Forensic Science International 255 (2015) 123–132

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

Extraction and analysis of the width, gray scale and radian in Chinese signature handwriting

Xiaohong Chen *

Institute of Forensic Science, Ministry of Justice, 1347, West Guangfu Road, Shanghai 200063, PR China

ARTICLE INFO

Article history:
Available online 13 July 2015

Keywords:
Stroke measurement
Identification
Pictograms
Statistical analysis
Dynamic time warping

ABSTRACT

Forensic handwriting examination is a relevant identification process in forensic science. This research obtained ideas from the process of features detection and analysis in forensic handwriting examination. A Chinese signature database was developed and comprised original signatures, freehand imitation forgeries, random forgeries and tracing imitation forgeries. The features of width, gray scale and radian combined with stroke orders were automatically extracted after image processing. A correlation coefficient was used to precisely characterize and express the similarities between signatures. To validate the differences between writers, a multivariate analysis of the variance was employed. The canonical discriminant analysis was performed between the original and non-original signatures; the cross-validation estimated the discriminating power of the width, gray scale and radian data. It is suggested that the extraction and analysis of these properties in Chinese signatures is reasonable. Meanwhile, forensic handwriting examination using the quantitative feature extraction and statistical analysis methods in this research could be performed with a satisfactory result in the discriminant analysis.

1. Introduction

Forensic handwriting examination is a relevant identification process in forensic science, and it is one important part of the tasks performed by forensic document examiners [1,2]. Forensic handwriting examination appears as an empirical discipline. Debates regarding the scientific basis of forensic handwriting examination have arisen in recent years [3,4]. The fundamental law considering the variability and individuality of handwriting were challenged for the lack of objectivity. Various studies have shown that trained forensic document examiners perform significantly better than lay people in the analysis, comparison and identification processes based on handwriting evidence [5]. Much research on the applications of quantitative extraction methods for handwriting features [6–10] and on the statistics methods for forensic handwriting examination assessment [11–16] has been conducted to reduce the subjective aspect in forensic handwriting examination.

Various features are extracted and applied for Chinese handwriting verification [17]. However, this research lacks the validation of the variability within-writer and the individuality between-writer based on the features that are used in comparison. Additionally, the differences between forensic handwriting examination and computer handwriting verification lead to verification methods that are not practical for forensic application. However, the quantitative feature extraction, statistical analysis and overall assessment are the trend of current forensic handwriting examination [18–20].

This study based on the process of feature detection and analysis in forensic handwriting examination. In addition to spatial features, forensic document examiners should recover the stroke orders and grasp the stroke contours, width and gray scale for comparison. The aim of this paper was to develop new quantitative and objective feature extraction and statistical analysis methods for Chinese signature handwriting examination.

A Chinese signature database was comprised of 981 signatures of 12 groups produced by 12 volunteers; each group consisted of original signatures, freehand imitation forgeries, random forgeries and tracing imitation forgeries. A threshold was applied in image binarization after the signatures were imported into the computer. The skeleton and the signature edges were extracted by image processing. Then, a program for stroke order recovery processed the skeleton of the signatures. The width, gray scale and radian values were automatically extracted in the stroke order.

* Corresponding author.
E-mail address: chenxh@ssfjd.cn

http://dx.doi.org/10.1016/j.forsciint.2015.07.008
0379-0738/

A dynamic time warping (DTW) method was applied to cope with the different writing speeds. Corresponding strokes were aligned for comparative analysis after DTW. In the correlation analysis, the correlation coefficient was used to precisely characterize and express the similarities between signatures.

In the statistical analysis, the mean and standard deviation of the width, gray scale and radian obtained in the correlation analysis were calculated as the data description. A multivariate analysis of the variance was employed to validate the differences between writers based on width, gray scale and radian. A discriminant analysis helped to distinguish different writers with the canonical discriminant functions.

2. Materials and data

2.1. Sampling

Chinese signatures were written by means of a ballpoint pen with black ink, on A4 paper (for original signatures, random forgeries and freehand imitation forgeries) printed with 12 squares and 195 mm × 271 mm highly transparent paper (for tracing imitation forgeries), with the signer in a sitting pose. Twelve volunteers who could produce skilled imitation forgeries were organized. The Chinese signature database was composed of 981 signatures of 12 groups produced by 12 volunteers; every groups contained 20–24 original signatures, 30–36 freehand imitation forgeries, 10–12 random forgeries and 10–12 tracing imitation forgeries.

