

Survey on the Features for Recognition of on-line Handwritten Uyghur Characters

Wujiahemaiti Simayi¹, Mayire Ibrayim¹, Dilmurat Tursun¹, Askar Hamdulla²

¹*Institute of Information Science and Engineering, Xinjiang University, China
hojaahmad2014@163.com*

²*School of Software, Xinjiang University, Urumqi, China 830046,
Corresponding Author: askarhamdulla@gmail.com*

Abstract

This survey attempts to provide a detailed overview of the previous work on Uyghur on-line handwritten character recognition technology and mainly focuses on investigating the feature extraction algorithms that have been presented or applied to Uyghur characters. Statistical, structural and combined-hybrid approaches for feature extraction technology for Uyghur characters are going to be explored. Besides the popular feature extraction methods which perhaps are granted to be taken into consideration for all letter kinds, the study on Uyghur handwritten recognition contributed new features and extraction algorithms for pattern recognition. Summarization on observing methods and extracted features from the published work and suggestions for further attempts would be valuable reference for the up-coming research activities on finding more efficient features.

Keywords: *Online handwriting recognition, Uyghur characters, Feature Extraction*

1. Background

A major goal of pattern recognition research is to recreate human perception capabilities in artificial systems. As a special aspect of visual perception, handwriting recognition is one of the main branches of pattern recognition technologies with major implications in automation and information handling. [1]

Handwriting recognition (HWR) is the ability of a computer to receive and interpret intelligible handwritten input from sources such as paper documents, photographs, touch-screens and other devices [2]. Handwriting recognition has two main directions in research and application, the off-line and on-line handwritten recognition. While the former one attempts to acquire the information of already formed handwritten or printed letters on paper using scanner or camera, the latter one focuses on recognizing the character forms which is being shaped on the pressure sensitive device's screen due to the pen or hand's touching on it. Both off-line and on-line recognition techniques can be applied to handwritten characters.

The natural and comfortable style in writing due to the online handwritten recognition technology effectively reduces difficulties of using computers for beginners, therefore, is of great social need and economic potential in global market. Advancement of modern electronics let handwriting recognition applicable in smaller digital tools such as handheld PDAs, tablet computers and service terminals, etc, for simple users and entrepreneurs as well. Data inputting methods without keyboards or key-pads are receiving more interest in both academic and commercial research communities. It is apparently showed that handwriting input is the most acceptable and welcomed input style, especially for those less-educated and minority ethnic groups who are not aware of using advanced devices in work, communication and everyday life.

The emerging of the first commercial products with handwriting recognition capability in 1980s made many scientists and people from different nations very much encouraged to start relative research on their own languages or letters. In China, much investment and effort has been invested on Chinese simplified and Chinese Complicated character forms and successful software modules have been available for public using such as Hanwang Pen [3]. As the economic and technological development and increasing needs of the people from various ethnic groups in western regions of China, the handwriting recognition for minority ethnic groups are becoming focus to accelerate the informatization process of these areas and people there. Uyghur is one of the main spoken languages in western China and Central Asia, furthermore, is anthropologically and linguistically connected to many ethnic groups not only in Central and Western Asia, but also even in Eastern Europe [4], which has been reflecting the need of building a complete handwriting recognition system with maximum reliability. The research on Uyghur Handwriting recognition technology and its applications can be important reference and sometimes strait technology base for the handwriting recognition of other related scripts. The reasonable ideas and effective methodologies from Uyghur Handwriting recognition will be the contribution for many application areas of pattern recognition, as well.

The remainder of this article is organized as follows. In the following section, some information for Uyghur characters and its handwriting properties are provided. Section 3 gives a brief introduction for the recognition units of on-line handwritten recognition system for Uyghur characters. Section 4 projects investigation on feature extraction method applied or presented for the recognition of Uyghur characters. Summarization on previous work is given in section 5. Then section 6 concludes the manuscript.


