Off-Line Verification for Chinese Signatures

HWEI-JEN LIN, FU-WEN YANG and SHWU-HUEY YEN

Department of Information Engineering, Tamkang University, Tamsui, Taipei Hsien, Taiwan, R.O.C.
E-mail: hjlin@cs.tku.edu.tw

Signature verification is one of the important methodologies for personnel identification. Because of its wide applications to security practices quite a number of signature verification techniques associated with various features have been proposed. The features utilized in a signature verification system directly affect the performance of the system. In this paper, we review a variety of recently proposed features of signature and present two new features. Each of these features is examined to evaluate its performance for Chinese signature verification. Then we choose and combine the features that establish good verification rates. As a result, we propose an off-line Chinese signature verification system based on this combination of features. The experimental results show that the proposed system achieves a good performance in term of verification rate.

Keywords: signature verification, slant, moment, projection moment, threshold, weighted Euclidean distance.

1. Introduction

Usually, one’s own signature is a mature form after experience over a quite long period of time, which involves some significant characteristics. Because of the effectiveness of these characteristics and the ease of process, signature verification is widely applied to authorize personal identification in a variety of organizations. With the advances in computer technology and the high speed required in modern business transactions, the need for automatic signature verification is increasing.

The objective of this paper is to propose an efficient and reliable off-line verification system for Chinese signatures. Verification for Chinese signatures is much more difficult than other languages due to the complicated structure of Chinese characters, which is the main factor that increases the variations in writing. To cope with this difficulty, we focus on finding effective features from Chinese signatures that are not much affected by these variations, for the proposed system.

In order to find effective features in Chinese signatures, a variety of recently proposed signature features for any language were surveyed and some new features for our own were presented. Secondly, we examined the effectiveness of each of these features extracted from Chinese signatures. Finally we selected the features that establish higher verification rates. As a result, we propose an off-line Chinese signature verification system on the basis of a combination of selected features.

The main job of an automatic signature verification system is to distinguish between genuine signatures and forged
signatures. The forged signatures can be coarsely divided into two classes, freehand forgeries and simulated forgeries. The latter class is much more difficult to verify than the former. For the purpose of efficiency, the proposed system processes in two stages, where the first stage discriminates simple freehand forgeries, and the second stage processes the more complicated case.

The most common measure of the performance of a signature verification system is a pair of statistics known as false rejection rate (FRR) and false acceptance rate (FAR). To set the desired balance between FRR and FAR, many systems have variable thresholds. Usually, zero percent FAR is required for the sake of security.

This paper is organized as follows: Section 2 tells how to capture signature data. In section 3, we describe the preprocessing steps utilized in our work. Features and dissimilarity measure are introduced in section 4. Features are examined in section 5. Some experimental results and discussions are given in section 6, with conclusion in section 7.

2. Signature Data

Our signature data consists of 400 genuine signatures from 20 persons and 320 forgeries from 20 persons. To acquire the genuine signatures, every person signed his own signature 10 times using a ballpoint pen on two different occasions. Five genuine signatures were randomly selected at each occasion. The remaining signatures were used as testing samples. Signatures were subjected to three kinds of forging, which are freehand, simulated, and traced [13] signatures. Signatures forged by a professional forger and traced signatures are very difficult to detect. Therefore, in general, the off-line signature verification system is designed to detect freehand signatures or simulated signatures by amateurs. To obtain simulated forgeries, forgers were allowed to practice ahead of time. Among 16 forgeries (8 each for the freehand and simulated) from each forger, 5 freehand and 5 most-similar simulated forgeries were chosen to make up the 200 forged test samples. All samples were written in a limited space and horizontally oriented. Each sample was scanned with a scanner and expressed as a 256 gray-level image with a resolution of 300×120 pixels. Figure 1(a) shows an example of these signatures.

3. Preprocessing Steps

The features to be examined in this paper are extracted from some preprocessed versions of the original images. In this section, we shall describe the preprocessing steps required for our work.

