Creation of Uyghur Off-Line Handwritten Signature Database

Ahat Yimin[1], Zulpiya Gheni[1], Mahpirat[2], Yunus Aysa[1] and Kurban Ubul[1,*]

[1]School of Information Science and Engineering, Xinjiang University, Urumqi 830046, China
[2]Education Section of Xinjiang University, Urumqi 830046, China
*Corresponding author

Abstract—**Until now, there has not been any major international effort that aims at comparing different signature verification methods systematically. However, most of the disclosed handwritten signature database is based on Latin, Chinese, Arabic and other languages. So creation of Uyghur off-line signature database plays very important role in estimating and comparing research results that was achieved by different groups of researcher. In this paper, the signature samples are collected by different age's people. For off-line Uyghur signature recognition and verification research, there are 500 persons genuine signature. (20 samples of each person), 185 persons simple forgery signature and skilled forgery signature (20 samples of each person), in total 17400 signature samples were collected. Each handwriting signature samples is scanned, stored with bmp file, and pre-processed (such as graying, noise reduction and binarization etc).The Uyghur signatures database very useful for designing a signature recognition and verification system that is optimized to verify Uyghur signatures.**

Keywords-uyghur; off-line; handwritten signature; database

I. INTRODUCTION

Over the past few decades, researchers around the world are widely studied automatic technologies for handwritten signature recognition and identification. Signature recognition and identification is different with the character recognition, because signature is often unreadable, and it is an image with some particular curves that represent the writing style of the person.

Signature database is indispensable for the research and development of signature recognition and identification algorithms. For a long time, various research institutions use self-built signature database for research and testing. Due to the use of the database is not unified, thus, not simply according to their accuracy results to judge the merits of the algorithm. A number of standard databases have been developed by the signature recognition community, off-line handwriting signature databases being the most popular ones.

Recently, some research institutions have published their own signature database and provide reference for other researchers to use [1~4]. The following is a brief introduction of these databases.

A. GDPS Database

This database is an off-line signature database established from Spain Las Parma province Gran Canaria University (ULPGC). This database contains 300 off-line signature classes, each classes consists of 24 genuine signatures and 30 skilled forgery signatures [5].

B. MCYT Database

This database is a multi-modal database established by the Spanish ATVS research group. This database consists of several sub databases, such as fingerprint, online signature, and off-line signature. The off-line signature sub-database is composed of 75 volunteers with 2250 signatures. Each class contains 15 true signature samples and 15 skilled forgery signature samples [6].

C. MYIDEAR Database

This is a multi-modal database that established by the Fribourg University of Fribourg, the Paris Institute of engineering and the French National Institute of telecommunications. Which includes face, voice, signature, hand writing and hand shape. The off-line handwritten signature database is composed of 630 genuine signature samples, 3780 skilled forgery signature samples and 14490 random forgery signature samples [7].

D. SVC Database

SVC database collected for First International Signature Verification Competition (SVC 2004). SVC contains signatures collected from 100 individuals, where there 20 genuine and 20 forgery signatures collected for each individual. Signatures in this database are not real signatures, in contrary these were made up by signers just for the sake of contributing to the database [8].

E. CEDAR Database

CEDAR is a research center at the University at Buffalo, State University of New York. This is a multi-modal database that established by the University at Buffalo (UB). Writers were asked to sign their signatures each on 2×2 inches paper. A total of 24 signature samples are collected from each of 55 writers. Some skilled forgers tried to forge the signatures in the database of 55 writer's original signatures. So, the total size of the signature database is 2640 signatures [9].

In general, the signature types involved in the off-line signature database are relatively small. Signature database were generally composited by western. Meanwhile, about signature recognition and verification research are mostly based on Latin [10], Chinese [11] and Arabic [12] handwritten signature and other languages, but one of them related with Uyghur handwritten signature. A lot of research has been carried out in

Uyghur handwritten signature recognition [13, 14] and verification [15].