2. Handwriting Properties of Uyghur Characters

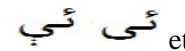
Dozens of kinds of scripts has been used in the long history of Uyghur civilization. The modern Uyghur letters in practice today are some vise adaptation of Arabic and Farsi characters into Uyghur language. Uyghur language belongs to Turkish family of Altaic language system, while Arabic and Farsi are the main languages in Semit-Hemit and Idian-Europe language system. Uyghur is alphabetic and fundamentally different from Chinese and English both in writing styles and linguistic aspects. There are some similar character forms which are being shared with Arabic and Farsi, however, quite differences form them are easily found. Followings are some attributes to take into consideration for successful recognition of handwritten samples of Uyhgur letters as well as the printed ones.

1. Uyghur is alphabetic and has 32 basic characters with total 128 writing forms. See Table 1.

1. Uyghur is written from right to left, and from top to bottom.
2. The characters in one word are connected using different forms and almost each character has its isolated, beginning, intermediate (or middle), ending forms, with the exception of some characters which only have the two or three forms mentioned above and some have eight writing forms. For example,

Characters  etc. only have two forms;

Characters  have four standard forms;

Characters  etc have eight writing forms. See Table 1.

3. Almost all characters are composed of its main body part and additional parts which provide differentiation from similar characters or character forms. Usually, the main part is formed first and the additional parts are put over, under or beneath the character. The numbers and forms of additional parts of characters are different from each other. For example,

Characters ر س د have not any additional parts;
Characters ن خ ه have additional parts over the main part; Characters
چ ك ا have additional parts located beneath the main part of character;
Characters پ ئ ي hold additional parts under their main part, etc.

4. Some characters share the same main part and can be grouped into some special subdivision. There a number of character groups which is differed only by the additional part. For example:

- 1) ا ء
- 2) ر ز ر
- 3) خ چ ج
- 4) ءو ءو ءو ءو
- 5) ئ ي
- 6) ب پ ت
- 7) س ش
- 8) ك ك

etc. are the some similar characters groups.

2. Some additional parts are always written in other forms due to the cursive characteristic of natural handwriting.
3. In actual handwriting, many of the standard writing rules are always invaded and characters are written in personal ways of the writers.

Table 1. Uyghur Alphabet

Character No.	Isolated Form	Ending Form	Intermediate Form	Beginning Form
1	ئا، ا			ئا، ئا
2	ئە، ە			ئە، ئە
3	ب	ب	ب	بب
4	پ	پ	پ	پپ
5	ت	ت	ت	تتا
6	ج	ج	ج	جج
7	چ	چ	چ	چچ
8	خ	خ	خ	خخ
9	د			د
10	ر			ر
11	ز			ز
12	ژ			ژ
13	س	س	س	سس
14	ش	ش	ش	شش
15	غ	غ	غ	غغ
16	ف	ف	ف	فف
17	ق	ق	ق	قق
18	ك	ك	ك	كك
19	گ	گ	گ	گگ
20	گ	گ	گ	گگ
21	ل	ل	ل	لل
22	م	م	م	مم
23	ن	ن	ن	نن
24	ھ	ھ	ھ	ھھ
25	ئو، و			ئو، ئو
26	ئۇ، ۇ			ئۇ، ئۇ
27	ئۆ، ۆ			ئۆ، ئۆ
28	ئۈ، ۈ			ئۈ، ئۈ
29	ۇ			ۇ
30	ئى، ى	ئى، ى	ئى، ى	ئى، ى
31	ئى، ى	ئى، ى	ئى، ى	ئى، ى
32	ي	ي	ي	يي

3. Introduction to On-line Handwritten Recognition Framework

This section gives a simple introduction for the main function units of On-line handwritten character recognition system, including preprocessing on the acquired first-hand sample data, feature extraction, classification and training, then testing on untrained samples. The function units are interdisciplinary that each unit has direct influence on the latter units successful performance.

Pre-processing the writing pad records the pen's trajectory on the screen, usually using the two dimensional (x,y) coordinates of the pen's touching location and the coordinates are recorded corresponding to the time series spent to write it. The noise attributed to the device resolution and hand fluctuation writing process should be treated first. The noisy sequence in addition to the information sequence may severely affect the on-line data sequence. Furthermore, the first-hand samples of various writers are quite different in size, distortion angle and location on screen that normalization process cannot be neglected, as well.

