International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.7, No.5 (2014), pp.409-424
http://dx.doi.org/10.14257/ijsip.2014.7.5.35

Research on On-line Uyghur Handwritten Character Recognition Technology Based on Modified Center Distance Feature

Askar Hamdulla, Wujiahemaiti Simayi, Mayire Ibrayim and Dilmurat Tursun

Xinjiang University, Urumqi Xinjiang 830046, P.R. China
askar@xju.edu.cn

Abstract

Through the analysis on the unique characteristics of Uyghur characters, in order to further improve the recognition rate, this paper developed the Center Distance Feature (CDF) to its modified form which is named as Modified Center Distance Feature (MCDF). By combination with some low dimensional features including stroke number feature, additional part's location feature, shape feature, bottom-up and left-right density feature(BULR) in experiments, MCDF gifted robust recognition accuracy of 98.77% for the 32 isolated forms of Uyghur characters. MCDF increased the recognition accuracy by 4.51 points comparing with the result from the combination of CDF with the same low dimensional features mentioned above, which is 94.16%. This paper used the samples from 400 different volunteers. The recognition system is trained using 70 percent of 12800 samples from 400 different writers and tested on the remained 30 percent.

Keywords: *Uyghur characters, On-line handwriting recognition, Low dimensional features, Modified center distance feature*

1. Introduction

Character recognition is one of the most important areas of pattern recognition, with major implications in automation and information handling [1]. Character recognition can be categorized into two types, such as the printed and handwritten character recognition technology [2]. Handwritten character recognition technology itself can be divided into two sub fields, namely off-line and on-line character recognition [3]. The off-line character recognition mainly deals with the recognition of character patterns in a scanned digital image [4]. In On-line handwriting recognition, the transducer device converts the pen's writing movement into a sequence of signals and then sends the information to the computer or the intelligent system [5]. This complies with today's tendency to develop machines and computers with interfaces acquiring human-like skills.

Uyghur is one of main languages in Altaic language system. Uyghur is an alphabetical scripture that consists of 32 basic characters including eight vowels and 24 consonants [6]. Uyghur characters haven't upper and lower cases, instead each character has various writing forms, with at least two and at most eight forms. Except from the isolated form for each character, most of the characters have certain forms according to the character position within a word and different forms of characters are put at corresponding position. Most of the characters are comprised of main part and additional part which are critical to distinguish similar characters from each other. Although there are some regulations to write characters correctly, various kinds of writing styles cannot be prevented in actual handwriting. Writing style and order of

ISSN: 2005-4254 IJSIP

handwritten characters are influenced by different writers and the writer's mood, writing context, *etc*.

With populous users, Uyghur character recognition is also receiving great demand and enjoying more and more significance in daily life of people in western china. There was successful work on the printed Uyghur character recognition, but still not much work has been done on the handwritten character recognition. The research on Online Uyghur handwritten recognition technology is late started and not meeting the urgent needs from people. Finding out the effective features of characters and the applicable feature extraction method is the very beginning demand and is critical to improve the recognition accuracy and efficiency of the character recognition system [7]. Our work is also an endeavor to develop the character recognition technology for Uyghur letters to the level that people can conveniently communicate with the implementation of this technology. The research and implementation of online Uyghur handwritten character recognition technology is not only beneficial to promote the modernization process of Uyghur people, but also helpful for other relative groups who use similar letters.

2. On-line handwritten character recognition system

The function units of the handwritten character recognition system applied for Uyghur characters in this paper is shown in Figure 1. As shown in Figure 1, the system consists of five function units, such as data collecting, preprocessing, feature extraction, training and testing. The results from each function unit are greatly dependent on completing the tasks in previous function units, and influence the efficiency of proceeding works.

Data Collecting The first-hand data collected from different writers are inputted to the system at first. In online handwritten recognition system, the first-hand data are usually the (x, y) coordinates of the pen's trajectory on writing pad.

Pre-processing The pre-processing unit mainly deals with the noise elimination and normalization for the raw materials, so that many conveniences are provided for the latter stages in recognition system [8]. Linear and non-linear combined method based on point density is applied to normalize the raw data of character samples. Take the resolution of the sample collecting device into consideration, the 96*96 lattics are used to normalize the handwritten samples.

Feature Extraction One of the most critical works is processed during the feature extraction, as the whole recognition system and the recognition results are dependent on the effective features that can express the unique characteristics of characters. This unit provides the feature vectors for model building and testing. The extracted and applied features in this paper will be introduced in section 4.

Training and Model Building It is very clear that the characters are differed from each other by their uniqueness in structure and writing order, *etc.*, the data of the features extracted from the characters are classified to several sub groups and build the template library for each character by training. In this paper, the quick-search classification method [13, 14] is applied for model building in training part. At first, the characters are categorized into several sub-templates according to stroke number feature and additional part's location feature. Then the feature vectors of characters are classified for each sub-template. Hence, we can get several templates for one character.

