Improved Offline Signature Verification Scheme
Using Feature Point Extraction Method

Debasish Jena¹, Banshidhar Majhi², Saroj Kumar Panigrahy³, Sanjay Kumar Jena⁴
¹Centre for IT Education, Bhubaneswar, 751010, Orissa, India
²,⁴National Institute of Technology Rourkela, 769008, Orissa, India
³Roland Institute of Technology, Berhampur, 761008, Orissa, India
debasishjena@hotmail.com, bmajhi@nitrkl.ac.in, skp.nitrkl@gmail.com, skjena@nitrkl.ac.in

Abstract

In this paper a novel offline signature verification scheme has been proposed. The scheme is based on selecting 60 feature points from the geometric centre of the signature and compares them with the already trained feature points. The classification of the feature points utilizes statistical parameters like mean and variance. The suggested scheme discriminates between two types of originals and forged signatures. The method takes care of skill, simple and random forgeries. The objective of the work is to reduce the two vital parameters False Acceptance Rate (FAR) and False Rejection Rate (FRR) normally used in any signature verification scheme. In the end comparative analysis has been made with standard existing schemes.

Index Terms: Offline signature, Geometric centre, Feature point, Forgeries, Euclidean Distance Model, FAR (False Acceptance Rate), FRR (False Rejection Rate).

1. Introduction

Signature verification is an important research area in the field of personal authentication. The recognition of human handwriting is important concerning about the improvement of the interface between human-beings and computers [1, 8]. If the computer is intelligent enough to understand human handwriting it will provide a more attractive and economic man-computer interface. In this area signature is a special case that provides secure means for authentication, attestation authorization in many high security environment. The objective of the signature verification system is to discriminate between two classes: the original and the forgery, which are related to intra and interpersonal variability [1]. The variation among signatures of same person is called Intra Personal Variation. The variation between originals and forgeries is called Inter Personal Variation.

This paper is organized in the following sections. Section 1.1 provides the different types of forgeries. Section 2, introduces new feature extraction method. Section 3 discusses classification based on Euclidean distance model. Section 4 discusses about threshold
selection. Section 5 shows training, testing and results and Section 6 gives concluding remarks.

1.1. Types of Forgeries

There are three different types of forgeries to take into account. First one is random forgery which is written by the person who doesn’t know the shape of original signature. The second, called simple forgery, which is represented by a signature sample, written by the person who knows the shape of original signature without much practice. The last type is skilled forgery, represented by a suitable imitation of the genuine signature model [4]. Each type of forgery requires different types of verification approach [5]. Hybrid systems have also been developed [10]. Figure 1 shows the different types of forgeries and how much they vary from original signature [1].

Figure 1. (A) Original signature, (B) Random forgery, (C) Simple forgery, (D) Skilled forgery

2. Feature Extraction

The geometric features are based on two sets of points in 2-dimensional plane [7]. The vertical splitting of the image results thirty feature points(v1,v2,v3,........,v30) and the horizontal splitting results thirty feature points(h1,h2,h3,........,h30).These feature points are obtained with relative to a central geometric point of the image. Here the centered image is scanned from left to right and calculate the total number of black pixels. Then again from top to bottom and calculate the total number of black pixels. Then divide the image into two halves w.r.t. the number of black pixels by two lines vertically and horizontally which intersects at a point called the geometric centre. With reference to this point we extracted 60 feature points: 30 vertical and 30 horizontal feature points of each signature image.

2.1. Processing of the Signature

The geometric features proposed here are based on two sets of points in two-dimensional plane. Each set having thirty feature points which represent the stroke distribution of signature pixels in image. These sixty feature points are calculated by Geometric Center. Vertical Splitting and Horizontal Splitting are two main steps to retrieve these feature points. Before finding feature points we have to do some adjustments to the signature image [1]. The processing of the signature is discussed below.

2.2 Moving signature into the centre of image

The signature is moved to the centre by taking the signature image into a fixed calculated frame and the unnecessary white spaces are removed without affecting the signature image such that the image is in the middle of the frame. For this first we divide the whole frame of the signature into 10*10 square row-wise and column-wise and find the variance(signature is considered to be binary and consists of only black and white pixels).If a square block has a zero variance we remove that square, otherwise restore. Thus squares of unnecessary white spaces are removed and then the image is restored in the fixed frame as shown in Figure 2.