Feature Extraction being the most critical factor to achieve high recognition performance in all pattern recognition, feature extraction is very important in handwritten recognition technology from the beginning. The objective of feature extraction is to capture the essential characteristics of the symbols. It produces the feature vector that will be characterized and evaluated to determine the character at last. In order to focus on the essentials of handwriting, deriving reasonable features is mandatory. Elegant feature selection can greatly decrease the workload and improve the recognition accuracy.

Classification

Classification is carried out on the basis of stored features in the feature space. Characters with similar characteristics (or feature values) are categorized into subgroups and further single characters are determined according to the classification rule. Many different classification strategies could be used to recognize handwritten characters. The classifier perhaps will be the popular ones such as Neural Network, Support Vector Machine, K-Nearest Neighbors, Bayesian Classification, and Decision Tree Classification, also can be matched ones for task context and even perhaps the wise combination of more classifiers. In most character identification has training and testing blocks which use same classification rule.

Training Some portion of the collected data are used to build model feature vector and establish the model library as decision base to evaluate the similarity of new input to the character which the model vector stands by.

Testing The feature vector of each testing character forms are compared with model feature vector and calculate the similarities according to the chosen rule. The most similar character in model data set is confirmed for the testing character.

4. Previous Work on Feature Extraction Techniques

Feature extraction is one of the essential steps of character recognition process affecting the overall accuracy of the recognition system. Extraction of features for achieving maximum accuracy and performance with minimum error rate is a difficult task from the beginning. This paper gives a detailed overview of different feature extraction for recognition process of Uyghur handwritten recognition technology.

Features for online handwriting recognition can broadly be classified into the following categories [9, 24, 30-31]:

Statistical features,

Statistical feature based recognition is derived from the probability and statistics theories. With strict mathematical base, this method is showing its reliability in pattern recognition. Statistical feature based recognition usually doesn't care about the basic pattern units. Instead, the global feature of pattern is observed with statistical computation, and the features with good classification performance and stable statistical characteristics are selected as feature vector to represent the pattern. Statistical feature based recognition approach is favored with its good resistibility to noise, simplicity in classification and matching algorithm, therefore, can be easily implemented in real tasks. However, the weakness on observing specific information of pattern and differing similar characters are the downside of this method.

Structural features

Structural features based recognition focuses on processing the structural information of recognition object, has reached substantial references and development. In structural feature based recognition, a pattern is observed as one kind of different organization of basic pattern units or pattern sub-divisions. For example, a character is divided into smaller character parts or even the basic units, such as strokes. The basic units are recognized using template matching algorithm and at last the original complicated character form is indentified. So, from the point of pattern recognition, structural feature can express the essential characteristics of character form. Structural feature based recognition has the advantages of good adaptability to character font change, strong differentiation similar characters. However, it is weak to resist the interference.

Hybrid-Combination approaches

Combination of statistical and structural feature aims at maximally benefiting from both kinds. Structural features can deliver the most distinguishing properties of pattern or character form, but with less resistibility to noise. Although, Statistical features always overlook the specific information of character form, but provide its good guard against interference. It is seen that combination can benefit the strengths of both structural and statistical feature based recognition, and simultaneously prevent the short-sides of them.

Some common features are always used in feature selection in handwritten recognition technology. Such as beginning and ending of character of stroke, number of strokes, time order of strokes, existence and number of loops and curves, number of intersections, number of free ends, horizontal and vertical symmetry etc. are attained from the basic sequential characteristics of character sequence. Mathematical observation and calculation always results in statistical features which are hidden from human-mind. Directional feature is one of the most common and visible one in statistical feature selection approaches. However, many statistic characteristics which are hidden from human mind are reached only from the repeated mathematical calculation. For example, bottom-up, left-right density ratio, width-height ratio etc. It is believed that every language has special characteristics in speaking as well as in writing. Both the local and global features are variably used from one application to another. It is apparent that higher the dimensions used in a feature selection, greater the time complexity. But, higher the dimensions put in in feature vector, greater the confirmation stability and recognition accuracy usually is.

Limited work has been done on Uyghur handwritten recognition technology, the researchers reached preliminary achievements. In spite of late beginning and lack of research staff and standard data-base for Uyghur on-line handwritten recognition, the researchers made their best effort to acquire adequate sample data and establish reliable research platform.

