 179-184.

[3]R. Bajaj and Chaudhary S., 1997. "Signature verification using multiple neural classifiers". Pattern Recognition, Vol. 30, pp. 1-87.

[4] A.P. Shankar and A. N. Rajagopalan, 2007. "Off-line signature verification using DTW". Pattern Recognition Letters, Vol. 28, pp. 1407-1414.

[5] E.J.R. Justino, F. Bortolazzi and R. Sabourin. 2005. "A comparison of SVM and HMM classifiers in the off-line signature verifications". Pattern Recognition Letters, Vol.26. Issue 9, pp.1377-1385.

[6] H.H. Bock and E. Diday, Analysis of symbolic data, Springer Verilag publication, 1999.

[7] D.S. Guru. and H. S. Nagendraswamy, 2006. "Symbolic representation of two-dimensional shapes". Pattern Recognition Letters, Vol.28, pp.144-155.

[8] H.N. Prakash and D.S. Guru. 2009. "Relative orientations of Geometric Centroids for Off-line Signature Verification", International Conference on Advanced Pattern Recognition, (ICAPR-2009),ISI, Kolkata, India pp. 201-204.

[9] N. Otsu, 1994, "A threshold selection method from grey level histogram". IEEE Transactions on Systems, Man and Cybernetics, Vol. 9, pp.62-66.

[10] D.S. Guru and H.N. Prakash, 2009 "On-line signature verification and recognition: An approach based on symbolic representation". IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 31(6), pp. 1059-1073.

[11] J.C. Bezdez, Pattern Recognition with Fuzzy Objective Algorithms, Plenum, New York, 1981.

[12] S.N. Srihari, Aihua Xu and M.K. Kalera, 2004. "Learning strategies and classification methods for off-line signature verification". Proceedings of 6th International Workshop on Frontiers in Handwriting Recognition (IWFHR), Tokyo, Japan ,2004, IEEE, pp. 161-166.

[13]B. Fang, Leung C. H.,Tang Y. Y., Tse K. W, Kwok P. C. K and Wong Y. K.,2003. "Off-line signature verification by tracking of feature and stroke position". Pattern recognition, 2003, Vol. 36, pp.91-101.

[14]B. Mahji, Y. Santhosh Reddy. D. Prasanna Babu. "Novel features for off-line signature verification", International Journal of Computers and Control, Vol. 1, 2006, pp. 17-24.