- Background reduction: This can be done by using the filter defined in Eq. (1).
\( p''(i,j) = \begin{cases} p'(i,j) & \text{if } p'(i,j) > 0 \\ 0 & \text{otherwise} \end{cases} \) (1)

where \( p'(i,j) = p(i,j) - \frac{1}{M} \sum_{l=1}^{M} p(l, j) \),

\( p(i,j) \) is the gray value of the original image with \( 1 \leq i \leq M, 1 \leq j \leq N \), and \( M \times N \) is the size of the image.

- Noise reduction: This is simply performed by a lowpass filter.
- Binarization: We use Otsu’s method [15] to binarize a gray-leveled signature image.
- Segmentation: Chinese signatures usually contain three Chinese characters with little or no connection between every two adjacent Chinese characters. This nature often causes significant variation in writing as a whole. In order to reduce such instabilities, we segment the signature to separate each character, so that the local features from each character can be extracted. This can be done by simply scanning the image vertically and horizontally to find out the boundary lines of each character.
- Thinning: Zhang-Suen/Stentiford/Holt’s method [16] is utilized to obtain the skeleton of the binarized image.
- Contouring: We obtain the contour of the binarized image by simply collecting all black pixels adjacent to a white pixel.
- Filling outline: This can be done by scanning the binarized image from 4 different directions, each scan line stop while reaching a black pixel. After scanning, all unscanned pixels are collected, and the result forms a filled outline of the image.
- Original density restoration: This can be obtained by simply restoring each of the black pixels in the binarized image into its original gray level [1].

Figure 1 shows the results of these preprocessing steps.

4. Features Extraction

In this section, we survey a variety of recently proposed signature features and propose some new good features. We also examine each of these features to evaluate its performance in Chinese signature verification.

4.1. Recently Proposed Signature Features

There are three main categories for the recently proposed signature features: geometric, moment-based, and pressure-like features, which are described as follows.

- Geometric features:
  (1) The width and height of the binarized image of each segmented character, which form a vector denoted by WH.
(2) Normalized width and height, which are the width and height evaluated from (1) divided by the maximal value of widths and heights, the resulting vector is denoted by NWH.

(3) Center of gravity: Centers of gravity of the binary and filled outline images are denoted by BG and OG, respectively. These features are normalized versions; that is, for each segmented Chinese character in the image, the x- and y-coordinates of the centers are divided by the width and height, respectively.

(4) Centroid, which is the mean of the X-Y coordinates of the black pixels in the reference image. These features are extracted from the thinned and contour images, and denoted by TC and CC, respectively. These features are also normalized versions.

(5) Baseline proposed by Ammar [1]: This feature is extracted from the binarized image and denoted by BL.

(6) Normalized baseline: To normalize the baseline, it is divided by the height of the character and denoted by NBL.

(7) Slant Features proposed by Ammar [1]: Slant features are extracted from the thinned image, which is defined as follows: For a black pixel in the thinned image, its black neighbors in four directions, negatively, vertically, positively, and horizontally slant pixels, are denoted by HSP, VSP, NSP and PSP, respectively. These slant features are measured on each signature character as a whole and then normalized with respect to the area (or the total number of black pixels in the thinned image). The vector formed by these features is denoted by S.

(8) Slant Features of Contour proposed by Dimauro [6]: This feature is extracted from the contour image and denoted by CS.

- Moment-Based Features:
  (1) Moments proposed by Hu [8]: These features are extracted from the thinned, binarized, contour, and filled outline images, and denoted by TM, BM, CM, and OM, respectively.
  (2) Projection Moments proposed by Madhvapathy [12]: These feature are also extracted from the thinned, binarized, contour, and filled outline images, and denoted by TPM, BPM, CPM, and OPM, respectively.

- Pressure-Like Features:
  (1) High Density Feature proposed by Ammar [1]: This feature is denoted by HDF.

The features described in this section are tested based on the same signature samples, and the results established by FRR and FAR are shown in Table 1.
Table 1. FRR and FAR based on different features.