In 2003, some preliminary preparations for Uyghur handwritten input methods were made by Muhtar etc. [5]. In their basic research, they had observed and analyzed some online handwritten character recognition methods. They recommend BP Neural Networks will be a promising method for this task. They discussed the difficulties of handwritten input method for Uyghur characters specifically and put some forward suggestions for later work. They wished for a adequate data library with enough volume [6].

Yusup.etc. in 2006, developed a handwritten input system for Uyghur characters including different variations according to position within words. Their work had recognized some characters by 90% accuracy. The distance between neighbor strokes was concerned for feature extraction method. [7].

Cao Zhi-hong, Li Yuan [8, 10], Wan Feng [9], in 2007~2009, conducted the exploration to the characteristics and recognition process on Uyghur characters. They introduced the preliminary technology and developed the experimental character recognition system using Visual Basic 6.0 and Visual Studio C++ 6.0 software environment respectively. The Stroke four direction features, point information characteristic code, stroke number feature, connectivity feature, penetrating characteristics

were applied to build the model library. Minimum distance classifier was implemented to put the characters into categories. This experimental system has carried out successfully; recognition rate was higher than 86% or 90%, to the Uyghur handwriting characters.

Ren Hong-yu, Yuan Bao-she, Tian Yuan [11] observed the handwriting properties of Uyghur characters from structural and statistical aspects. The combination of structural characteristics with statistical information such as pen's trend information granted a gladly and stable recognition rate. They used stroke number, main and additional strokes position relationship, loop on main stroke features to describe the structural characteristics, and took the information of main stroke beginning position, main stroke ending position, pen's trend after main stroke began, pen's trend after main stroke end, number of intersection on horizontal axis. Number of pen's change on vertical direction, Number of pen's change on vertical direction into consideration for statistical features. BP Neural Networks carried out good performance and stable recognition rate with the extracted features.

Mutallip (2010) used the gradient directional feature method for feature extraction of on-line handwritten Uyghur characters [12]. Support vector machine (SVM) for classification was proved helpful that as training data volume increased, the recognition rate for untrained input also improved drastically in his experiment.

Alimjan (2010) selected some commonly used features including direction feature, loop and angle features to observe the characteristics of Uyghur characters [13]. Yuan Baoshe, etc. put 21 kinds of features in feature vector to build the Uyghur written language character recognition data sample template storehouse[14].

Ranagul (2011) discussed the grid directional feature for Uyghur characters. The characters are investigated using eight directions within each divided grid unit [16]. Tan Fuxiu [17] conducted some research feature extraction, which contains structural features of stroke and meshing direction feature. By the analysis of the experimental results, the reason that results low recognizable rate of the structural features of stroke is contributed to the cursive writing of additional parts(points , strokes) and reluctant writing of main part(main stroke). Han Linfeng, etc applied the invariant moment feature in combination method [18].

Piguilin (2012) propose a delay stroke processing method which contains seeking and projection of delay strokes for Uyghur characters [19].

Another SVM based character recognition experiment reached good results by Han Lin-feng, especially for multi-stroke characters using some common structural features and directional features on Uyghur isolated character forms [20].

In 2013, Jiang Wen, etc used directional element features with KNN classifier for Uyghur character recognition technology. Although, the overall recognition rate was not high enough, the top three or five candidates showed remarkable accuracy [22].

In 2013, Jiang Wen etc. Applied the Gabor filter to extract the real value energy feature of Uyghur handwritten characters. In this method, character image is divided into small pieces at first, and the energy of each piece is extracted. The formed real value energy matrix is inconvenient to handle, so the matrix dimension is reduced and put as feature data. KNN classifier is utilized for training and testing [23].

In 2011, Mayire analyzed the characters with equal sampling feature [15]. 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.

Mayire conducted a comparative research on the efficiencies of some popular techniques of normalization, feature extraction and classification. Specifically, eight normalization techniques and the normalization cooperated feature extraction (NCFE) method with different settings were used. Four classifiers were used for classification including the modified quadratic discriminant function (MQDF), the discriminative learning quadratic discriminant function (DLQDF), the learning vector quantization

(LVQ) classifier, and the support vector classifier with RBF kernel (SVC-rbf). Furthermore, geometric features which characterize the spatial context in handwritten documents were extracted for enhancing the recognition performance. In experiments on 38 400 test samples of 128 classes of Uyghur character forms, the proposed approach in the paper achieved an accuracy of 89.08% [21].

