

OFF-LINE UYGHUR SIGNATURE RECOGNITION BASED ON MODIFIED GRID INFORMATION FEATURES

Kurban Ubul^{1,3}, Andy Adler², Gulirana Abliz¹, Maimaitijiang Yasheng², Askar Hamdulla^{1,3}

¹School of Information Science and Engineering, Xinjiang University, 830046, Urumqi, China

²Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada

³Xinjiang laboratory of Multi-language information technology, Xinjiang University, 830046, Urumqi, Xinjiang, China

ABSTRACT

Many techniques have been published on handwriting signature recognition, but none of these techniques presented are about Uyghur handwritten signature due to its complex nature. In this paper, we propose methods for off-line signature recognition for Uyghur handwriting first time. The signature images were pre-processed based on the nature of Uyghur signature. The preprocessing included noise reduction, binarization and normalization. Then multi-dimensional modified grid information features were extracted according to the character of Uyghur signature and its writing style. Finally, three kinds of classification techniques were used: Euclidean distance (ED) classifier, K nearest neighbor (K-NN) classifier and Bayes classifier. Experiments were performed using Uyghur signature samples from 50 different people with 1000 signatures. A promising result of 93.53% average correct recognition rate was achieved.

1. INTRODUCTION

Handwritten signatures are widely accepted, because signature has long been established for personal identification and verification in daily life, including banking transactions, and other commerce applications

[1]. The aim of the recognition process is to identify the writer of a given sample that it is a 1:N classification problem; the purpose of the verification process is to confirm or reject the sample that it is belong to 1:1 classification problem.

Signature recognition can be divided in two classes: on-line and off-line signature recognition. On-line (or dynamic) systems use a digitizer or an instrumented pen to generate signals; while off-line (or static) systems produce an image of a signature with the help of a scanner or camera. One key difference with on-line signing action is that it allows capture of the handwriting trajectory. The features of handwritten signature, such as the writing style, writing order, writing-speed variation, and skillfulness, need to be recovered from the grey-level pixels. On-line signature recognition system has been achieved higher accuracy rates than off-line recognition with the time-related information and dynamic features like speed, pressure, acceleration stroke order etc. However, off-line signature recognition is more difficult than on-line signature recognition due to the lack of time-related and dynamic information.

Scientific reports about signature recognition and verification are mostly based on Latin (English) handwriting. Surveys for non-English and non-Latin (Chinese, Japanese, Arabic and Persian) signature verification systems are appear in [2]. Radmehr et al. [3] presented Radon Transform and Fractal Dimension (FD) based offline signature recognition system. Xiao et al [4] proposed a way of off-line Chinese handwritten

recognition based on wavelet packs and Gauss model in order to solve the problems of complicated data processing, difficult segmentation and deficient feature extraction. Neural network based signature recognition and verification methods indicated in [5], and grid information feature based approaches are presented in [6]. An off-line Arabic signature recognition based on invariant moment's properties is proposed in [7]. These reports about signature recognitions are mostly based on Latin handwriting [1, 3], and Chinese handwriting [4, 5], Arabic handwriting [7] even Persian handwriting [2]. However, none of methods and algorithms for automatic signature recognition that have been reported in the literature is based on Uyghur handwriting. It is challenging task to find and develop methods suitable for Uyghur signature recognition and verification. So there is still much research space for implementing existing algorithms or developing new effective algorithms and methods to fall in with the nature of Uyghur handwritten signature.

The work reported here is an off-line signature recognition approach using modified grid information features for Uyghur handwritten signature. Signature images were pre-processed based on the character of Uyghur signature in our database. Since grid information feature was used effectively and extensively for signature recognition [6], the feature is extracted having modified the method of grid information feature. Classification experiments were conducted to testing the feature's efficiency. It has been also tested on Latin (English) signature dataset. Forgery signatures were used to test the performance of the recognition system.

