Int. J. Advance. Soft Comput. Appl., Vol. 2, No. 2, July 2010
ISSN 2074-8523; Copyright © ICSRS Publication, 2010
www.i-csrs.org

Writer Identification for Chinese Handwriting

Wong Yee Leng & Siti Mariyam Shamsuddin

Soft Computing Research Group
Universiti Teknologi Malaysia,Skudai, Johor Bahru
e-mail: yeeleng28@gmail.com

Soft Computing Research Group
Universiti Teknologi Malaysia,Skudai, Johor Bahru
e-mail: mariyam@utm.my

Abstract

Chinese handwriting identification has become a hot research in pattern recognition and image processing. In this paper, we present overview of relevant papers from the previous related studies until to the recent publications regarding to the Chinese Handwriting Identification. The strength, weaknesses, accurateness and comparison of well known approaches are reviewed, summarized and documented. This paper provides broad spectrum of pattern recognition technology in assisting writer identification tasks, which are at the forefront of forensic and biometrics based on identification application.

Keywords: *Chinese Handwriting, Discretization, Feature Extraction, Pattern Recognition, Writer Identification.*

1 Introduction

Since the initiation of Multimedia Super Corridor (MSC) project in Malaysia, recent studies in the field of computer vision and pattern recognition show a great amount of interest in the content of Biometric Security, Information and Communication Technology. As a result, computer system as a tool for information and communication medium is becoming more vital since then. In MSC themes, there are few focus areas that have been identified, and these include Digital Content, RFID Technology, Advanced Materials and Biometric Technology, which has become an important research area now days.

A biometric system is one of the established systems that follow the concept of pattern recognition technology. It begins with the input from individual's biometric data, undergoing feature extraction and finally compare extracted feature against the model set in database to obtain the precise result. This tremendous growth resulted in many biometric applications being developed and commercialized for security and crime identification purpose. The related studies include Handwriting [1,2], Signature [3,4], Iris [5,6], Facial Features [7,8], Fingerprints [9] and others. Furthermore, pattern recognition as described by Anil, Robert and Jiangchang [10] is the most critical role in human decision making task.

Recently, interest in the area of pattern recognition has been well improved and exposed due to the emerging applications which are not only challenging but also attracted many researchers' attention. Such new and emerging applications include data mining, web searching, retrieval of multimedia data, face recognition, handwritten recognition, which require robust and intelligent solutions. Basically, the conventional process of pattern recognition involves three phases: data collection and pre-processing, data representation and decision making [11, 87]. Fig. 1 shows the conventional framework of pattern recognition system.

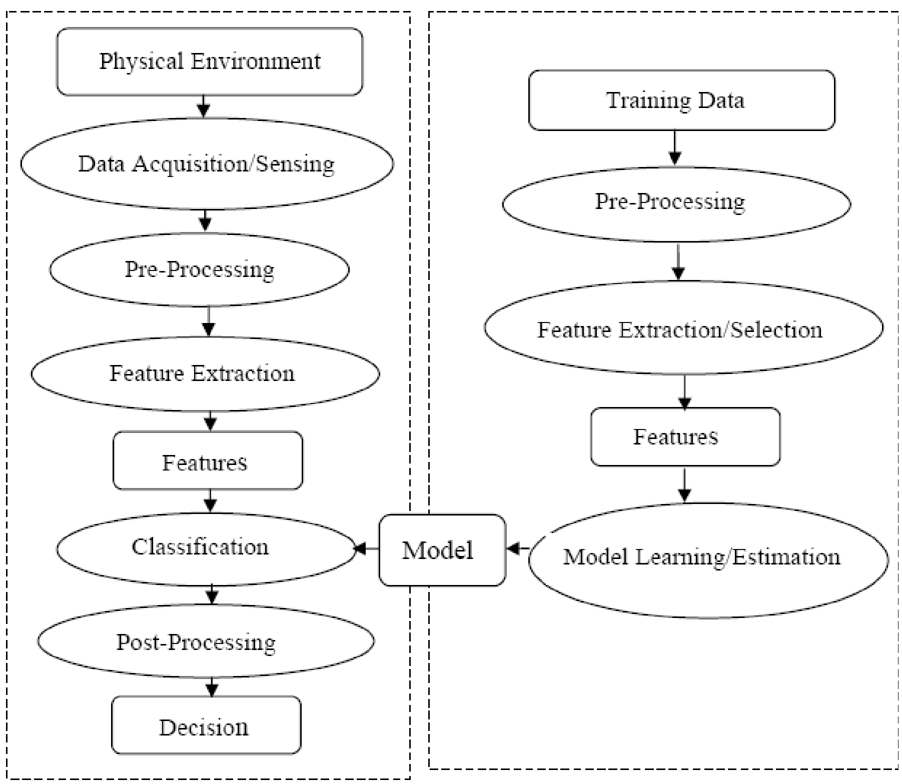


Fig. 1 Conventional Framework of Pattern Recognition System [11]