In 2013, Mayire formulated her work on the recognition of handwritten Uyghur characters and words into her PhD thesis. New approaches for character segmentation is presented and reached some pleasant results. According to her adopted two level character segmentation scheme, the word is segmented into conjoined section firstly, and then the conjoined section is cut into characters in the next step. For the difficulties in character segmentation and recognition which are mainly attributed to the cursive writing of Uyghur characters, an online recognition approaches based on a lexicon-driven, integrated segmentation and recognition have been were presented. Word recognition problem is transformed into matching optimization problems between the dictionary entry and the handwritten word image. Dynamic matching between characters in the lexicon entry and segment(s) of the input word image is used to rank the lexicon entries in order to get the best match [24].

Zulpiya (2013) discussed the peripheral contour feature and presented the shape feature for on-line Uyghur handwritten character recognition [25]. 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 [26]. 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.

In 2014, the Center Distance Feature (CDF) has been developed 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 [27-29].

5. Summarize and Suggestion

The practical outcome of this section is a set of guidelines and hints that, at least to the authors' minds, should be considered for future research and development in the field.

Handwriting recognition is notoriously a difficult task. Nevertheless, the work has been done on this technology is fruitful to some extent. Handwritten recognition technology for Uyghur characters is still very young not only in age, but also in maturing. Still many question marks and difficulties waiting for more intense research and creativeness.

5.1 Major Problems in Recognition of On-line Handwritten Uyghur Characters

Although online handwriting recognition is wished very much by people who are already accustomed to use pens, the accuracy problem of online recognition systems still

putting difficulties for widespread application of handwriting recognition technology. One of the main difficulties in on-line character recognition is, characters may be misclassified because of different sequential representation. This is not only the problem of Uyghur but also belongs to other scripts as well. Main problems associated with on-line handwritten character recognitions are classified into following categories:

• **Cursive Writing**

This is a common difficulty for many kinds of handwritten scripts. Uyghur is also very difficult to know the Numbers of strokes and distinguish the kinds of additional parts, such as the strokes' order, shapes and sizes, directions, skew angle etc. Additional parts in Uyghur formed by dot or dots, dash, special curve or angled small line including:



The special additional parts are always written variously from one to another. In addition, writer uses only one stroke to complete a character In cursive mode of writing, while more strokes should be used in original writing rule.

• **Similarity in Drawings among Many Classes of Characters**

Some Uyghur characters are in structure, especially in the forming of main part. In addition, writer may draw different characters in a similar way with the use of same number of strokes and in same order, which can be the most difficult part to classify correctly. So, it is very hard to analyze their dissimilarity from one another in case of natural handwriting. It is also no guarantee that same one always writes in a same way.

- Strokes may vary in direction in different peoples' styles.
- The length of a specific stroke is variable.
- The size (big or small) of a specific symbol is variable.
- The speed in writing is variable

• **Digitizer Resolution and precision**

The digitizer system resolution and precision is also one important factor that affects the obtaining the handwritten character forms adequately.

5.2 Review and Suggestion on Feature Extraction

Many an experiments has been conducted and some useful feature extraction algorithms have been proposed or applied to Uyghur handwritten characters successfully.

Table 2. Applied Features in Previous References

Ref. No.	First applied or suggested features	Classifier	Recognition rate	Feature type
5, 6	Character stroke length etc.	BP	Not Provided	Statistical
7	Distance between neighbor strokes etc.		90% for some characters	Statistical
8, 10	Direction of main stroke, Stroke number, stroke point information, connectivity feature, penetrating feature etc.	Minimum Distance Classifier	86%	Structural &Statistical
9,	Direction of main stroke , stroke number, stroke point number, stroke point position information, shape of couple points, position of beginning and ending point , direction change on additional stroke etc	<i>Piecewise Linear Classifier</i>	90%	Structural &Statistical
11	Structure feature—stroke number, main and additional strokes position relationship., loop on main stroke Stroke Trend feature ---- main stroke beginning and ending position, pen's trend after main stroke began or end , number of intersection on horizontal axis, number of pen's change on horizontal or vertical direction etc.	BP	>96.46%	Structural &Statistical
12	Gradient directional feature etc.	SVM	90%~96%	Structural &