Although, so far it has been established Uyghur handwritten document image database [16], but there is no a standard database for Uyghur handwritten signature. So, there is still much research space for creating or developing standard database that suitable for Uyghur handwriting signature recognition and verification field.

II.   THE NATURE OF UYGHUR HANDWRITTEN SIGNATURE

Today, Uyghur mainly live in the Xinjiang Uyghur Autonomous Region (hereafter: Xinjiang) in China. In the 19th century, after Uyghur accepting Islam, gradually using the script which based on the Arabic alphabet languages. This script has been continuously improved, so far it has become the common script language of the Uyghur people. Uyghur script is absorbed the features of Arabic and Persian characters to formed their own writing system. But Uyghur is different from Arabic and Persian in some respects. Modern Uyghur including eight vowels and 24 consonants, a total of 32 letters, on the morphological structure belongs to agglutinative language types. Because of the different position of Uyghur letter form, the letters differ slightly. 25 letters in the 32 letters have 4 kinds of different forms: "initial form", "intermediate form", "final form", and "isolated form". 5 letters in the 32 letters have two kinds of writing form: "initial form", and "final form". The other 2 letters have eight kinds of writing form, each writing form including two forms. Parts of the Uyghur character styles as the following table 1.

TABLE I. PARTS OF THE UYGHUR ALPHABET

final form	intermediate form	initial form	isolated form
سِّي سِّ	جِّ جِّ	سُّ ئِّ	نِّ ي
ت	ت	ت	ت
ـز			ز

The structure characteristics of Uyghur letter is complex:

1. whether in print or in writing, the writing direction is from right to left, and the letters are written down to the baseline, which can be connected to form one or more letters or sentences. It is different from Latin and Chinese. The writing direction as shown in figure 1:
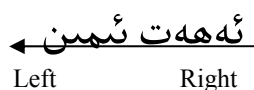
ئەھەت ئۇمىن

Left   Right

FIGURE I. WRITING DIRECTION

2. The Uyghur characters stroke is less, the main body of a lot of letters was same stroke, and adding some additional parts can constitute different letters, such as the letters ۋ and ۇ, the main body of these letters is ۋ. The number of strokes can be used to distinguish.

3. A lot of Uyghur characters with point stroke, the point strokes have three kinds of markers, like a point, two points, three points etc. The difference between individual letters is only determined by the different number of point strokes. For example: the letters ن and ت, the letter ب and the letter پ, the points mark located below and above of the letters. Sometimes it is divided into vertical and horizontal points. For example: the letter ي and the letter ي, these letter's main body is ى, each is a different letter. The difference between the key block and additional block for some of the letters is the position of the same attachment points. For example: the letters ب and ن, the letters ج and خ.

4. The main part of the Uyghur letters stroke is written along the baseline, additional strokes and other shape point, stroke point are located in higher or lower than the baseline, the baseline can be considered as a reference line to identify the position of adding portion. As shown in Figure 2.
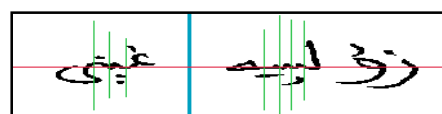
FIGURE II. UYGHUR HANDWRITTEN SIGNATURE (GREEN VERTICAL LINE IS THE LINKING POINT BETWEEN THE LETTERS, RED horizontal LINE IS THE BASELINE, BLUE VERTICAL LINE IS THE DIVIDING LINE BETWEEN THE UYGHUR FIRST NAME AND LAST NAME)

The characteristic form of the Uyghur signature can be summed up the following three points:

1. Writing form of structural unit. Uyghur handwriting signature with the word as a unit, according to the standardized writing Uyghur words, words of the same kind of chirography, various components have standard written form, However, due to the differences in writing habits, may show different forms of writing characteristics.

2. The relative position of the structural unit is not fixed. Write the various structural units of Uyghur word requires a reasonable match, but most people used to break this restriction, which characteristic form such as: too tight, too loose, left and right asymmetry, upper and lower size is not fine.