Figure 2. Captured signature (A) before adjustment and (B) after adjustment

2.3 Feature points based on Vertical Splitting

Thirty feature points are obtained based on vertical splitting w.r.t. the central feature point. The procedure for finding vertical feature points are given below:
Algorithm:
Input: Static signature image after moving it to the centre of the fixed sized frame.
The steps are:
1) Split the image with a vertical line passing through the geometric centre (v0) which divides the image into two halves: Left part and Right part.
2) Find geometric centers v1 and v2 for left and right parts correspondingly.
3) Split the left and right part with horizontal lines through v1 and v2 to divide the two parts into four parts: Top-left, Bottom-left and Top-right, Bottom-right parts from which we obtain v3, v4 and v5, v6.
4) We again split each part of the image through their geometric centers to obtain feature points v7, v8, v9, ..., v13, v14.
5) Then we split each of the parts once again to obtain all the thirty vertical feature points (as shown in Figure 3).

2.4 Feature points based on Horizontal Splitting

Thirty feature points are obtained based on horizontal splitting w.r.t. the central feature point. The procedure for finding horizontal feature points are given below:

Algorithm:
Input: Static signature image after moving it to the centre of the fixed sized frame.
Output: Horizontal feature points: h1, h2, h3, h4, ..., h29, h30.
The steps are:
1) Split the image with a horizontal line passing through the geometric centre(h0) which divides the image into two halves: Top part and Bottom part.
2) Find geometric centers h1 and h2 for top and bottom parts correspondingly.
3) Split the top and bottom part with vertical lines through h1 and h2 to divide the two parts into four parts: Left-top, Right-top and Left-bottom, Right-bottom parts from which we obtain h3, h4 and h5,h6.
4) We again split each part of the image through their geometric centers to obtain feature points h7, h8, h9, ..., h13, h14.
5) Then we split each of the parts once again to obtain all the thirty vertical feature points (as shown in Figure 3).

3. Classification

In this paper features are based on geometric properties. So we use Euclidean distance model for classification. This is the simple distance between a pair of vectors of size n. Here vectors are nothing but feature points, so the size of vector is 2. How to calculate distance using Euclidean distance model is described in the following Section. In threshold calculation these distances are useful.

3.1 Euclidean distance model

Let $A(a_1,a_2,....,a_n)$ and $B(b_1,b_2,....,b_n)$ are two vectors of size n. We can calculate distance(d) by using equation 1.

$$\text{distance}(d) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \quad (1)$$

In our application, vectors are feature points on plane. So $d$ is the simple distance between two points.

4. Threshold

We have calculated individual thresholds for vertical splitting and horizontal splitting. Here we proposed one method for threshold selection. Fig. 5 shows the variations in single corresponding feature points of training signatures. Let $n$ is the number of training signatures and $x_1, x_2, ..., x_n$ are corresponding single feature points of training signatures (taking one corresponding feature point from each signature). $x_{\text{median}}$ is the median of n features from n signatures.

Let $d_1, d_2, ..., d_n$ are distances defined here,

$$d_1 = \text{distance}(x_{\text{median}}, x_1)$$
$$d_2 = \text{distance}(x_{\text{median}}, x_2)$$
$$\ldots$$
$$d_n = \text{distance}(x_{\text{median}}, x_n) \quad (2)$$

Two main parameters we used in threshold calculation are $d_{\text{avg}}$ and $\sigma$. Equations 3 and 4 shows the calculation of these two parameters.

$$d_{\text{avg}} = \text{average}(d_1, d_2, ..., d_n) \quad (3)$$
$$\sigma = SD(d_1, d_2, ..., d_n) \quad (4)$$
Like this total thirty different feature points are there for both vertical and horizontal splitting based on average distance \(d_{avg}\) and standard deviation \(\sigma\). Equation 5 shows the main formula for threshold.

\[
\text{threshold}(i) = \sqrt{\sum_{t=1}^{30} (d_{avg,t} + \sigma_t)^2}
\]

(5)