Original Signature (OR): Each volunteer wrote 20–24 original signatures using their normal writing style in two writing sessions. One original signature was chosen as the model at random and was duplicated into a blank A4 paper for forgery production.

Freehand imitation forgery signature (FF): Three volunteers other than the writer of the original signature were chosen at random. After practicing for 2–4 days until the forgeries were produced reasonably well, each volunteer wrote 10–12 forgeries while attempting to imitate the model signature (FF1, FF2, and FF3, respectively).

Random forgery signature (RF): With the knowledge of only the name of the original writer, one volunteer other than writer of the original signature was chosen at random; this volunteer produced 10–12 signatures based on their own writing habits.

Tracing imitation forgery Signature (TF): One volunteer other than the writer of the original signature was chosen at random. The volunteer was asked to place a piece of highly transparency paper on top of the original signature model and carefully trace the model signature, attempting to imitate the model signature.

2.2. Image processing procedure and features extraction

The Chinese signatures were imported into the computer by means of an EPSON PERFECTION V700 PHOTO scanner with a resolution of 400 dpi. Matlab 7.0 software was applied for to extract the features. Stroke order or sequence is an important feature in handwriting. A program of stroke order tracing is referred to in the research of Oshiharu Kato and Makoto Yasuhara [21].

Firstly, a threshold was applied in the image binarization. Then, the skeletons and edges of the signatures were extracted by separately skeletonizing and edging the images (Fig. 1). The stroke order tracing was automatically used in the signature skeletons after a begin point was manually provided (Fig. 2). If any error occurred, it was corrected manually. The stroke order or the sequence of signatures (S) were assigned with x coordinates (X) and y coordinates (Y). Finally, the width (W), gray scale (G) and radian (R) values were automatically extracted in the stroke order (Fig. 3). The gray scale values of the points in the skeletons were used as the gray scale data. The width and radian data were calculated as the following functions, as shown in Fig. 3:

$$S = \sum_{i=1}^{n} (X_i, Y_i); \tag{1}$$

$$W = \sum_{i=1}^{n} (L_i \cdot \sin\theta_i); \tag{2}$$

$$R = \sum_{i=1}^{n} \theta_i \left(\frac{\pi}{180}\right); \tag{3}$$

$$F = \sum_{i=1}^{n} W_i, G_i, R_i; \tag{4}$$

where stroke order (S); x coordinate (X); y coordinate (Y); width (W), gray scale (G), radian (R); tangent line (T), n = length (X) or length (Y); i = 1,2,3. . .n.

2.3. Feature data preparation

2.3.1. Dynamic time warping application

Dynamic time warping is a well-known technique to find an optimal alignment between two time-dependent sequences under

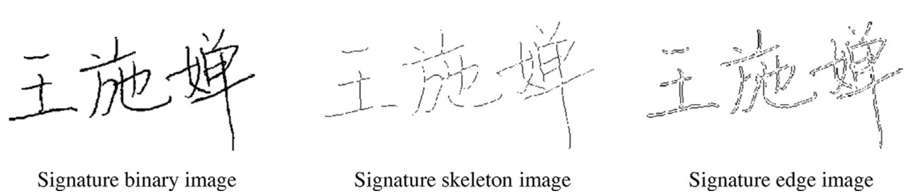

Fig. 1. Image pressing.

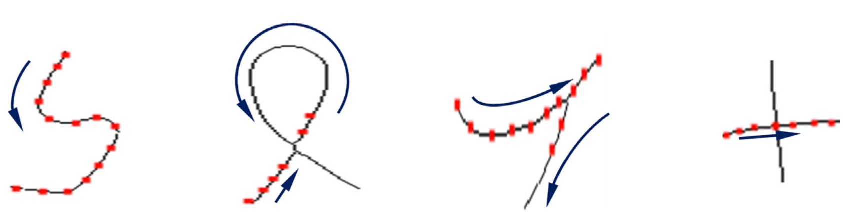

Fig. 2. Basic stroke order recovery processing image.