				Statistical
13	Eight-directional feature, angle, loop, stroke delay, point properties etc.	HMM	Not Provided	Structural & Statistical
14	position of additional stroke, straight line, quadrant position of beginning and ending point, side of beginning and ending point, pen's trend on beginning and ending point, number of direction change on X direction or Y direction, right-left-right direction change on X direction, left-right-left direction change on X direction, Up-Down direction change on Y direction, Down-Up direction change on Y direction, Down-up-down direction change on Y direction, the maximum crossing number on X direction, width-height ratio, Loop and intersection on main stroke, stroke number, stroke length etc.	XOR calculation	97%	Structural & Statistical
16	Stroke number, main and additional stroke position relationship, Eight grid directional feature etc.	Minimum Distance Classifier	60%	Structural & Statistical
17	Direction on main stroke, direction change on additional stroke, stroke number, stroke point number, stroke point position, shape of coupled points , grid directional element feature etc.	Minimum Distance Classifier(Euclidean)	75.26%	Structural & Statistical
18	features of Invariant Moment , Projection, directional feature, Loop, Intersection etc	HMM	Not provided	Structural & Statistical
19	Stroke delay , Loop, Directional feature etc	HMM, Logistic Tree	93.71%	Structural & Statistical
20	Stroke number, Diagonal slope, Axis intersection with strokes, pen's trend change, beginning and ending position difference, additional stroke position, four and eight directional feature etc	SVM	93%	Structural & Statistical
22	Eight directional feature, Directional Element trajectory feature etc	KNN	94.94% for top five candidates	Statistical
23	Gabor energy value etc	KNN	94.11% for top five candidates	Statistical
15, 21, 24	NCFE with eight direction, width, height, length of diagonal, width-height ratio, character center position etc	DLQDF	89.08%	Structural
15, 21, 24	Uniform sampling feature , eight direction feature, grid direction , grid density , two directional projection etc	Nearest Neighbor with DTW	75.06 %	Statistical
25	Stroke number, main and additional stroke position relationship, shape feature for additional stroke, bottom-up, left-right density ratio, peripheral contour feature etc	Minimum Distance Classifier	75.7%	Structural & Statistical
26, 29	Stroke number, main and additional stroke position relationship, , shape feature for additional stroke, bottom-up, left-right density ratio, Center Distance feature	Minimum Distance Classifier	94.50%	Structural & Statistical
27,28, 29	Stroke number, main and additional stroke position relationship, shape feature for additional stroke, bottom-up, left-right density ratio, Modified Center Distance feature	Minimum Distance Classifier	98.77% for Isolated character forms	Structural & Statistical

There are some summarizations over the feature extraction and classification methods presented by the past work.

- Structural features are always first taken into consideration by many researchers. Such as stroke number feature, lope feature, slope feature, distance between the beginning and ending point of stroke, connectivity feature, additional part's shape feature. The structural features are hoped to acquire the basic information of character shape and be the first decision base for preliminary classification in many ways.

- Statistical information is also obtained by observing the pen trajectory coordinates and their distribution on normalized lattices. Some calculations are also made to find new attributes of pen-tip movement. Equal sampling feature, grid directional feature with four or eight directions, direction code features, bottom-up and left-right density ratio, additional parts location feature played important roles to deliver more information about the construction of character form.

- The combination statistical and structural feature showed a great advantage to get more comprehensive information about the character shape and improve the recognition

accuracy, especially for some complicated characters with more strokes or additional parts.

The equal sampling feature, peripheral contour feature, center distance feature and its modified form (MCDF) showed great appealing to combine with other low dimensional features and gifted much increase in recognition rate. The main reason for the advantage of combination method perhaps is that the extracted high dimensional features can express the structural and statistical information of character by themselves to some degree, also provide a space for low dimensional features to play their role.