5. Experiments and Results

For experiment we took 24 original signatures from each person and selected 8 for training. These original signatures are taken in different days. Forgeries taken by three persons and 8 from each. Total 16 originals and 24 forgeries for each person signature are going to be tested. These original signatures are taken in different days. Forgeries taken by each person and selected 8 for training. These original signatures are taking for training from each person. There are 60 feature points from each original signature, 30 are taken by vertical splitting (Section 2.3) and 30 are taken by horizontal splitting (Section 2.4). Individual thresholds and patterns are calculated for vertical splitting and horizontal splitting. Pattern points based on vertical splitting are shown below:

\[
\begin{align*}
V_{pattern,1} &= \text{median}(v_1,1; v_2,1; \ldots; v_n,1) \\
V_{pattern,2} &= \text{median}(v_1,2; v_2,2; \ldots; v_n,2) \\
V_{pattern,3} &= \text{median}(v_1,3; v_2,3; \ldots; v_n,3) \\
V_{pattern,4} &= \text{median}(v_1,4; v_2,4; \ldots; v_n,4) \\
& \vdots \\
V_{pattern,29} &= \text{median}(v_1,29; v_2,29; \ldots; v_n,29) \\
V_{pattern,30} &= \text{median}(v_1,30; v_2,30; \ldots; v_n,30)
\end{align*}
\]

(6)

where \(v_i, v_2, \ldots; v_{30}\) are vertical splitting features of \(i^{th}\) training signature sample. Threshold based on vertical splitting is shown below. Now we have to calculate the \(V_d\) which is the distance of the first feature point off all the training signatures from the geometric centers.

\[
V_{d_{n}} = \text{Distance}(x_{median}, x_{n})
\]

So,

\[
\begin{align*}
V_{d1,1} &= \text{Distance}(V_{pattern,1}, V_{1,1}) \\
V_{d1,2} &= \text{Distance}(V_{pattern,1}, V_{1,2}) \\
V_{d1,3} &= \text{Distance}(V_{pattern,1}, V_{1,3}) \\
& \vdots \\
V_{d1,n} &= \text{Distance}(V_{pattern,1}, V_{1,n})
\end{align*}
\]

Therefore,

\[
\text{davg1} = \text{Average}(V_{d1,1}; V_{d1,2}; V_{d1,3}; \ldots; V_{d1,n})
\]

where \(\text{davg1}\) is the average of \(V_d\)'s for the first feature point of \(n\) signatures. Similarly we calculate \(\text{davg2}, \text{davg3}, \ldots, \text{davg29}, \text{davg30}\) for the 2nd, 3rd, \ldots, 29th, and 30th feature points.

Now we know that, Variance,

\[
\sigma = SD(d_1, d_2, \ldots, d_n)
\]

Therefore,

\[
\begin{align*}
\sigma_1 &= SD(V_{d1,1}; V_{d1,2}; \ldots; V_{d1,n}) \\
\sigma_2 &= SD(V_{d2,1}; V_{d2,2}; \ldots; V_{d2,n}) \\
& \vdots \\
\sigma_{30} &= SD(V_{d30,1}; V_{d30,2}; \ldots; V_{d30,n})
\end{align*}
\]

Hence,

\[
\text{vth_{threshold}} = \sqrt{\sum_{t=1}^{30} \left(d_{avg,t} + \sigma_t\right)^2}
\]

(7)

Now we apply the same procedure to calculate the \(h_{\text{threshold}}\) using the horizontal feature points.

\[
\begin{align*}
V_{h_{pattern,1}} &= \text{median}(h_1,1; h_2,1; \ldots; h_n,1) \\
V_{h_{pattern,2}} &= \text{median}(h_1,2; h_2,2; \ldots; h_n,2) \\
V_{h_{pattern,3}} &= \text{median}(h_1,3; h_2,3; \ldots; h_n,3) \\
V_{h_{pattern,4}} &= \text{median}(h_1,4; h_2,4; \ldots; h_n,4) \\
& \vdots \\
V_{h_{pattern,29}} &= \text{median}(h_1,29; h_2,29; \ldots; h_n,29) \\
V_{h_{pattern,30}} &= \text{median}(h_1,30; h_2,30; \ldots; h_n,30)
\end{align*}
\]

(8)

Where \(h_i, h_2, \ldots; h_{30}\) are horizontal splitting features of \(i^{th}\) training signature sample. Threshold based on horizontal splitting is shown below:

\[
\text{h_{threshold}} = \sqrt{\sum_{t=1}^{30} \left(h_{d_{avg,t}} + \sigma_{hi}\right)^2}
\]

(9)

We will store pattern points and thresholds of both horizontal splitting and vertical splitting. These values are useful in testing.