3. Even pen can be violation of the Uyghur word specification. The signature of a person is not consistent with the variance structure characteristics of the word rule. Following figure 3 showed as the different forms of signature.
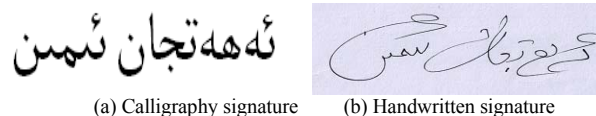
ئەھەتجان ئۇمىن

(a) Calligraphy signature   (b) Handwritten signature

FIGURE III. DIFFERENT FORMS OF UYGHUR SIGNATURE

III.   DESCRIPTION OF UYGHUR DATABASE

Signature collection is the first step for creating Uyghur off-line signature recognition and verification database and it is also the raw material for off-line signature research field .The signatures were collected from different gender and different age native Uyghur people. Each writer was asked to sign her/his signature on a white A4 sheet of a paper. Each paper

was divided into 21 same sized boxes and area of each box is big enough to write his/her natural signature. Thus, each signer need to write own signatures 21 times on per page. The signature data samples used the scanner to digital, which is used as the signature of the original data. In here using CanonMP810 scanner scanning each signature samples with 300dPi. These signature samples using the Photoshop to make size normalization (Width 384, High 96), and which is saved in a BMP format and Storage to the computer according to a certain naming rule.

A. Sub Database for Off-Line Signature Verification

The Uyghur handwritten signature database can be used in signature recognition and verification research. In verification field, the signature of each person was included genuine signature (original signature) and forgery signature. Meanwhile, the signatures of forgeries will be classified in two basic types.

(1) Simple forgery: the forger produce the signature in his/her own writing style rather than access the original signature sample.

(2) Skilled forgery: the forgery well known the shape of genuine signature and practicing it number of times.

For off-line Uyghur signature verification research, there are 500 people's genuine signature (20 samples of each person), 185 person's simple forgery signature and skilled forgery signature (20 samples of each person), in total 17400 signature samples were collected. Then the collected signature samples through a series of pre-processing, and which directly used to be feature extraction. Following figure 4 showed as the original signature image, skilled forgery signature image and simple forgery signature.
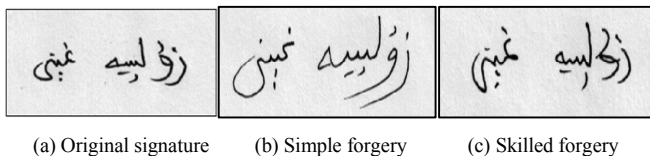

(a) Original signature   (b) Simple forgery   (c) Skilled forgery

FIGURE IV. SAMPLE OF UYGHUR HANDWRITING SIGNATURE

B. Sub Database for Off-Line Signature Recognition

The signature recognition is a multi-classification problem, and the signature verification is a two classification problem. So they have different requirements for the establishment of a handwritten signature database. That is, in the field of signature recognition, only need to collect the original signature of the signer and not need the forgery signature. In signature recognition part, collected a total of 500 individual's signature sample, each of the 20 samples, a total of 10000 signature samples.

IV. PRE-PROCESSING METHODS FOR SIGNATURE IMAGES

Before establishing a standard handwritten signature database, it is necessary to pre-processing at first. Here, the pre-processing step included graying, noise reduction and binarization.

A. Graying

The graying of the signature image is that, converting a 256 color BMP image into a grayscale image, that is:

$$\Gamma\rho\alpha\psi\,(\iota, \varphi) = 0.11*P(\iota, \varphi)+0.59*\Gamma(\iota, \varphi)+0.3*B(\iota, \varphi) \quad (1)$$

where i, j represents the location of a pixel point in the signature image. R, G, B represents the red component, green component and blue component, respectively. The following figure 5 showed as the original signature image and its graying signature image.
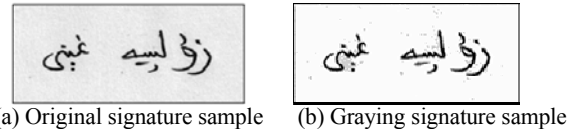