<table>
<thead>
<tr>
<th>Feature</th>
<th>FRR</th>
<th>FAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>WH</td>
<td>10.5%</td>
<td>32.5%</td>
</tr>
<tr>
<td>NWH</td>
<td>6.5%</td>
<td>43.5%</td>
</tr>
<tr>
<td>BG</td>
<td>19.5%</td>
<td>24.5%</td>
</tr>
<tr>
<td>OG</td>
<td>19.5%</td>
<td>24.5%</td>
</tr>
<tr>
<td>TC</td>
<td>20%</td>
<td>26%</td>
</tr>
<tr>
<td>CC</td>
<td>1.5%</td>
<td>67.5%</td>
</tr>
<tr>
<td>BL</td>
<td>14%</td>
<td>55.5%</td>
</tr>
<tr>
<td>NBL</td>
<td>7%</td>
<td>61%</td>
</tr>
<tr>
<td>S</td>
<td>7.5%</td>
<td>10%</td>
</tr>
<tr>
<td>CS</td>
<td>8.5%</td>
<td>19.5%</td>
</tr>
<tr>
<td>TM</td>
<td>17%</td>
<td>24%</td>
</tr>
<tr>
<td>CM</td>
<td>19.5%</td>
<td>30.5%</td>
</tr>
<tr>
<td>BM</td>
<td>24%</td>
<td>26%</td>
</tr>
<tr>
<td>OM</td>
<td>19%</td>
<td>32%</td>
</tr>
<tr>
<td>TPM</td>
<td>20%</td>
<td>16.5%</td>
</tr>
<tr>
<td>CPM</td>
<td>22%</td>
<td>10%</td>
</tr>
<tr>
<td>BPM</td>
<td>17%</td>
<td>11.5%</td>
</tr>
<tr>
<td>OPM</td>
<td>21.5%</td>
<td>10%</td>
</tr>
<tr>
<td>HDF</td>
<td>11.5%</td>
<td>50.5%</td>
</tr>
</tbody>
</table>

4.2. The Proposed Signature Features

We present two features in this paper, both belonging to the category of geometric features, as follows:

1) Normalized slant features: The slant features described in the category of geometry is modified to become more stable features under different writing conditions as follows:

\[ NHSP = \frac{HSP}{\text{width}}, \quad NVSP = \frac{VSP}{\text{height}}, \quad NNSP = \frac{NSP}{\text{Min}(\text{width}, \text{height})}, \quad NPSP = \frac{PSP}{\text{Min}(\text{width}, \text{height})}. \]

The vector formed by the resulting features is denoted by NS.

2) Pixels/area ratio: The ratio of the number of pixels in the character to the area of the rectangle where the character is inscribed. These features extracted from the binarized, contour, and filled outline images are denoted by BAR, CAR, and OAR, respectively.

The proposed features are tested based on the same signature samples, and the results established by FRR and FAR are shown in Table 2.
Table 2. FRR and FAR based on Proposed features.

<table>
<thead>
<tr>
<th>Feature</th>
<th>FRR</th>
<th>FAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS</td>
<td>4%</td>
<td>12.5%</td>
</tr>
<tr>
<td>BAR</td>
<td>9.5%</td>
<td>49%</td>
</tr>
<tr>
<td>CAR</td>
<td>7%</td>
<td>41.5%</td>
</tr>
<tr>
<td>OAR</td>
<td>9.5%</td>
<td>48%</td>
</tr>
</tbody>
</table>

5. Dissimilarity Measure and A Near-Optimal Combination of Features

5.1. Dissimilarity Measure

The weighted Euclidean distance proposed by Ammar[1] is utilized to measure the dissimilarity of two signature images, which is defined as follows:

\[
D = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( \frac{F_i - u_i}{\sigma_i} \right)^2}
\]

where \(F_i\) is the \(i\)th feature of the sample to be verified.

\(u_i\) is the mean of the \(i\)th feature of the reference genuines.

\(\sigma_i\) is the standard deviation of the \(i\)th feature of the reference genuines.

\(n\) is the number of features.

\[
u_i = \frac{\sum_{k=1}^{m} R_{ki}}{m}
\]

where \(u_i\) is the mean of \(i\)th feature of the reference genuines.

\(R_{ki}\) is the \(i\)th feature of the \(k\)th reference genuine.