Testing The tested characters are compared with the templates and the recognition result is obtained from recognition unit. The characters are identified by the minimum distance classifier using Euclidean distance in recognition unit.

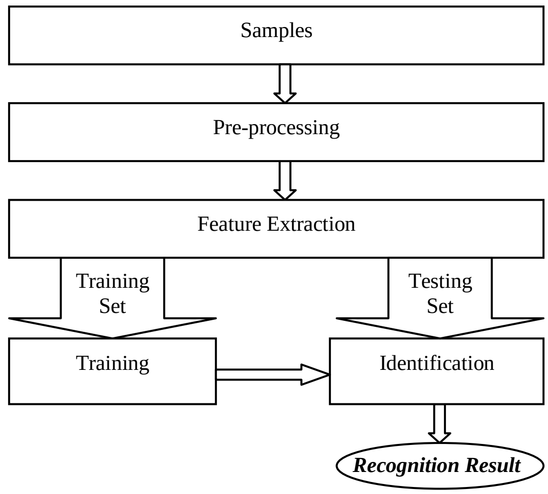


Figure 1. Block Diagram of On-line Handwritten Uyghur Character Recognition System

3. Related Work

On-line Uyghur handwritten recognition technology is just started and still in its preliminary stage. Although it is highly demanding to get higher accuracy in on-line handwritten recognition, considerable work has been undertaken on this field and made some improvements to some extent. Tsinghua University and Xinjiang University contributed most efforts with distinguished results [9]. Comparing with the other function units, the feature extraction is perhaps more challenging and innovative. Our records from previous works focus on the feature extraction algorithms and the classification methods which are fit for the specialty of Uyghur characters.

Mayire (2009) analyzed the characters with equal sampling feature [10]. According to this method, the points are selected from the strokes of character shape in equal distance calculated by the number of points, and the selected points are recorded as the feature of character.

Mutallip (2010) used the gradient directional feature method for feature extraction of on-line handwritten Uyghur characters [11]. Support vector machine (SVM) for classification was proved helpful in experiments in his paper.

Alimjan (2010) selected some commonly used features including direction, loop and angle features to observe Uyghur characters [12].

Ranagul (2011) discussed the grid directional feature for Uyghur characters. The characters are observed using eight directions within each divided grid unit [13].

Mayire (2012) and Kamil applied a quick-search classification method for model building [13, 14]. The classification method consists of two acceptable sub-steps. In the first step, low dimensional features are used to classify the characters. After the pre-classification from the first step, the other features of characters are classified for model building and identification.

Piguilin (2012) propose a delay stroke processing method which contains seeking and projection of delay strokes for Uyghur characters [15].

Zulpiya (2013) discussed the peripheral contour feature and presented the shape feature for on-line Uyghur handwritten character recognition [16]. The former approach involves the statistical and structural information of character while the latter one is describes the special symbols of Uyghur scripture. The shape features focus on finding and distinguishing the special symbols of Uyghur characters. The shape feature is low dimensional and easily combined with other features. It delivered obvious increasing in different feature combined experiments.

Wujiahemaiti (2013) presented the center distance feature (CDF) with its three different implementations (CDF-2, CDF-4 and CDF-8) to extract the features of Uyghur characters [17]. The center distance feature combines the structural and statistical characteristic of characters and used evenly and unevenly grid division method. The unevenly grid division method was one of the main contributors to improve the recognition accuracy in experiments.

This paper applies the multiple features combination method and proposed the modified center distance feature (abbreviated as MCDF) which showed much progress in recognition accuracy. Only the isolated forms of characters are participated into experiments.

4. Feature Extraction Algorithm

Selection of a feature extraction is probably the single most important factor in achieving high recognition performance [18]. Researchers in the field of handwritten recognition have proposed different approaches for feature extraction, such as statistical, structural, and neural network approaches [19~21]. Appropriate and effective feature selection can increase the recognition accuracy very much. Feature extraction for online handwritten recognition is highly dependent on handwriting process. While writing on the sensible pad, the pen's trajectory is recorded. Pen trajectory is usually described by its position, the (x, y) coordinates, and the states of pen's lifting and touching movements.

This paper takes Uyghur characters' characteristics and writing rules into consideration, according to the principle of feature extraction, chooses the bottom-up and left-right density feature(BULR), shape feature , stroke number feature, additional part's location feature to combine with the center distance feature(CDF) and its modified form(MCDF) with three different implementations (MCDF-2,MCDF-4 and MCDF-8). This section gives a brief introduction for the features used in experiments and their feature extraction methods.

4.1. Center Distance Feature-CDF with its Three Different Implementations (CDF-2, CDF-4 and CDF-8)

The center distance feature presented in reference [17] analyzes the characters by dividing the rectangular character shape into 32 grids. It describes a character by the distances from the gravitational centers of grid units to the gravitational center of character shape.