2. DATA ACQUISITION AND PRE-PROCESSING

The steps of signature recognition commonly include data acquisition, pre-processing, feature extraction and classification. Uyghur people are selected to give their natural signatures on paper for data acquisition. The signatures are digitized with a scanner. The signature image must be preprocessed to reduce noise and account for different sizes. The nature of Uyghur is described first of all, because, this paper take some methods suitable for the character of Uyghur signature in preprocessing and feature extraction steps.

2.1. The nature of Uyghur handwriting text [8]

The Uyghurs are one of the Turkic-speaking ethnic group inhabiting Central and Eastern Asia. Today, Uyghurs mainly live in the Xinjiang Uyghur Autonomous Region (hereafter: Xinjiang) in China. Arabic based Uyghur script (some letters similar to Persian) is an official writing system in Xinjiang, while Cyrillic-based Uyghur script is still used by Uyghurs in former Soviet Union Republics and Latin-based Uyghur script are also in use¹. The handwritten signature of Arabic-based Uyghur script (hereafter: Uyghur) used widely in Xinjiang area is studied in this paper.

Uyghur character is composed of 32 letters including 8 vowel letters and 24 consonant letters, besides 4 kinds of different forms for each character. Thus, 32 letters become more than 120 character styles.

(1) The Uyghur character is written from right to left direction, and line is progressed from left down right direction. There are 4 different writing forms for Uyghur letter: "initial form", "intermediate form", "final form", and "isolated form". Thus, they constitute 126 character styles from 32 letters.

(2) The vocabulary of Uyghur character is composed of one or several letters. According to rules of writing, these letters will form one or several letter passages by initial and suffix connections. For a block letter or handwriting, the letters are connected along a certain level, which is called base line.

(3) Unequal width of letters. This phenomenon happens not only on different letters, but also on the 4 different forms of certain letters.

(4) The word of Uyghur language is consists of syllables. It normally comprises the vowel and consonant combinations while vowel is the centre. There is a blank space between two vocabularies.

(5) The stroke of Uyghur character is not fixed. The numbers of strokes for the same Uyghur word are different from person to person. Especially, they are different with position, size, longitude, slant angle and structure.

¹ Uyghur alphabet. See http://en.wikipedia.org/wiki/Uyghur_alphabet

2.2. Data acquisition

The signatures were collected using black ink from 195 Uyghur people, on a white A4 sheet of a paper, with 21 signatures per page. Each paper was divided into 21 same sized boxes into which the person was asked to sign his/her signature. The area of each box is big enough to give enough space to use his/her natural signature, and to allow size deviation of signature. The signatures are scanned and inputted to the computer using Canon MP810 scanner with .bmp format and 300 dpi resolutions in 256 grey levels. A part of the scanned image is indicated as Figure 1.

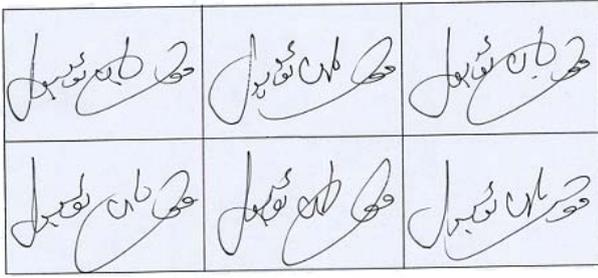


Figure 1. A part of the sample signatures

2.1. Pre-processing

The off-line signature recognition requires several preprocessing steps. These pre-processing steps prepare the signature images for the different type of processing required at the feature extraction phases. Pre-processing steps included: removing blank edges of the signature images (or data-area cropping), binarization (or thresholding), noise reduction, removing textured backgrounds, size normalization, thinning (or skeletonization), close contour tracing, blurring, and rotation of the signature images. For the Uyghur signature images, the preprocessing steps were noise reduction, binarization, and size normalization.