1.1 Pattern Recognition Technology (Identification, Recognition and Verification)

Generally, Pattern recognition in handwriting covers all types of application including handwritten for identification [2], verification [1], authentication [12,13] and character recognition [14]. Each of the proposed work has their own approaches to achieve their goal. Guo, Christian and Alex [2] proposed a character prototype approach as a model template to assist their alphabet knowledge base. These proposed approaches present additional information to support writer identification process and simultaneously preserve the style of handwriting's individuality. Their experimental results have successfully increased the accuracy from 66% to 87%. There are other studies that focus on pre-processing phases such as normalization, and feature extraction for better recognition and identification. It is vital to understand the history and complexity of the handwriting structure prior to classification and categorization.

In automatic handwriting recognition, variations between different handwritings are eliminated [15]. In other words, the system is able to classify the shapes of each characters and words effectively. On the contrary, handwriting identification requires dissimilarity or uniqueness information, i.e., writer's characteristics to determine the identity of a writer [16, 17]. The differences procedure between both patterns can be done and illustrated at early stage including pre-processing and feature extraction approaches of the handwriting. Detailed information on pre-processing and feature extraction process will be given in next section. Fig. 2 depicts the standard framework of writer identification [18].

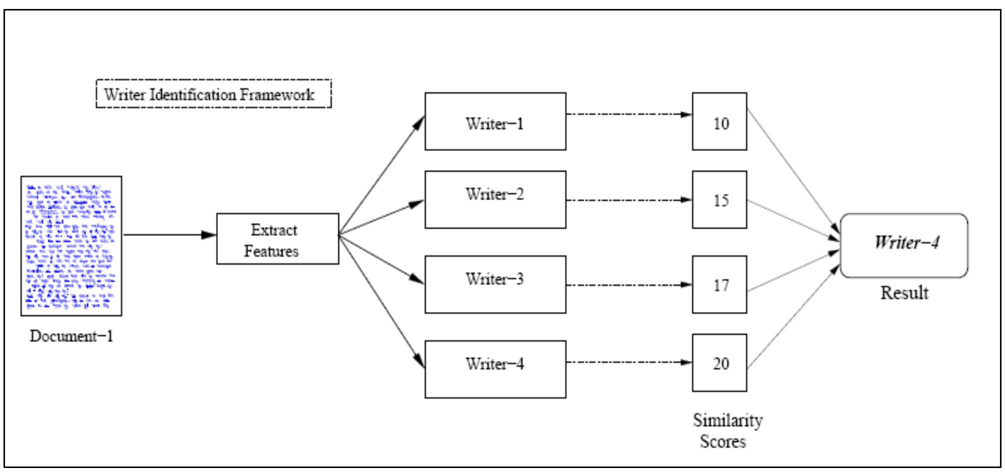


Fig. 2 Writer Identification framework [18]

A writer identification system begins with a search 'one to many' in a large database with known authorships samples and in return, is a list of candidates containing the handwriting. The hit lists are then examined by a human expert. On the other hand, writer verification is belonging to yes or no decision making and one to one comparison process. Fig. 3 shows the common writer verification system. In summary, handwriting recognition is the task of transforming a language in its spatial form into its symbolic representation. While, handwriting identification is the task of determining the author of a sample of handwriting from a set of writers, assuming that each person's handwriting is individualistic. On the contrary, handwriting verification is tasks of determining whether or not the handwriting is belong to the right owner [1].