The three implementations of the center distance feature are mainly differed in the numbers of bins used for grid division. As the three methods are same dimensional which is 32, the numbers of grid units in each bin are different according to the method applied. The grids are unevenly divided, so that the grid units are various in sizes.

4.2. Modified Center Distance Feature-MCDF and its Comparison with Center Distance Feature-CDF)

In fact, there is not very much difference between the MCDF and CDF features in extracting algorithms. MCDF uses all the divided grids as effective feature blocks no matter the grid does or does not have any points from the pen's trajectory, while CDF only uses the pen trajectory located grids. The three implementations of MCDF are differed by the number of bins used for grid division.

A. Modified Center Distance Feature with Four Bins MCDF-4

The MCDF-4 feature is obtained by following steps:

Step 1: Calculate barycentric coordinates P(CentX, CentY) (gravitational center of character shape) from every character point sequence according to the following expression

$$\text{CentX} = \frac{1}{n}\sum_{i=1}^{i=n} x_i \quad \text{CentY} = \frac{1}{n}\sum_{i=1}^{i=n} y_i \tag{1}$$

(n represents the total points of a character shape)

Step 2: The Center of gravity P(CentX, CentY) divides the character's external rectangular shape into the left and right bins(B_L and B_R) by CentX at first. The left and right bins are further divided into two sub-bins respectively by (CentX)/2 and (96+CentX)/2. Such 96*96 rectangular shape of the character can be processed in four bins. Each bin provides upper and lower parts by the barycentric coordinate of character CentY. So, there are eight parts in the rectangular shape of character. See Figure 2. (a).

Step 3: Each part is evenly split into four units as sub-graph in vertical direction. See Figure 2(b), so that 32 grids are obtained from the rectangular shape of character. See Figure 2 (b).

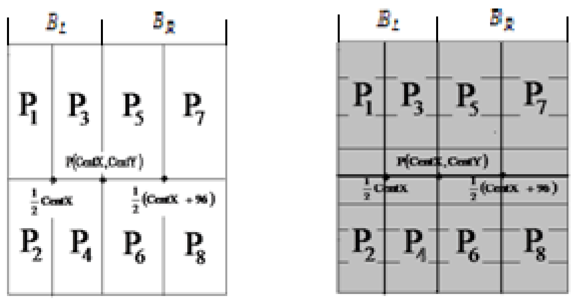


Figure 2. (a) Eight parts Figure 2. (b) 32 Grids

Step 4: Calculate barycentric coordinates $P_j(cenx_j , ceny_j)$ (gravitational center) from the points in each grid using following expressions

$$\text{cenx}_j = \frac{1}{\text{m}}\sum_{j=1}^{j=\text{m}} x_j \quad \text{ceny}_j = \frac{1}{\text{m}}\sum_{j=1}^{j=\text{m}} y_j \tag{2}$$

(m represents the total points in a grid). If there is not any point in the grid, the barycentric coordinate is recorded as -1 which represents that there is no real barycentric coordinate of this grid.

Step 5: Calculate the distance from the barycentric coordinate of each grid to the gravitational center of character shape by following expression

$$d_j(p_j, \text{P}) = \sqrt{(cenx_j - CentX)^2 + (ceny_j - CentY)^2} \tag{3}$$

If there is not any point in the grid, the distance is recorded by the distance from the gravitational center of character shape to the origin of coordinate O(0，0) according to the expression(4). See Figure 2 (c)

$$D(O, P) = \sqrt{(CentX - 0)^2 + (CentY - 0)^2} \tag{4}$$

Thus 32 dimensional center distance feature with four bins is produced by 32 grids. See Figure 2 (c)

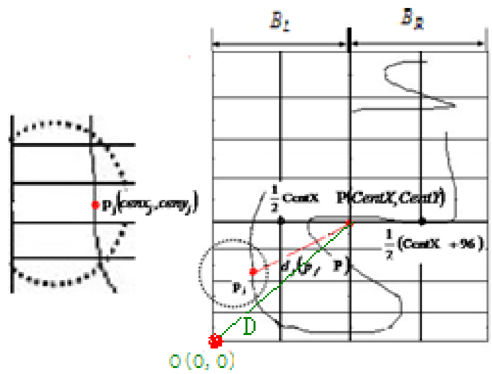


Figure 2. (c) MCDF-4 Feature Extraction

B. Modified Center Distance Feature with Two bins-MCDF-2

The clear difference between MCDF-2 and MCDF-4 is the number of bins to generate the grids in rectangular shape of character. MCDF-2 uses two bins while MCDF-4 applies four bins to make grids.

Almost every step to extract the MCDF-2 feature is similar to the MCDF-4. The different points in extraction are as follows:

In step 2: MCDF-2 only uses two bins divided by the barycentric coordinate. The Center of gravity P(CentX, CentY) divides the character's external rectangular shape into two bins and altogether four parts. (left and right bins by CentX, upper and lower parts for each bin by CentY, respectively), see Figure 3.