2.2.1. Noise reduction

Signature image noise presents as random pixels that are not part of the image particulars. Scanned signature images may have image noise caused by the sensors of the scanner. If these are not eliminated before the feature extraction and classification, the signature image may mistakenly be analyzed or interpreted. For each Uyghur

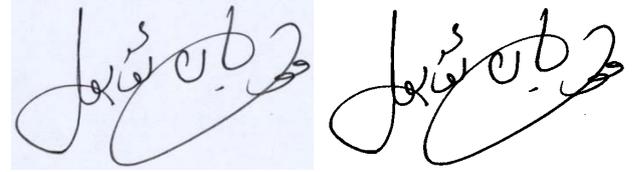
signature, a recursive algorithm is used to calculate the minimum target length which may be noise or incidental component of Uyghur signature in the image. An 8 pixel threshold is selected after statistical calculations for the incidental component of the Uyghur signature. Noise is considered discrete if the connected black pixels of observed points' is less than 10 and all the discrete noise are filled in white.

2.2.2. Binarization

A threshold filter is selected for binarization in this paper. This filter is used to convert image to black (0) and white (255) and to remove unnecessary information by using specific threshold:

$$b(i, j) = \begin{cases} 255 & \text{if } I(i, j) \geq T \\ 0 & \text{if } I(i, j) < T \end{cases} \quad (1)$$

where $I(i, j)$ is the gray level of the pixel and T is a specific threshold value. An original image and its binary images are showed as the following Figure 2.



(a) Original image

(b) Binarized image

Figure 2. Original image and binarization

2.2.3. Size normalization

As the sizes of signatures in our database are different, normalization of these images with respect to their height and width is necessary, which enforces a fixed height and width by stretching or cropping. The size of signature image is normalized to 384×96 based on the nature of Uyghur signature. The sizes of images are normalized based on distribution of black pixels in vertically and horizontally. The gravity of signature G_I, G_J is calculated first (Eq. 2), the divergence σ_I, σ_J is evaluated using Eq. 3, then the lattice σ_I^2, σ_J^2 which can linearly amplify or reduce the signature image based on the proportion [9].

$$\begin{cases} G_I = \frac{\sum_{i=A}^B \sum_{j=L}^R i \cdot c(i, j)}{\sum_{i=A}^B \sum_{j=L}^R c(i, j)} \\ G_J = \frac{\sum_{i=A}^B \sum_{j=L}^R j \cdot c(i, j)}{\sum_{i=A}^B \sum_{j=L}^R i \cdot c(i, j)} \end{cases} \quad (2)$$

$$\begin{cases} \sigma_i^2 = \sum_{i=A}^B \left(\sum_{j=L}^R c(i, j) \right) \cdot (i - G_i)^2 / \sum_{i=A}^B \sum_{j=L}^R c(i, j) \\ \sigma_j^2 = \sum_{j=L}^R \left(\sum_{i=A}^B c(i, j) \right) \cdot (j - G_j)^2 / \sum_{i=A}^B \sum_{j=L}^R c(i, j) \end{cases} \quad (3)$$

where, $c(i, j)$ indicates signature point matrix, A, B, L and R are distances above border, below border, left border and right border of signature respectively. The center point of normalized signature H matches the centre of gravity point P . The relation between the original coordinates (i, j) and transformed coordinates (x, y) is calculated by Eq. 4. The original signature and normalized signature is showed as the following Figure 3.

$$\begin{cases} i = A + \text{ROUND}(G_i - A / (\frac{L_{AB}}{2} - 1) * (x - 1)) \\ j = L + \text{ROUND}(G_j - L / (\frac{L_{LR}}{2} - 1) * (y - 1)) \end{cases} \quad (4)$$

where, $A \leq i \leq G_i, L \leq j \leq G_j, 1 \leq x \leq L_{AB}/2, 1 \leq y \leq L_{LR}/2$