- Deep learning approach is also a very promising method for feature extraction and counted as the state-of-art in this field. New feature extraction algorithms with deep learning approach in urgent need to build flexible model data library which can adapt itself automatically corresponding to writers. Deep learning approach can be a wise solution to the diversity of handwriting styles and writers in spite of its greater time consuming.

To sum up, the Combination method of statistical and structural features can be considered the most successful approach to date for online handwritten character recognition. The time spent on it will be longer than other simple, rule based feature extraction methods still take advantages in time costing. Providing some handy algorithms to train on character features will save the processing time and increase its accuracy. Comparability of results is only possible when a standard widely admitted dataset is available for research community.

6. Conclusions and Future Direction

This paper reviews the past work on Uyghur handwriting recognition technology, listed some algorithms or methods used or proposed for feature extraction and classification tasks. Standard and reliable sample is mandatory to get a reliable experiment result. Structural, statistical and combined methods are explored on feature extraction. Some common difficulties for accurate recognition of Uyghur characters are categorized. Combination method for feature extraction is advised for it improved the system recognition confidence and efficacy. Deep Learning approach with quick algorithm is hoped as it is the state-of-art technique in this field. More intense and detailed research on Uyghur handwritten technology is still in great demand.

Acknowledgements

This work is supported by Natural Science Foundation of China (No. 61263038 and No. 6146208) and Natural Science Foundation of Xinjiang (No.2015211C285) and PHD Start Foundation of Xinjiang University (No.BS130117).

References

- [1] S. Theodoridis and K. Koutroumbas, "Pattern Recognition", Fourth ed., Academic press, USA, (2009).
- [2] wikipedia-http://en.wikipedia.org/wiki/Handwriting_recognition
- [3] W. Zhang Guo, "Science & Technology Industry of China", vol. 92, no. 2, (1997), pp. 45-46.
- [4] K. Mahmut, "Modern Uyghur Language", Xinjiang People's Press, (2003).
- [5] M. Mamut and M. Shemshidin, "Bais Research on Uyghur Handwriting IM", Computer Engineering and Applications, vol. 29, (2003).
- [6] R. Muhtar, M. Mutallip and Shemshidin, "Analysis on Difficulties on Uyghur Handwriting IM", the 15th China National Conference on Computer Science and Technology Applications, (2003), YinChuan, China, September 23-24, pp. 108-112.
- [7] Y. Abaydul and M. Mosajan, "The Research Entering Electronic Text and Written Style Editor of Uyghur Language", The Proceedings of 2nd Chinese Conference on Human-Computer Interaction (CHCI'06), (2006), pp. 66-71.
- [8] Z.-h. Cao, "Research and Realization of On-line Uyghur Handwriting System", Ms thesis, Xinjiang University, (2007).