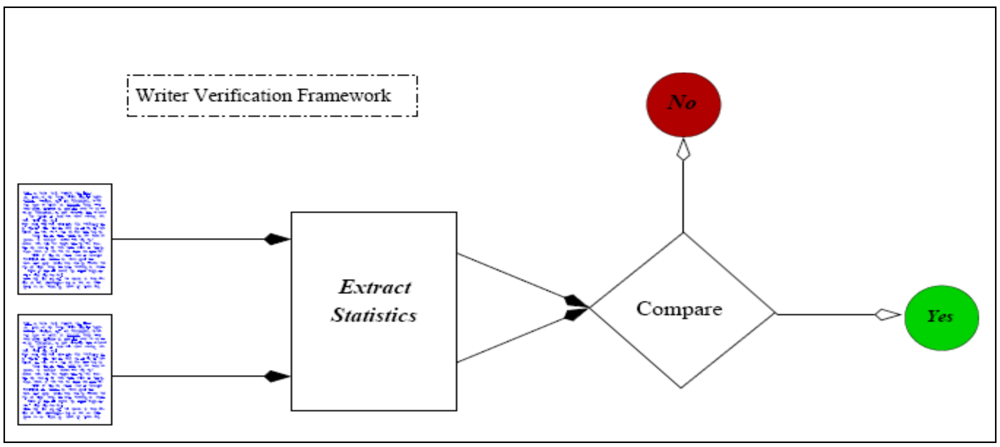


Fig. 3 Writer Verification framework [18]

Identification and verification [19] commonly play their role in forensic analysis to determine the special nature contained in a writing of a specific writer [20]. Handwriting recognition is responsible to remove the dissimilarity in handwriting and to interpret the idea of the message. The hierarchy in Pattern Recognition is summarized in Table 1.

Table 1: Hierarchy in Pattern Recognition

Features	**Handwriting Recognition**	**Handwriting Identification**	**Handwriting Verification**
Task	Transform a language in its spatial form into its symbolic representation	Determine the author of a sample handwriting from a set of writers, assume each writer's handwriting is unique	Determine whether handwriting is belonging to that person or the other one
Responsibility	Eliminate the variation between different handwriting (remove dissimilarity)	Require dissimilarity, which are the characteristic of the handwriting sample	-
Application	-Commonly used to classify the shapes of each characters and words effectively Attempt to determine the idea behind the message	Commonly used in forensic analysis to determine the special nature contained in a writing of a specific writer [20].	
Writer	-	In writer identification -one to many searching -Who write the sample of handwritten text?	In writer verification -one to many comparison -Were these two samples of handwritten text written by the same person? -Yes and No decision

2 Overview of Chinese Handwriting Identification

Chinese characters are ideographic in nature with more than 50,000 characters, of which 6000 characters are commonly used and have a wide range of complexity [21, 22]. Chinese character can be expressed in two common styles, block or in cursive handwriting. The Chinese characters have an average of 810 strokes in block style. According to Fang [22], the complication structure in Chinese character mostly affected by multi stokes of each character. The characters may consist of one to thirty or more distinct strokes due to the variety of handwriting style. His proposed algorithms successfully unite some sub strokes into the proper and complete stroke to determine a Chinese character. Fig. 4 shows similar Chinese character with different writing styles which has created a different number of strokes, called multi strokes.

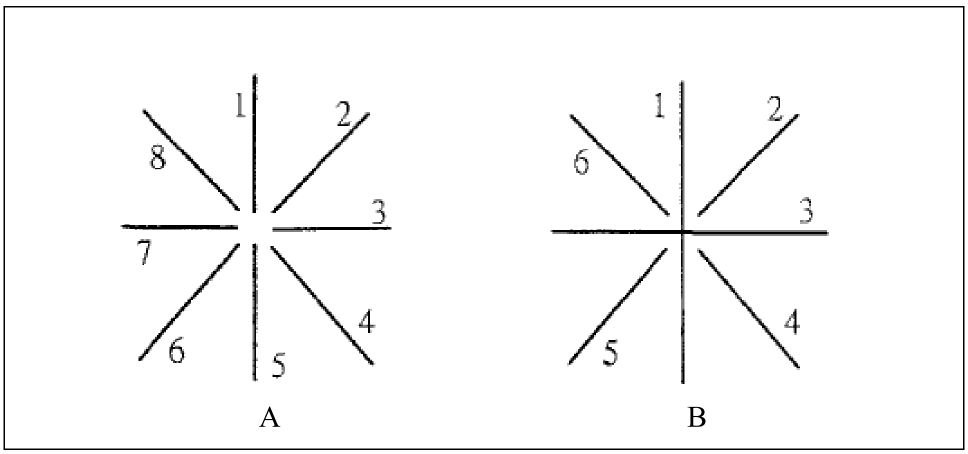


Fig. 4 Same Chinese character (米) with different number of strokes [22]