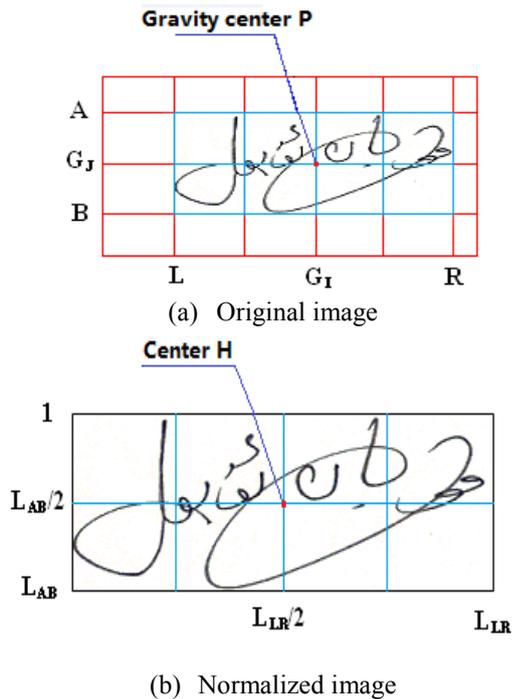


Figure 3. Size normalization of signature

3. FEATURE EXTRACTION

The selection of efficient features is fundamental in signature recognition systems. According to the extracted features, the classifier output the recognition rate. There are various kinds of features for signature recognition.

Grid features have been used extensively in the off-line signature recognition and verification approach [6]. The signature image is divided into many rectangular segments (often $12 \times 8, 15 \times 8$ segments), and for each segment, the area (the sum of foreground pixels) is calculated. Results are normalized so the rectangle with the highest number of black pixels is taken as one, and the rectangle with the smallest number of black pixels is taken as zero. Thus, there are resulting 96 features form the grid feature vector. In our approach, the signature image is divided into several rectangular segments in horizontal and vertical direction separately. The feature extraction method of modified grid information features is explained to take 8×8 segments as a case. The extraction procedure as following:

- Step 1: The signature image is horizontally divided into 8 segments indicated as horizontal blue line in Figure 4. A first feature in each grid is that the sum of the white pixels calculated from right point to black pixel firstly met with initially in the right-left direction. Next, the second feature is that the sum of the white pixels calculated between first and second block of black pixels in same direction. Thus, 16 features are extracted in the left - right direction in these 8 rectangular grids.
- Step 2: Another 16 features extracted from left to right direction indicated as Step 1.
- Step 3: The signature image is vertically divide into 8 segments indicated as vertical red line in Figure 4, and it is taken as a first feature in each grid that the sum of the white pixels calculated from highest point to black pixel firstly met in up- down direction. Then, it is taken as a second feature that the sums of the white pixels are calculated between first and second block of black pixels in same direction. Thus 16 features can be extracted from up-down direction.
- Step 4: Another 16 features extracted from down to up direction indicated in Step 3.

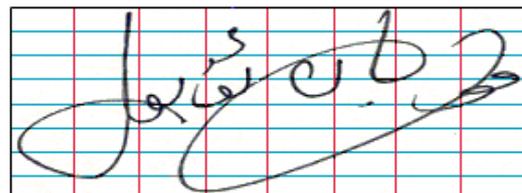


Figure 4. Segmentation of the signature

In this method, it is extracted 64 and 96 features for dividing image into 8×8 segments and 12×12 segments respectively. Consider with the nature of Uyghur signature which there are more strokes in horizontal direction than vertical direction, it is changed to take 3 features in each grid in Step1 and Step 2. Thus, it is get 80 and 120 features for dividing image into 8×8 segments and 12×12 segments respectively.

4. EXPERIMENTAL RESULTS

The ED classifier, K-NN classifier and Bayes classifier is used in our experiment. The databases used for experiments include 1000 signatures collected from 50 Uyghur persons were tested in our experiment. Two kinds of experiments are performed based on the training data set. That is, 800 images from the database were used for training data set, 200 images were used as the testing data set in the first kind of experiment. 500 images from the database were used for training dataset and testing data respectively in the second type of experiment. 96.0% of highest recognition rate was achieved using K-NN classifier taking 120 features indicated in Figure 6. The horizontal and vertical axes indicated K value of K-NN classifier and identification rate respectively in Figure 5 and Figure 6.