- [9] W. Fang, "Research and Implementation of the On-line Uyghur Handwriting Recognition Technology", Ms thesis, Xinjiang University, (2007).
- [10] Z.-h. Cao and Y. Li, "Research and Realization of Feature Extraction on On-line Uyghur Handwritten Characters", *Technological Trend*, vol. 14, (2009).
- [11] H.-y. Ren, B.-s. Yuan and Y. Tian, "On-line & Handwritten Uyghur Character Recognition Based on BP Neural Network", *Microelectronics & Computer*, vol. 27, no. 8, August (2010), pp. 239-241.
- [12] S. Mutallip, "Research on the Technology of Trainable Online Uyghur Handwritten Character Recognition", M. S. thesis, Dalian Maritime University, Dalian, China, (2010).
- [13] A. Yasin, "Preprocessing and Feature Extraction of On-line Uyghur Handwriting Recognition", M.S. thesis, Xinjiang University, Urumqi, China, (2010).
- [14] B.-s. Yuan and S. Wshour, "Handwriting Recognition Algorithm of Uyghur Letter", *Computer Engineering*, vol. 36, no. 2, (2010).
- [15] I. Mayire, "Research on the Technology of Online Handwritten Uyghur Characters Recognition", M.S. thesis, Xinjiang University, Urumqi, China, (2011).
- [16] Ranagul, "Research on the Key technologies of Online Handwritten Uyghur Word Recognition", M.S. thesis, Xinjiang University, Urumqi, China, (2011).
- [17] F. Tan, "On-line Uyghur Handwritten Character Recognition Based on Mobile Platform", Ms thesis, Xidian University, (2011).
- [18] L.-f. Han and H. Zhao, "The Method of Preprocessing and Feature Extraction for Online Handwriting Uyghur Characters", *Computer Knowledge and Technology*, vol. 7, no. 7, (2011), pp. 1607-1609.
- [19] Piguilin, "HMM Based Recognition Method Research on Online Handwriting Uyghur Word", M.S. thesis, Xinjiang University, Urumqi, China, (2012).
- [20] L.-f. Han and H. Zhao, "SVM Based Recognition Algorithm of Online Uyghur Handwritten Characters", *Computer Applications and Software*, vol. 29, no. 3, (2012) March, pp. 179-182.
- [21] I. Mayire, H. Zhang, C.-l. Liu and A. Hamdulla, "A Method for Online Handwritten Uyghur Character Recognition, Pattern Recognition and Artificial Intelligence", (Dec,2012) Vol.25, No.6, pp: 979-986.
- [22] Jiang Wen, Lu Zhao-yang, Li Jing, Recognition of Uyghur Handwritten Character Recognition Based on Directional Element Feature, *Micro-Electronics & Computer*, (Oct,2013), Vol.30, No.10, pp: 97-100.
- [23] Jiang Wen, Liu Li-kang, Recognition of Handwritten Uyghur Character Based on Real Value Gabor Energy Feature, *Computer Engineering and Design*, vol. 34, no. 9, (2013) September, pp. 3271-3278.
- [24] M. Ibrayim, "Key Technologies for Recognition of On-line Handwritten Uyghur Characters and Words", PhD thesis, Wuhan University, (2013) May.
- [25] K. Zulpiya, "Research on Online Uyghur Handwritten Character Recognition Based on Feature Combination", M. S. thesis, Xinjiang University, Urumqi, China, (2013).
- [26] 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)*, (2013), Athens, Greece, Dec.12-15; SP-6046, pp. 000293 – 000298.
- [27] A. Hamdulla, W. Simayi, M. Ibrayim and D. Tursun, "Research on On-line Uyghur Handwritten Character Recognition Technology Based on Modified Center Distance Feature *International Journal of Signal Processing*", *Image Processing and Pattern Recognition*, vol. 7, no. 5, (2014) October, pp. 409-424.
- [28] A. Hamdulla, W. Simayi, M. Ibrayim and D. Tursun, "MCDF Based On-Line Handwritten Character Recognition for Total Uyghur Character Forms, *Pattern Recognition*", *Proceedings of 6th Chinese Conference, Part II*, (2014) , Changsha, China, November 17-19, pp. 474-480.
- [29] W. Simayi, "Online Uyghur handwritten character Recognition technology based on Center Distance Feature", Ms thesis, Xinjiang University, (2014).
- [30] R. Plamondon and S. N. Srihari, "On-line and Offline Handwriting Recognition: A Comprehensive Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, (2000) January, pp. 63-84.
- [31] A. Raj, R. Srivastava, T. Patnaik and B. Kumar, "A Survey of Feature Extraction and Classification Techniques Used In Character Recognition for Indian Scripts", *International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958*, vol. 2, no. 3, (2013) February, pp. 566-568.

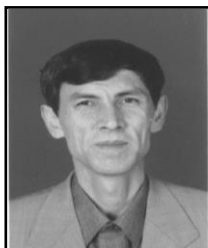
Authors



Wujiahemaiti Simayi, he has received his B.E. and M.S. degree in Electronics, Signal & Information processing from Xinjiang University, China, in 2009 and 2014, respectively. Currently, he is a PhD candidate in Computer Applications in Xinjiang University, and working as research assistant at the Key Laboratory of Intelligent Information Processing, Xinjiang University, China. His research interests include feature extraction and classification techniques, handwritten character recognition, computer simulation for atmospheric pollution dispersion modeling.



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.



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. Currently, he is a professor in the School of Software Engineering, Xinjiang University. He has published more than 140 technical papers on speech synthesis, natural language processing and image processing. He is a senior member of CCF and an affiliate member of IEEE.