In step 3: each part is evenly split into eight units as sub-graph in vertical direction. See Figure 3, so one character's external rectangular shape is divided into 32 grids, see Figure 3.

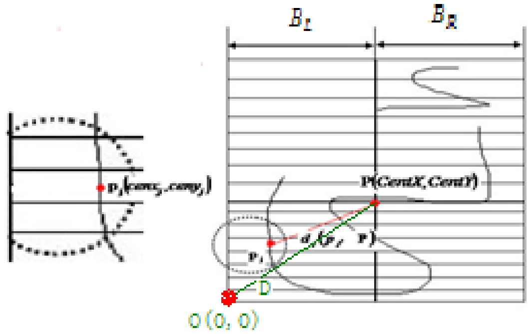


Figure 3. MCDF-2 Feature Extraction

C. Modified Center Distance Feature with eight bins—MCDF-8

MCDF-8 is a further modified method which used eight bins to generate the grid units. Similar to above two forms of Modified center distance feature, 32 dimensional features are produced from MCDF-8 method with following differences to the extraction of MCDF-4:

In Step 2: The left bin is divided into four sub-bins respectively by (CentX)/4, (CentX)/2 and (CentX)*3/4; and right bin is divided into another four sub-bins respectively by the points (3*CentX+96)/4, (96+CentX)/2 and (CentX+3*96)/4. Such the rectangular shape of character can be processed in eight bins. Each bin provides upper and lower parts by the barycentric coordinate of character CentY. Now, 16 parts are in the rectangular shape of the character. See Figure 4.

In Step 3: Each part is evenly split into two units as sub-graph in vertical direction. See Figure 4, 32 grids are obtained and 32 dimensional Center distance feature with eight bins is produced.

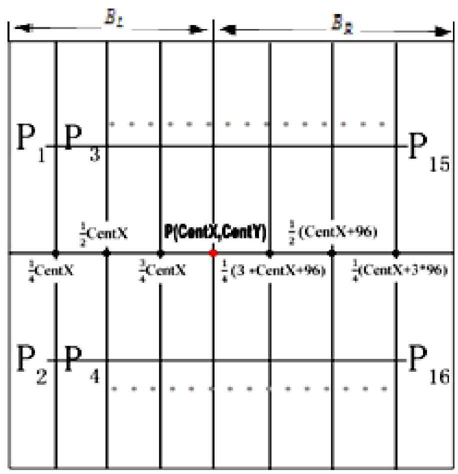


Figure 4. Eight Bins of MCDF-8

D. Difference between CDF and MCDF

Center distance feature CDF uses the Figure 1 for the grids without pen trajectory and makes those grids useless to describe characters [10], while MCDF uses all divided grids to describe the statistical and structural information of characters. MCDF lets the empty grids without pen trajectory points also describe characters by recording the location information of the character's gravitational center. The recorded same information for empty grids makes the similar characters nearer in feature domain, and simultaneously puts them far from different characters. As a result, the grids which couldn't participate to distinguish characters in CDF method become useful in MCDF method and make their contributions to improve the recognition rate. See Figure 5.

D	D	D	D
D	D	dj	dj
D	D	D	dj
dj	dj	dj	dj
dj	dj	D	D
d;	dj	dj	D
dj	D	dj	D
dj	dj	dj	D

Figure 5(a). Feature by MCDF-4

−1	−1	−1	−1
−1	−1	dj	dj
−1	−1	−1	dj
dj	dj	dj	dj
dj	dj	−1	−1
dj	dj	dj	−1
dj	−1	dj	−1
dj	dj	dj	−1

Figure 5(b). Feature by CDF-4

D refers the distance from the gravitational center of character shape to the origin of coordinate O(0，0), while dj refers the distance from the barycentric coordinate of each grid to the gravitational center of character shape

4.3. Stroke Number Feature

Stroke number feature refers to the number of strokes that form the handwritten character trajectory. Stroke number feature is recorded by the number of pen's leaving and touching to the recorder device screen. For example, character ۋ has three strokes.

4.4. Additional Part's Location Feature

Uyghur characters are usually composed of one or more strokes. The longer or the longest one among strokes is called main stroke. Most characters appear with one main stroke and a few additional strokes. The additional strokes besides the main stroke are called the additional part of character and the main stroke is named as main part. The additional part is located over or within or under the main part. So we have to take the following cases into account and some simple figures below are used to represent the different cases.