Moreover, most of Chinese character has similar characteristics but define a different meaning. Fig. 5 illustrates similar characteristics of Chinese character. Optical Character Recognition (OCR) [23,24] is a most popular tool used to determine a Chinese character. As usual, in OCR system, input characters are read and digitized by an optimal scanner. Each character is located, segmented and finally the produced matrix is fed into pre-processor for other process such as smoothing, normalization or thinning for noise reduction. Normalization or thinning techniques is needed due to the complex structure of Chinese and can appear in different sizes and fonts. Chinese characters also have a large number of mutually similar characters. Brief description of normalization and other techniques for noise reduction can be found in subsection 2.4.1.1.

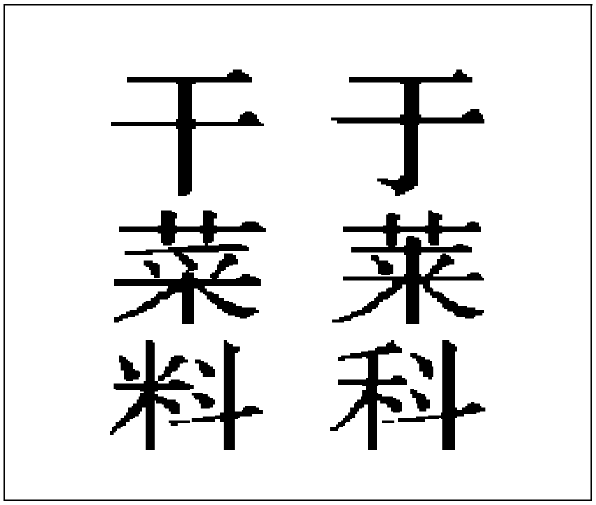


Fig. 5 Examples of similar Chinese character with different meaning [25]

Another attractive part of Chinese languages is where each Chinese word has its own characteristic and uniqueness in forming another meaningful compound word by combining related words with its neighboring character. For instances, Fig. 6 shows a combination word with its new meaning derived from two Chinese characters. In Chinese sentences, there is no explicit separator and spaces between Chinese words to indicate its boundaries. Each word is written continuously with equal spaces between them.

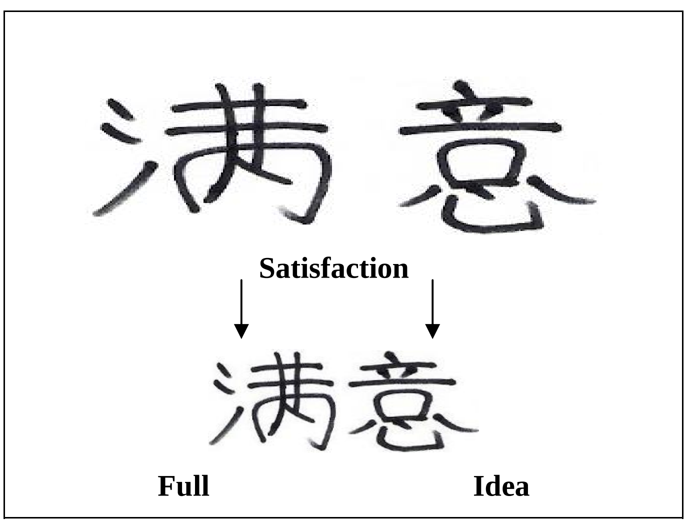


Fig. 6. Examples of new created word and its meaning

Recently, many identification systems have been developed for numerous languages including Arabic, Chinese, Japanese, Korean and others. However, the most difficult character are Chinese due to the enormous number of over 50 000 and has a complex stokes structural character. As such, it involves a different handwritten processing and feature extraction method in which this method is not applicable in some type of languages. The differences of identification accuracy on different language of handwriting can be seen in the Table 2. Thus, we can state that different handwriting of spoken languages leads to different approaches in order to achieve competent identification.