5.2. Testing

When new signature comes for testing we have to calculate features of vertical splitting and horizontal splitting. Feature points based vertical splitting are \(v_{new};1, v_{new};2, v_{new};3, v_{new};4, \ldots, v_{new};29, v_{new};30\). Distances between new signature features and pattern feature points based on vertical splitting are shown below:

\[
\begin{align*}
v_{d_{new};1} &= \text{distance}(v_{pattern};1, v_{new};1) \\
v_{d_{new};2} &= \text{distance}(v_{pattern};2, v_{new};2) \\
v_{d_{new};3} &= \text{distance}(v_{pattern};3, v_{new};3)
\end{align*}
\]
vdnew;4 = distance(vpattern;4;vnew;4) (10)

......

vdnew;29 = distance(vpattern;29;vnew;29)

vdnew;30 = distance(vpattern;30;vnew;30)

For classification of new signature we have to calculate vdistance and compare this with vthreshold. If vdistance is less than or equal to vthreshold then new signature is acceptable by vertical splitting.

vdistance = \sqrt[29]\sum_{i=1}^{29} v_{i}^2 \quad (11)

Feature points based vertical splitting are hnew;1, hnew;2, hnew;3, hnew;4, ...... hnew;29; hnew;30. Distances between new signature features and pattern feature points based on horizontal splitting are shown below:

hdnew;1 = distance(hpattern;1;hnew;1)

hdnew;2 = distance(hpattern;2;hnew;2)

hdnew;3 = distance(hpattern;3;hnew;3)

hdnew;4 = distance(hpattern;4;hnew;4)

......

hdnew;29 = distance(hpattern;29;hnew;29)

hdnew;30 = distance(hpattern;30;hnew;30)

(12)

For classification of new signature we have to calculate hdistance and compare this with hthreshold. If hdistance is less than or equal to hthreshold then new signature is acceptable by horizontal splitting.

hdistance = \sqrt[30]\sum_{i=1}^{30} h_{i}^2 \quad (13)

New signature features have to satisfy both vertical splitting and horizontal splitting thresholds.

5.3. Results

False Acceptance Rate (FAR) and False Rejection Rate (FRR) are the two parameters using for measuring performance of any signature verification method. FAR is calculated by equation 14 and FRR is calculated by equation 15.

\[ \text{FAR} = \frac{\text{number of forgeries accepted}}{\text{number of forgeries tested}} \times 100 \quad (14) \]

\[ \text{FRR} = \frac{\text{number of originals rejected}}{\text{number of originals tested}} \times 100 \quad (15) \]

<table>
<thead>
<tr>
<th>Forgery Type</th>
<th>Existing Scheme</th>
<th>Existing Scheme(12 feature point)</th>
<th>Proposed Scheme(60 feature point)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RandoM</td>
<td>5.61</td>
<td>2.08</td>
<td>0.43</td>
</tr>
<tr>
<td>Simple</td>
<td>16.39</td>
<td>9.75</td>
<td>0.98</td>
</tr>
<tr>
<td>Skill</td>
<td>19.3</td>
<td>16.36</td>
<td>2.08</td>
</tr>
</tbody>
</table>

Table 2: Comparative analysis of FRR

6. Conclusion

The Algorithm which is based on the 60 feature points is more efficient and gives more accurate results than the existing Techniques and survives against the skilled forgeries. The algorithm results the FAR which is very much less as compared to the FARs of the previously existing techniques. We compared our algorithms with other techniques based on feature extraction(12 feature points) and techniques based on Polar and Cartesian coordinates. But as our algorithm takes 60 feature points for threshold calculations, a small variation of a signature results in a large change in the values of threshold distance from the geometric center. Therefore in our algorithm the FRR value is increased. So it is important for a user to sign his signature with utmost care so that there is not a large variation of his signature to his training signatures. Otherwise there is a probability of rejection of an original signature. Moreover, since we have extracted 30 feature points by vertical splitting and 30 by horizontal splitting for the calculation of the threshold value, the time complexity is higher than the time complexity of the existing technique which uses 12 feature points for threshold calculations.

10. References