- 0 represents that the character has not any additional part, such as د.
- 1 represents that the additional part is over the main part, such as ذ.
- 2 represents that Additional part is within the main part, such as ج .
- 3 represents that the Additional part is under the main part, such as ب.
- 4 represents that the Additional strokes are both at over and under the main part, such as . ئې

4.5. Bottom-Up (BUDR) and Left-Right (LRDR) Density Ratio

This method was applied for Arabic characters [20]. Uyghur characters are related to the Arabic characters in many ways. So the feature extracted by this method is also participated into the experiments in this paper.

Many Uyghur letters have a noticeable property that the points from the pen trajectory are not equally distributed on the normalized lattics, See Fig.6. If the normalized sample is observed with 5*5 square grids, we easily calculate the ratio between the pixels of the written letter in the first two rows and in the last two rows; or between the first two columns and the last two columns. The first case is called bottom-up ratio while the second one is called left-right ratio. Every row or column contains five equal size cells. In order to calculate bottom-up or left-right density ratios we use the following expressions:

$$BUDR = points \ \ in \ (R1 + R2)/points \ \ in \ (R4 + R5) \tag{5}$$

a. If BUDR>1, then recorded as 1, means up-oriented;
b. If BUDR<1, then recorded as 0, means bottom-oriented

$$LRDR = points \ \ in \ (C1 + C2)/points \ \ in \ (C4 + C5) \tag{6}$$

a. If LRDR >1, then recorded as 1, means left-oriented:
b. If LRDR<1, then recorded as 0, means right-oriented:

Above two kinds of ratios give us two dimensional features that BUDR is charge of the first dimension and LRDR is for the latter dimension.

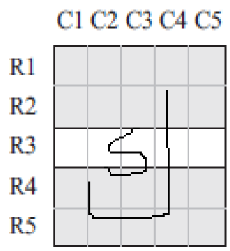


Figure 6(a) Bottom oriente

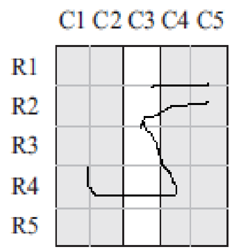


Figure 6(b) Right oriented

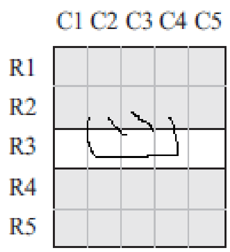


Figure 6(c) Up oriented

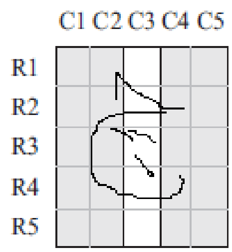


Figure 6(d) Left oriented

Figure 6. Orientation Behavior of Characters

4.6. Shape Feature of Additional Strokes

The strokes after the first stroke are called the additional strokes or the additional part. The additional part includes a dot or a group of dots and five special symbols such as ء , ۇ , ٚ , ا , ـ . Figure 6 shows some examples of characters with some kinds of additional parts.

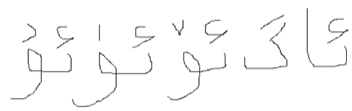


Figure 6. Handwritten Characters with Additional Parts

The shape feature of additional stroke is extracted by following steps [16]:

1. After the first stroke finished, visit the coordinates of each point from additional stroke, and calculate the slope of the line generated by the first and last point of the stroke.
2. Extract the directional feature through visiting each stroke's trajectory. Estimate the value of angle that made up of the directions on this and former points. If the angle value is greater than 90^o it means that there is a change in direction. Then the number of changes is recorded.
3. Judge the shape of additional stroke by the numbers of change.
1. If the number of changes is two, the shape ء is favored;
2. If the number of changes is one, the shape can be searched among ٚ or ۇ, so there should be a further judgment for the two shapes:
a. If the slope $|k|>1$, shape ۇ is favored;
b. If the slope $|k|<1$, shape ٚ is favored;
3. If there is no changing recorded, the shape can be searched among ا or ـ , so there should be a further judgment using slope:
a. If the slope $|k|>1$, shape ا is favored;
b. If the slope $|k|<1$, shape ـ is favored;

4. Count the number of shapes from the additional strokes and record for shape feature.

The extracted shape feature is five dimensional. Each position in feature code represents for one special shape of the additional strokes mentioned above. Fig.7 shows the arrangements of shapes in feature code.

Shapes					
Codes (yes/no)	***1/0***	***1/0***	***1/0***	***1/0***	***1/0***

Figure 7. Arrangements of Shapes in Feature Code

For example, the code 10100 means that the current character has one shape and one shape.

5. Experimental Results and Analysis

A total of 400 different writers contributed the 12800 different samples for the experiments in this paper. The system is trained using 8960 samples from 280 writers, which are 70 percent of total samples. And 3840 samples from 120 writers which are 30 percent of total samples are participated into identification test.

The character recognition experiments are conducted using the introduced features above in following function orders:

- The stroke number feature and the additional part's location feature are used for pre-classification in training and testing.
- The Three implementations of modified center distance feature (MCDF-2, MCDF-4 and MCDF-8) are used as main features, as well as the CDF for comparison.
- The shape feature and the BULR (bottom-up, left-right density ratio) features are treated as low dimensional feature (LDF) in feature combination method.