Table 2: Feature Extraction and Writer Identification Methods on Multiple Languages

Authors+ References	**Feature + Identification methods**	**Language**	**Summary**	**Identification Results**
Bangy Li et al. [26]	Hierarchical Structure in Shape Primitives + Fusion Dynamic and Static Feature for writer identification	English and Chinese text	Experimental result shows that writer identification performance in English better than Chinese text. Reason: English text contains more oriental information than Chinese text.	Chinese accuracy>90 % English accuracy>93 %
Guangyu Zhu et al. [27]	Shape Codebook + Multi Class SVM Classifier	Arabic, Chinese, English, Hindu, Japanese, Korean, Russian and Thai script	Confusion of template (mismatch) occurred. -matching between Japanese and Chinese. Reason: vary number of characters.	Chinese identification rate significantly lower (55.1%) compared to others eight languages.

Anoop Namboodiri and Saphin Gupta [28]	Shape based Curve Extraction + Clustering for writer identification	Arabic, Roman/Chinese, Cyrillic, Hebrew and Devanagari	Shape based primitive proved bad choice for Chinese script. Reason: shapes extracted were straight lines and do not contain much individuality information's.	Accuracy of Roman or Chinese achieved only 50% and others achieved > 85%
Judith Hochberg et al. [29]	Means, Standard Deviation and Skew of five Connected Component feature + Linear Discriminant Analysis for classification	Arabic, Chinese, Roman, Cyrillic, Japanese and Devanagari script	Some document misclassified (tended to confuse) -Between Roman and Cyrillic. -Between Chinese and Japanese.	Through this method, Japanese achieved highest matched (100%) meanwhile others script achieved matched accuracy > 75%

3 The Methodology of Chinese Handwriting Identification

Handwriting is personal to individual. The most important task in handwriting identification is to determining, then extracting handwriting samples that has distinguishing features and finally deal with the machine to provide the best decision. To illustrate this, a number of questions need to be answered; for instances: How can a valid identification be made, given the differences in the same character written by the same person at different times? These are the common questions and issues in which most of the expert and researches have concerned themselves in recent years.

Chinese handwriting for writer identification has been well known as a lively research concern in image processing, pattern recognition and computer vision area [30]. Recently, many innovative methods and approaches have been developed for writer identification using Dynamic Feature on Strokes [31], Fusion on both Dynamic and Static Features [32], and Textual Analysis [16]. Those approaches managed to overcome the complexity of Chinese character to make identification task easier. However, each of the method has advantages and disadvantages.

Table 3 summarizes the comparison on advantages and disadvantages of Identification methods. Meanwhile appropriate approaches developed for character recognition on Chinese handwriting are said to be useful for earlier understanding such as Hidden Markov Model (HMM) based approach [14][33]. They have been proved well and known as established approaches but some weaknesses still present. This approach often required large data in computation process and was known as a multifaceted method to develop classifier for recognizing each character from large character set.

However, to have a smooth and efficient identification task, a well concern needed to be placed at the beginning of the process. The approach of this study can be classified into four processes: pre-processing, feature extraction, discretization and classification for writer identification phase. Discretization approaches is applied to assist classification task. According to Azah *et. al* [34], discretization provides loads of advantages when dealing with huge and continues data. The authors have made some comparison on post-discretized and pre-discretized data. Their significant result shows an achievement accuracy of 99.9% on writer identification using post-discretized data instead of pre-discretized data. Detail on discretization can be read in section 3.4.

Table 3: Advantages and Disadvantages of Feature Extraction and Identification

Authors	**Approaches**	**Pros**	**Cons**	**Result (%)**
Tong Hua Su et al. [14]	Segmentation Free Strategy based on HMM	-state of art strategy.	-very depending on CPU time and memory. -very depending on the experiences of a designer.	Statistically significant with confidence more than 99%.
YuChen Yan et al. [35]	Spectral Feature Extraction Method based on Fast Fourier Transformation	-successfully reduced randomness in Chinese character. -obtained stable feature -feasible for large volume of data set.	-required higher computation costs.	Handwriting samples achieved from 100 persons=98% 500 persons=95%