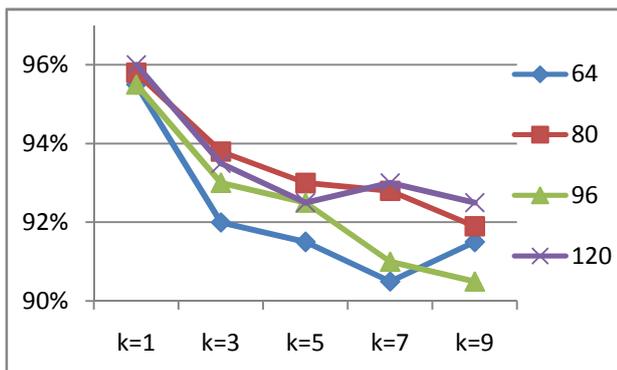


Figure 5. Identification rate of using K-NN classifier with different dimensional features (training 800)

The recognition rate is directly affected by the training number of dataset. So, the identification rates in Figure 6 are lower than in Figure 5. The more features, the higher identification rates. It is important for K-NN classifier to choose suitable value for K, because the identification rates are different in choosing different K.

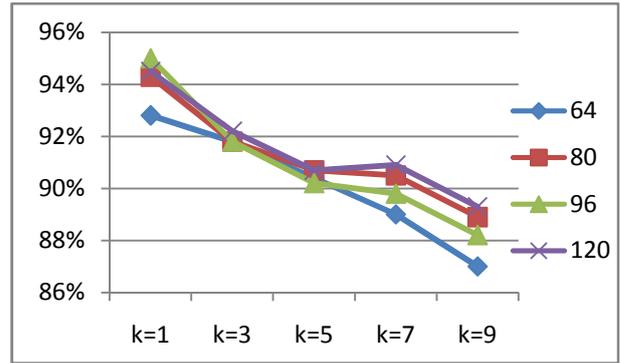


Figure 6. Identification rate of using K-NN classifier with different dimensional features (training 500)

Latin signatures are also tested to examine the effect of the features. Latin (English) databases used for the experiments include 500 signatures collected from 25 persons were tested in our experiment. 250 images from the database were used for training data set, 250 images were used as the testing data set. 95.20% of accuracy rate was obtained with K-NN classifier when extracting 120 dimensional features and others indicated in Figure7.

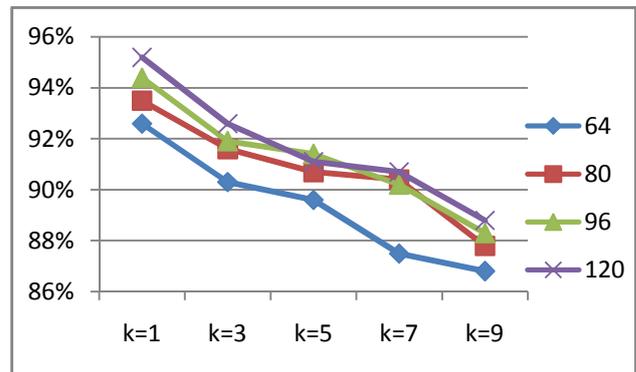


Figure 7. Identification rate of using K-NN classifier with different dimensional features (training 250)

Three kinds of different classifiers (ED, K-NN, and Bayes classifier) were used in the first kind of experiment, it is get 93.53% of average identification rate using K-NN classifier and higher than ED and Bayes classifier's as indicated in the following Table1.