(a) Original signature sample   (b) Graying signature sample

FIGURE V. A SAMPLE OFGRAYING UYGHUR SIGNATURE

B. Binarization

The main purpose of binarization is that to separate the background and foreground of the signature image. There are selecting threshold method used to be signature binarization. That is, the value of the threshold used to change signature image to black (0) or white (255), its formula is:

$$G(i,j) = \begin{cases} 255 & , \ f(i,j) \geq T \\ 0 & , \ others \end{cases} \quad (2)$$

where, T is the value of specific threshold which have been selected; $f(i,j)$ represents the gray level of the pixel of original signature images. The following figure 6 showed as the graying signature image and binarized signature image.
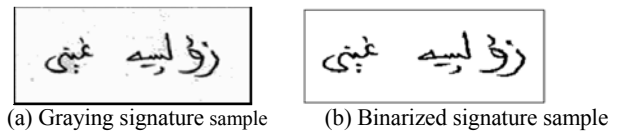

(a) Graying signature sample   (b) Binarized signature sample

FIGURE VI. GRAYING SIGNATURE IMAGE AND BINARIZATION

C. Binarization

In the process of obtaining, transmission and storage, a great deal of noise is involved in and meanwhile scanned signature images may have image noise caused by the sensors of the scanner, so the quality goes bad. Signature image noise reduction refers to the use of various filter models. Here, use the standard median filter for noise reduction. Standard median filtering algorithm is divided into the following four steps:

1. Select the appropriate filtering window as the solid neighborhood of the pixel point p(i, j). In here, the 3*3 square filter windows were used.

2. The gray value of all pixels in the filter window is arranged according to the order from small to large.

3. After sorting the gray value of middle position, that is, as the median value.

4. Replace the gray value of the pixel point p(i, j) with the value of step(3).

Standard median basic formula is as follows:

$$f(i,j)_{SM} = median\{ f(r,s) , \quad f(r,s) \in N_{f(i,j)}\} \quad (3)$$

In the formula, $f(r,s)$ represents the gray value of any pixel in the filter window, and $N_{f(i,j)}$ represents the solid neighborhood of $f(i,j)$

Standard median filter can not only remove or reduce the random noise and pulse interference, but also can keep the information of the image edge. The following figure7 showed as the binarization signature image and noise reduction signature image.
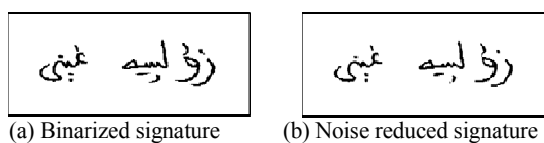

(a) Binarized signature (b) Noise reduced signature

FIGURE VII. A SAMPLE OF NOISE REDUCTION FOR UYGHUR SIGNATURE

V. CONCLUSION AND FUTURE WORK

In this work, there are 17400 signature samples collected by 870 native Uyghur writers of different ages, different gender and different education backgrounds. During the signature collection process, each Uyghur writer filled a total of 21 forms on a white A4 sheet of a paper firstly. Then, the collected signatures was scanned (300dpi) and size normalized (Width 384, High 96) artificially. Next, each signature samples was pre-processed based on the nature of Uyghur handwritten signature. The pre-processing step included graying, noise reduction and binarization. Since the differences of type and nature of signature samples in handwritten signature recognition and verification research, where, in order to make it easy to distinguish and research, the whole handwritten signature database was divided into two parts, which is database for off-line signature recognition field and database for off-line signature verification field. For the signature recognition research, collected a total of 1000 signature samples .In signature verification part, in addition to 1000 original signatures (genuine signatures) were collected in recognition database, there are 3700 forgery signature samples and 3700 skilled forgery signature samples was collected in different time by different signer. So that, this is the first Uyghur off-line signature database that available freely for worldwide research.