5.1 Experimental Results and Facts

This paper uses the combination of several features in experiments, See Table 1. Table 1 gives the recognition results from the experiments that CDF and MCDF features used alone and combined with some low dimensional features including shape feature and BULR feature. The CDF (Center Distance Feature) and MCDF (Modified Center Distance Feature) are participated into the recognition experiments as main feature, because CDF and MCDF features have been showing reasonable recognition rates in former and these experiments.

Table 1. Average Recognition Rate of CDF and MCDF in Feature Combinations (%)

Feature	**CDF-2**	**MCDF-2**	**LDF and CDF-2**	**LDF and MCDF-2**
Average	**78.17%**	**72.29%**	**87.16%**	**98.33%**
Feature	**CDF-4**	**MCDF-4**	**LDF and CDF-4**	**LDF and MCDF-4**
Average	**90.47%**	**87.08%**	**90.28%**	**98.62%**
Feature	**CDF-8**	**MCDF-8**	**LDF and CDF-8**	**LDF and MCDF-8**
Average	**94.50%**	**92.60%**	**94.16%**	**98.77%**

LDF is the abbreviation for the low dimensional features
From the table above, following facts are easily observed:

1. As increasing the number of bins used, the three forms of center distance feature (CDF-2, CDF-4 and CDF-8, center distance feature with two, four and eight bins respectively) show improvement in recognition rates accordingly. The CDF-8 method can describe characters with more detailed information and contributed the highest recognition of 94.50% among the three implementations of CDF in Table 1.
2. The modified center distance feature (MCDF) also comes up with up-going trend as increasing the number of bins taken. The MCDF-8 (modified center distance feature with eight bins) holds the highest point among the three forms of MCDF again, which is 92.60%.

   However, each form of MCDF named after the same Figure (refers the number of bins used) is lower in recognition rate when compared with CDF forms. The more the number of bins applied, the less the gap got between the recognition rates of CDF and MCDF methods.
3. Among the combinations of CDF with low dimensional features, only the CDF-2 assembled feature increased a few points comparing with its alone application, while the other cases of combinations turned backward in recognition rate. We can see that CDF is not very much appealing to associate with the features used.
4. MCDF contributed robust increase in recognition accuracy on each of the three implementations (MCDF-2, MCDF-4, MCDF-8) combined with the same low dimensional features. The implementation of MCDF with more bins holds higher recognition accuracy than the ones with fewer bins. Each of the three combinations uttered remarkable recognition result. We can put the statement that MCDF is very much appealing to combine with the low dimensional features.

5.2 Analysis on the Experimental Facts

Both the CDF and MCDF methods describe characters with structural and statistical information. The feature extraction algorithms for the both methods are similar. An idea with just little modification gifted robust progress in recognition accuracy within combined method. MCDF uses all the divided grids to describe the character shape, while CDF only uses the pen trajectory located grids. CDF records -1 for the no trajectory located grids, so that the grid units without pen trajectory points become useless to distinguish characters. As for the MCDF method, it records useful information (the distance from the gravitational center to the origin) for the no trajectory located grids to distinguish the character. As a result, we receive the following advantages and disadvantages for both methods respectively when they are used alone and in combination. Table 2 gives more specific information about the recognition accuracy of each character. In Table 2, only the eight bins forms of both CDF and MCDF method were used.

1. In the case of alone using:

- As for CDF method, the useless grid units decrease the volume of information to differ characters and more specific details from more bins show higher accuracy;
- According to MCDF method, the no trajectory located units also describe one of the main characteristics of character shape. This makes similar characters become difficult to be differed. For example: ج, چ, خ character group; the group د, ر, ز, ژ and س, ش group; and ف, ق characters group.
- we have to notice that there is one exceptional group including the characters with much and complex special symbols such as ئو, ئۇ, ئۆ, ئۈ (25~28, Table 2). In reference [17], experiments show that eight bins form of CDF is not very much favorable for these characters. Too much bins easily spoil the special symbols of characters and make

backwards in recognition rate. However, the no trajectory grid units which record one of the main characteristics of character shape according to the MCDF method affect to recognize characters with positive result.

2. In another case : when CDF and MCDF methods combined with other features such as shape feature which is remarkable for the special symbols of characters, we can see some different situations:

- CDF cannot use the potential of no trajectory located grid units. Therefore, CDF cannot provide much environment or space for the effective features to play their functions to differ the similar shaped characters. So the combination of CDF with other methods shows only a little progress compared with the alone using.
- As for the MCDF, the no trajectory located grid units make the similar characters even similar in feature domain, but put them far from the different groups. These results in those effective features for special symbols or characteristics can easily distinguish the characters within the group which consisted of similar characters in shape.
- As MCDF mainly aims to provide better environment for the special symbols of characters. The simplest characters with the least special symbols such ژ ز د and slightly lowered in recognition rate by MCDF combined method comparing with CDF combination with low dimensional features.