Table.1 The recognition rate using different classifier

Classifier	Average recognition rate	
	Dimensional numbers of features	
	64	120
ED classifier	78.82%	86.45%
Bayes classifier	88.50%	89.26%
K-NN classifier	91.65%	93.53%

Random forgery Uyghur signatures are tested in first kind of experiment. 96 dimensional modified grid information features are extracted in this test. The individual recognition rate of the system to the random forgery is 0% (not match, can't find the class type correctly) when testing the forgeries without training. The overall recognition rates of the system using K-NN classifier is 92.15%. The individual recognition rate of the system to the random forgery is 100% (match, can find the class type correctly) when testing the forgeries with training. It is obtain 82.80% of overall recognition rates using K-NN classifier when each class contain 2 forgeries, in which one is used for training and the other is used for testing .While, the overall recognition rates of the system using K-NN classifier is reduced to 74.66% when each class contain 5 forgeries, in which four are used for training and one is used for testing. Experimental results indicate that the overall recognition rate is above 92% for genuine Uyghur signatures and below 75% for forgeries. So, the recognition system indicates quite strong robustness for Uyghur signature.

5. CONCLUSION

A Uyghur off-line signature recognition system was presented in this paper. Signature images were preprocessed according to the character of Uyghur signature. Then, modified grid information features were extracted based on the structure of Uyghur signature. Two kinds of experiments were performed for variations in the number of training and testing datasets. Experiments indicate that K-NN classifier has obtained more accurate classification rate (highest is 96%, and average is 93.53%) than ED and Bayes classifier. The feature was also extracted from Latin signature, and obtained 95.20% of accuracy rate. It can be concluded from the forgery test that the recognition system shows quite strong robustness for Uyghur signature. Experimental results indicated that the modified grid information features can capture the character of Uyghur signature and its writing style.

In future work, we plan to enlarge our database. We will explore more efficient features suitable for the nature of Uyghur signature, and perform more experiments for other methods and algorithms.

6. ACKNOWLEDGMENT

The National Natural Science Foundation of China (No. 61163028) and the Open Project of Xinjiang Laboratory of Multi-language Information Technology (No. 049807) supported this work.

REFERENCES

- [1] R. Plamondon, *Progress in Automatic Signature Verification*, World Scientific Publ., Singapore, 1994.
- [2] S. Pal, M. Blumenstein, U. Pal, "Non-English and non-Latin signature verification systems: a survey," *in proc of first International Workshop on Automated Forensic Handwriting Analysis (AFHA 2011)*, pp. 1-5, September 2011.
- [3] M. Radmehr, S. M. Anisheh, M. Nikpour, A. Yaseri. "Designing an Offline Method for Signature Recognition," *World Applied Sciences Journal*, vol. 13, no. 3, pp. 438-443, March 2011.
- [4] X. Chun-jing, L. Chun-li, Q. Yong-wei, "Wavelet packs and Guass model for off-line handwritten signature recognition," (in Chinese), *Computer Engineering and Applications*, vol. 45, no. 36, pp. 161-164, Dec. 2009.
- [5] W. Chun, J. Bin, F. Xun-sheng, M. Shi-ming, "Off-line Chinese signature recognition system based on neural network," (in Chinese), *Computer Application*, vol. 28, no. 9, pp. 2389-2391, Mar. 2008.
- [6] M. A. Abdala, N. A. Yousif, "Offline signature recognition and verification based on artificial neural network," *Eng & Tech. Journal*, vol. 27, no.7, July, 2009.
- [7] A. H. Rassoul, Y. C. Bucheet and M. I. Abd-Elmajied, "Off-line Arabic signature recognition based on invariant moments properties," *Eng. & Tech. Journal*, vol. 29, no.10, pp. 1936-1954, 2011.
- [8] K. Ubul, A. Hamdulla, and A. Aysa et al., "Research on Uyghur Off-line Handwriting-based Writer Identification," *in Proc. of the 9th International Conference on Signal Processing*, pp.1656-1659, Beijing, China, October 26-29, 2008.
- [9] J. Yuan. Research of Off-line Signature Verification System by Computer (in Chinese), MS thesis, Xinjiang University, Urumqi, Xinjiang, China, 2006.