WenFeng Jin et al. [36]	Sum Rule, Weighted Sum Rule and User Specific Sum Rule applied to combine dynamic and static features	-reduce the number of character.	- focused only on 12 primary strokes for Chinese character. -more complex fusion method, more training data needed; hence decrease the performance of identification.	With dynamic features, achieve accuracy of 80 sample characters same as 120 characters. Combination features perform better accuracy.
Bangy Li and Tieniu Tan [31]	Temporal Sequence Code (TSC) and Shape Codes (SC)	-required small number of character. -effectively work for English and Chinese character.	-depending on emotions and physical state of writers, which will affect the individual performance on handwriting due to the environment changes and etc.	Chinese template achieved >90% English template achieved >95%
He and Tan [37]	-Using Gabor Filters+Autocorrelation Function	-applicable to both text dependent and text independent Chinese handwriting.	-required long calculation time. -high computation cost.	Top 1=90% Top 3=96% Top 5=100% Top 10=100%
Hang Joon Kim et al. [33]	Hidden Markov Model	-work well with variety of cursive strokes. -required small memory	-required large data set (18,000 handwriting character). -not applicable for real time applications if Chinese character set are too large. -The search algorithm has a time complexity depend on the level of input pattern.	Accuracy of recognition rate of 90.3% and a speed of 1.83 s per character

3.1 Variation of Chinese Character

Chinese handwriting can be expressed in variety of style, which is commonly based on the alignment and formation of multi stokes pattern. Naturally, the style of Chinese character can be seen differently based on static, dynamic [38] and geometrical features. Static or stationary features are essential and it tells apart each character from other character. Dynamic features on the other hand represent generative aspect of the characters. Slant, ornamentals, aspect ratio, relative position and size of the strokes, corners and retraces included in geometrical structures. Because of these numerous variation of stroke construction to write a certain Chinese character, the number of writing style of a Chinese character can be indefinite. These features are those important points to be thought of when we want to determine a writer of handwriting sample from a set of writers. The main challenges for Chinese handwriting based writer identification arise when we are dealing with different style of Chinese character that written by the same author. Thus, the ultimate goal of designing an Identification System based on handwriting with an accuracy rate of 100% is quite illusionary, because even human beings are not able to identify every handwritten character of the writer without any uncertainty. For example, most people cannot even read their own handwritten notes. However, there is interesting paper that would change our mind on this matter and help identification work well. Xin Li and Xiaoqing Ding [39] carried out experiment on two Chinese handwritings written by one writer and another experiment with two Chinese handwritings written by two writers. Their result clearly shows that the differences measurement between handwritings of two writers is higher than the measurement between handwritings of the same writer. From their observation, it can be summarized that although handwriting style may vary from same writer, but with assistant of soft computing techniques, we still can differentiate handwritings of those writers.

3.2 Pre-Processing Phase

Chinese handwritten character can be written in variances of stroke thickness and pattern depends on the writer's writing pressure or style of handwriting. This disparity of style, stroke width, increases the variances between different samples of handwritten Chinese character. Those were the features considered by researches to distinguish the writer of the handwriting. Thus, the best processing approaches on Chinese character is to normalize all character into a uniform size of pixels, binarized them or remove the background noise of the character in order to smooth the progress of the identification. In contrary, this approaches work differently for handwriting recognition.

As for handwriting recognition, they do not care on the distinctive features of the character but similarity of the features between two characters. Thus, the complexity of the Chinese character structure, the large number of characters used, and misrepresentation of the character due to multiple styles of writing were the most challenging issues for Chinese handwriting recognition system. Andreas and Horst Bunke note that “Any potential reduction in handwriting individuality is compensated for by a gain in recognition performance” [40]. They applied three common normalization techniques, namely slant correlation, width normalization and vertical scaling to their writer identification system through Hidden Markov Model (HMM) based recognizer. Their experiment shows that if the combination of three normalizations or slant correlation alone were omitted, then the word recognition rate drop but writer identification rate significantly increased. And regards to this issues, many other researches attempts to solve Chinese character problem by undergoing many studies on pre-processing methods for character recognition purpose including shape normalization [41] and stroke extraction [42,43], which were commonly implemented along with recognition system based on unsupervised classification [44], Artificial Intelligence and Neural Network approaches [45]. The importance of Pre processing technique is well described by Bing et al. [46]. In this section, we discuss thinning and normalization techniques, which are the most common techniques used in pre processing part. As for our work, normalization is applied to remove the irrelevant variability noise occurring in the raw coordinate sequence of our Chinese handwritten character, binarization and conversion of the input character into the uniform size.