Table 2. Recognition Rates of Center Distance Feature with Different Number of Bins (%)

Features / Uyghur Characters		CDF-8	MCDF-8	LDF and CDF-8	LDF and MCDF-8
1.	ئا	97.50%	91.67%	96.67%	100.00%
2.	ئە	90.83%	95.83%	91.67%	100.00%
3.	ب	87.50%	91.67%	89.17%	94.17%
4.	پ	95.00%	98.33%	91.67%	100.00%
5.	ت	91.67%	94.17%	95.83%	97.50%
6.	ج	94.17%	85.00%	94.17%	95.00%
7.	چ	96.67%	91.67%	93.33%	100.00%
8.	خ	90.00%	96.67%	93.33%	97.50%
9.	د	95.83%	84.17%	97.50%	93.33%
10.	ر	97.50%	89.17%	96.67%	97.50%
11.	ز	100.00%	90.00%	97.50%	96.67%
12.	ژ	99.17%	91.67%	97.50%	99.17%
13.	س	95.00%	88.33%	96.67%	99.17%
14.	ش	95.83%	93.33%	98.33%	99.17%
15.	غ	97.50%	96.67%	92.50%	100.00%
16.	ف	95.00%	90.00%	92.50%	100.00%
17.	ق	94.17%	97.50%	86.67%	100.00%
18.	ك	95.00%	94.17%	94.17%	98.33%
19.	گ	95.83%	95.83%	96.67%	99.17%
20.	ڭ	98.33%	95.83%	96.67%	100.00%
21.	ل	99.17%	95.00%	99.17%	100.00%
22.	م	99.17%	96.67%	95.83%	100.00%
23.	ن	95.83%	98.33%	94.17%	99.17%

Features / Uyghur Characters		CDF-8	MCDF-8	LDF and CDF-8	LDF and MCDF-8
24.	ھ	**93.33%**	**93.33%**	**93.33%**	**99.17%**
25.	ئو	**94.17%**	**96.67%**	**97.50%**	**100.00%**
26.	ئۇ	**92.50%**	**94.17%**	**93.33%**	**99.17%**
27.	ئۆ	**81.67%**	**94.17%**	**82.50%**	**100.00%**
28.	ئۈ	**89.17%**	**97.50%**	**93.33%**	**100.00%**
29.	ۋ	**98.33%**	**97.50%**	**98.33%**	**99.17%**
30.	ئې	**94.17%**	**92.50%**	**94.17%**	**98.33%**
31.	ئى	**93.33%**	**65.00%**	**100.00%**	**100.00%**
32.	ي	**90.83%**	**90.83%**	**82.50%**	**99.17%**
Average		**94.50%**	**92.60%**	**94.16%**	**98.77%**

6. Conclusions

Thanks to the idea of making the similar ones even nearer in feature domain and putting them far from the other groups, so that the effective features for special symbols which are critical to differ the similar characters find themselves free to play their functions. Although MCDF declines in recognition accuracy when it is applied without partners comparing with CDF; MCDF gifts much benefit to differ the similar characters when some other effective features for special symbols come together. Obviously, the combination of MCDF with the low dimensional features used in experiments makes a great contribution to improve the recognition accuracy. However, the beginning form, middle form and ending form of Uyghur characters are still waiting for an intensive research.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (61263038 and 61462081) and Program for new century Excellent Talents of the ministry of education (NCET-10-0969).

References

[1] S. Theodoridis and K. Koutroumbas, Pattern Recognition, fourth ed., Academic press, USA, **(2009)**.
[2] N. Arica and F. T. Yarman-Vural, Student Member, IEEE Senior Member, IEEE, An Overview Of Character Recognition Focused On Off-line Handwriting，C99-06-C-203, **(1999),** pp. 1-22.
[3] W. Shangqing and Z. Fenghao, "Off-line handwriting Chinese character recognition based on Bayesian grid", Computer-Aided Engineering, vol. 15, no. 3, **(2006),** pp. 72-74.
[4] K. B. M. R. Btuwita and G. M. D. C. Bandara, "An online adaptable fuzzy system for off-line handwritten character recognition", Proceedings of the Eleventh International Fuzzy Systems Association World Congress, **(2005).**
[5] M. Ammar, R. A. Majed, "Online handwritten recognition for the Arabic letter set", World Scientific and Engineering Academy and Society (WSEAS), vol. 11, no. 5, **(2011),** pp. 42–49.
[6] M. Ibrayim, Z. Heng, L. Cheng-Lin and A. Hamdulla, "A Method for Online Handwritten Uyghur Character Recognition", Pattern Recognition and Artificial Intelligence, vol. 25, no. 6, **(2012)** December 3, pp. 979-986.
[7] C. Mohamed, K. Nawwaf, "Character Recognition Systems", A Guide for Students and Practitioners, New York, JOHN WILEY & SONS, **(2007).**
[8] Z. Meng and Y. Zhongqiu, "Image preprocessing research in Handwriting numeral recognition", Micro-Computer Information, vol. 22, no. 6, **(2006),** pp. 256-258.