The further work of the database is likely to include an increased number of signer's handwritten signature and more standardized signature samples in order to suitable for the nature of Uyghur handwritten signature. Meanwhile, the signature images in complex background or different situations, such as checks, a bank notes, forms, etc. will be collected in the future work. The database, described in this paper will be available for the researchers worldwide.

REFERENCES

[1] J. Wen, M. H. Chen, and J. X. Ren, "Off-Line signature verification based on local structural pattern distribution features," Pattern Recognition: 6th Chinese Conference, CCPR 2014, Changsha, China, November 17-19, 2014,pp. 499—507.

[2] R. Doroz, P. Porwik, and K. Wrobel, "Signature recognition based on voting schemes, " Biometrics and Kansei Engineering (ICBAKE), 2013 International Conference on, Tokyo, 2013, pp. 53-57.

[3] M. Zalasiński , K. Cpałka, Y. Hayashi, "A new approach to the dynamic signature verification aimed at minimizing the number of global features," 15th International Conference on Artificial Intelligence and Soft Computing (ICAISC 2016), pp.218−231.

[4] M. M. Kumar and N. B. Puhan, "Inter-point envelope based distance moments for offline signature verification," 2014 International Conference on Signal Processing and Communications (SPCOM), Bangalore, 2014, pp. 1-6.

[5] J. F. Vargas, M. A. Ferrer, C. M. Travieso, and J. B.Alonso,"Off-line handwritten signature GPDS-960 corpus," Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), Sept.2007: 764-768.

[6] J. O. Garcia, J. F. Aguilar, and D. Simon, "MCYT baselinecorpus:a bimodal biometric database," IEE Proc. Vision Image Sensor Process, 2003, 150(6): 395-401.

[7] B. Dumas, C. Pugin, J. Hennebert, D.P.Delacrétaz, A .Humm, F.Evéquoz, R.Ingold, D.Von Rotz, "MyIdea-Multimodal biometrics database, description of acquisition protocols," Proc. of. 3rd COST,Hatfield, U.K., Oct.2005:.59-62.

[8] Dit-Yan Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto and G. Rigoll "SVC2004: First International Signature Verification Competition", Biometric Authentication: First International Conference, ICBA 2004, pp.16-22.

[9] CEDAR Signature Database: http://www.cedar.buffalo.edu/about.html

[10] A. Foroozandeh, Y. Akbari, M. J. Jalili, and J. Sadri, "Persian signature verification based on fractal dimension using testing hypothesis," Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on, Bari, 2012, pp. 313-318.

[11] Z. S. Lin and L. F. Liang, "Off-line handwritten Chinese signature verification based on support vector machine multiple classifiers," Advances in Electric and Electronics, 2012, Springer Berlin Heidelberg, pp.563-568.

[12] S. M. A. R. Ahmed, "Off-line Arabic signature verification using geometrical features, " WIAR '2012; National Workshop on Information Assurance Research; Proceedings of, Riyadh, Kingdom of Saudi Arabia, 2012, pp. 1-6.

[13] K. Ubul, A. Adler, and N. Yadikar, "Effects on accuracy of Uyghur handwritten signature recognition, " In: Liu, C.-L., Zhang, C., Wang, L. (eds.) CCPR 2012. CCIS, vol. 321, pp. 548–555. Springer, Heidelberg (2012).

[14] G. Abliz, K. Ubul, and K. Moyidin, "Research on off-line Uyghur signature recognition technology based on multiresolution geometric features," Computer Engineering and Applications, 2013, 49（16）: 168-171.

[15] K. Ubul, T. Yibulayin, and A. Aysa, "Off-Line Uyghur handwritten signature verification based on combined features," 6th Chinese Conference on Pattern Recognition (CCPR 2014), Changsha, China, November 17-19, 2014, Springer Berlin, pp.491—498.

[16] K. Ubul, M. Zunun, A. Aysa, N. Yadikar, and U. Yunus, "Creation of Uyghur offline handwritten database," 2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA) Algiers, 2013, pp. 291-295.