\(m\) is the number of the reference genuines.

For comparison of two signatures, we need to determine a threshold \(T\) on the dissimilarity measure. Here we set the value of \(T\) to be the maximal distance of all pairs of reference genuines multiplied by a parameter \(C\) (1≤\(C\)≤2). The value of \(C\) is determined from our experiment. The details are described in the next subsection.

5.2. Feature Selection and Combination

According to the results of feature examination in the previous section, we select the features that established lower error rates for classification. All possible combinations of these candidate features associated with operations “AND” (denoted by
“&”) and “OR” (denoted by “|”) are tested. The best combination, with FRR/FAR crossover of 3.5%, turns out to be

\[(S \& CC \& NWH \& BL \& CAR) | (CS \& CC \& NWH \& CAR \& BPM) | (NS \& CC \& NWH \& OAR \& TM) | (NS \& CC \& NBL \& CAR \& HDF)\]

The further experimental results show that the performance of the feature CS is not as good as NS, OAR is not better than CAR, and HDF, NBL and BL show poor performances. Thus, we get rid of these features, CS, OAR, HDF, NBL and BL, to produce a fair combination with a smaller set of features and thus yielding fast computation, the resulting simplified combination is \((NS | S) \& (NWH | CAR) \& CC \& (TM | BPM)\); its FRR/FAR crossover is 4.5%.

6. Experimental Results and Discussion

In this paper, we proposed an off-line Chinese signature verification system using a combination of features which show good performance. Figure 2 shows the flow chart for this system. 200 genuine signatures are taken for training and 400 signatures (200 genuine and 200 forgeries) for testing. Some of these samples are shown in Figures 3 and 4. In the previous section, we discussed how to determine an appropriate threshold value on dissimilarity measure for comparison. That is, the greater the threshold, the greater decrease in the Type I Error Rate and the greater increase in the Type II Error Rate. Table 3 shows the verification results with respect to the values of C between 1 and 2. Based on the combination of features described in section 5. It can be seen that C=1.1 is the best choice, in which case the sum of Type I and Type II error rates is minimal.

<table>
<thead>
<tr>
<th>C</th>
<th>1.0</th>
<th>1.1</th>
<th>1.2</th>
<th>1.3</th>
<th>1.4</th>
<th>1.5</th>
<th>1.6</th>
<th>1.7</th>
<th>1.8</th>
<th>1.9</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRR</td>
<td>8.5%</td>
<td>4%</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
<td>2%</td>
<td>0.5%</td>
<td>0.5%</td>
<td>0.5%</td>
<td>0.5%</td>
</tr>
<tr>
<td>FAR</td>
<td>1.5%</td>
<td>5%</td>
<td>8.5%</td>
<td>10%</td>
<td>12.5%</td>
<td>14.5%</td>
<td>17.5%</td>
<td>22%</td>
<td>24.5%</td>
<td>28.5%</td>
<td>30.5%</td>
</tr>
</tbody>
</table>

7. Conclusion

In this research, some new features of signatures are presented and a variety of recently proposed features are reviewed and analyzed. Features showing good performance are selected, and finally a near-optimal combination of such features in terms of time efficiency and high verification rate is derived. The weighted Euclidean distance is utilized for dissimilarity measure. The experiments demonstrated satisfactory results for the progress of off-line Chinese signature verification. Further improvement in our system, like successfully discriminating between skilled forgeries and genuines, and searching for features that are independent of sorts of pens is being investigated.
8. References


Figure 1. (a) original image (b) background-reduced image (c) noise-reduced image (d) binarized image (e) segmented image (f) thinned image (g) contour (h) filled outline (i) original density restored image
input the reference signature

Preprocessing

Feature Extraction

Create the database of the reference signatures

Similarity Measure

Create the database of prototype features and thresholds

The database of the prototype features and thresholds of the reference signatures

Signature Verification

Verification Result

Figure 2. The off-line Chinese signature verification system

(a) Create the database of the prototype features and thresholds for the reference signatures.

(b) Signature Verification.
Figure 3. Samples of genuines

Figure 4. Samples of forgeries