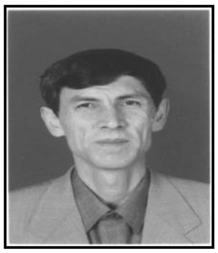
[9] H. Linfeng and Z. Hui, "On-line handwriting Uyghur character recognition based on support vector machine (SVM)", Computer Applications and Software, vol. 36, no. 3, **(2012),** pp. 179-182.
[10] I. Mayire, "Research on the technology of online handwritten Uyghur characters recognition", M.S. thesis, Xinjiang University, Urumqi, China, **(2011).**
[11] S. Mutallip, "Research on the Technology of Trainable Online Uyghur Handwritten Character Recognition", M. S. thesis, Dalian Maritime University, Dalian, China, **(2010).**
[12] A. Yasin, "Preprocessing and feature extraction of on-line Uyghur handwriting recognition", M.S. thesis, Xinjiang University, Urumqi, China, **(2010)**.
[13] Ranagul, "Research on the key technologies of online handwritten Uyghur word recognition", M.S. thesis, Xinjiang University, Urumqi, China, **(2011)**.
[14] I. Mayire and M. Kamil, "Multi-classifier combination scheme for recognition of online handwritten Uyghur characters", Computer Engineering and Applications, vol. 48, no. 31, **(2012),** pp. 140-145.
[15] Piguilin, "HMM based recognition method research on online handwriting Uyghur word", M.S. thesis, Xinjiang University, Urumqi, China, **(2012)**.
[16] K. Zulpiya, "Research on Online Uyghur Handwritten character recognition based on Feature combination", M.S.thesis, Xinjiang University, Urumqi, China, **(2013)**.
[17] W. Simayi, M. Ibrayim, D. Tursun and A. Hamdulla, "Research on On-line Uyghur Character Recognition Technology Based on Center Distance Feature", Proceedings of 13th IEEE International Symposium on Signal Processing and Information Technology (ISSPIT 2013), SP-6046, **(2013)**.
[18] V. D. Trier, A. K. Jain and T. Taxt, "Feature Extraction Methods For Character Recognition—A Survey", Pattern Recognition, vol. 29, no. 4, **(1996),** pp. 641-662.
[19] C. C. Tappert, C. Y. Suen and T. Wakahara, "The state of the art in on-line handwriting recognition", IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 12, no. 8, **(1990),** pp. 787-808.
[20] A. T. Al-Taani, "Recognition of On-line Arabic Handwritten Characters Using Structural Features", Journal of Pattern Recognition Research, **(2010),** pp. 23-37.
[21] Y. Junchuan, T. Yu and H. Zhisong, "Character Feature Extraction Method based on the BP Neural Network", Proceedings of 2011 3rd IEEE International Conference on Information Management and Engineering, ICIME, vol. 04, **(2011)**.

Authors

Askar Hamdulla, he received B.E. in 1996, M.E. in 1999, and Ph.D. in 2003, all in Information Science and Engineering, from University of Electronic Science and Technology of China. In 2010, he was a visiting scholar at Center for Signal and Image Processing, Georgia Institute of Technology, GA, USA, tutored by Professor Biing-Hwang (Fred) Juang. Currently, he is a professor in the School of Software Engineering, Xinjiang University. He has published more than 120 technical papers on speech synthesis, natural language processing and image processing. He is a senior member of CCF and an affiliate member of IEEE.

Wujiahemaiti Simayi, he received his B.E. and M.S. degree in electronics, signal and information processing from Xinjiang University, China, in 2009 and 2014, respectively. His scientific interest includes pattern recognition. Currently, he is a research assistant at the Key Laboratory of Intelligent Information Processing, Xinjiang University, China. His research interests include feature extraction and selection techniques, handwritten character recognition

Mayire Ibrayim, she received her M.S. and Ph.D. degree in communication engineering, communication and information system from Xinjiang University and Wuhan University, China, in 2009 and 2013, respectively. Since 2013, he has been working as a teacher in School of Information Science and Engineering, Xinjiang University, and now she is doing her postdoctoral research work in a workstation of Computer science and Technology of Xinjiang University. Her research interests include image processing and pattern recognition.



Dilmurat Tursun, he received his B.E. in 1983 in Electrical Engineering, from Xinjiang University of China. Currently, he is a professor in the Institute of Information Science and Engineering of Xinjiang University. He has published more than 50 technical papers on experimental phonetics and image processing.