3.2.1 Normalization

The handwritten text is originally saved as BMP format file, and converted into grey level image in order to assist identification task later. As we acknowledged, the handwritten text contains spaces between each character words, letters or lines. Beside this, character may appear in different of sizes between words or lines. Thus, normalization techniques need to be applied to an original text line to eliminate noise and transform them into a standard clean character representation. Normalization text begins by determining the location of the text line by using horizontal projection profile [47]. The lines and spaces between character words are eliminated by text line and spaces normalization. Finally, the spaces between words or lines are set to the predefined size of text padding. More information about normalization procedure can be read in Nafiz and Fatos [48] and Villegas and Aviles [49]. Below shows the summarization of the common and basic steps for normalizing handwritten character;

Step 1: ***Project the text line***

The location of the Chinese text line is determined using Horizontal Projection Profile (HPP).

Step 2: Normalize the text line

Original handwritten Chinese text contains different point size of character or words, variety of handwriting style and font type. Thus, it is essential to normalise and represent this Chinese character into a standard representation. Once the text line is located, each of the character will be normalized into same equal size.

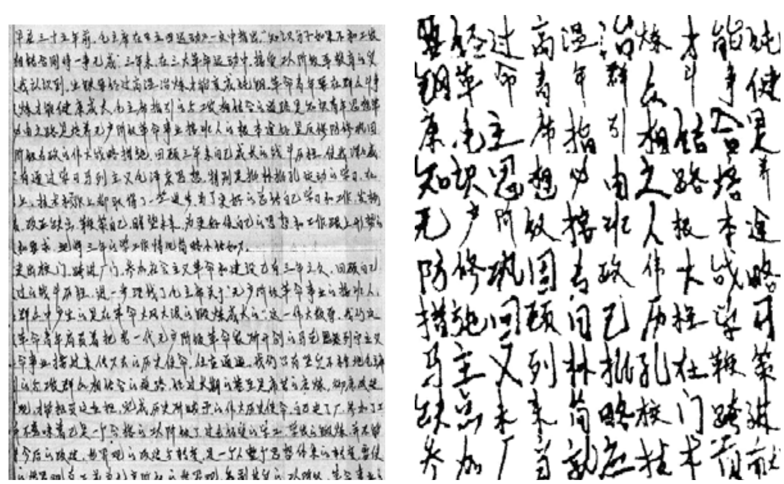
Step 3: Normalize the space between Chinese text lines

Undesired spaces between each Chinese text line can be reduced and eliminated using Vertical Projection Profile (VPP).

Step 4: Text padding

Construct a cleared texture handwritten image by text padding.

However, normalization method needs to be done carefully to see its effect on the handwritten text since some normalizations may swab away writer's information. Not all normalization techniques can ensure high accuracy of writer identification as mentioned earlier. Fig. 10 illustrates an original Chinese handwritten image and the resulting texture image after applying the normalization process.



A) Original handwritten text B) Texture image

Fig. 10 An example of the original image and the normalisation result of the Chinese handwritten text [35]

3.2.2 Thinning

Thinning is a process of reducing the length of character image from a several pixels to a single pixel. The final image after thinning process is called skeleton. Rosenfield [50] examined the criteria that need to be achieved in order to reduce the noise of binaries handwritten images. Fig. 11 illustrates an original scanned image and the resulting skeleton after applying the thinning process.

Fig. 11 An example of the original image and the thinning result of the Chinese character [21]

3.3 Feature Extraction

Many researches have been well studied and focus on finding the best feature extraction methods to enhance the effectiveness of writer identification task [51]. As we acknowledged, feature extraction methods cannot be applied directly to any text. For instances, feature extraction methods for text dependent system should be done in a lower level considering character, sub-character or at word level. A review on script identification techniques for multiple script documents was well written by Ram Sarkar et al. [52], Abirami and Manjula [53] considering the handwritten level of identification.

Xin Li and Xiaoqing [39] have identified the major issues in Chinese character, mostly affected by variety of writing style. The authors have proposed a histogram based feature to assist writer identification task which was known as grid microstructure feature. The uniqueness of their proposed method is due to the grid microstructure feature that generated from a size-adjustable grid. This method focuses on writing tendency of the writer in local region. The experiment result successfully showed that their recognition accuracy seems better compared to the existing text independent methods for Chinese writer identification. Meanwhile, Yuchen Yan et al. [35] proposed spectral feature extraction by implementing mathematical expectation value to the image's spectral features to enhance writer identification accuracy. Their method successfully eliminated effect of randomness spectral features and significantly improves identification accuracy.

In [54], the authors' implemented moment based feature extraction approaches on training phases to produce feature vectors. The features are extracted during testing phase for classification. The results showed that moment based feature computation are faster than texture based methods. This advantage significantly drives to the classifier accuracy over 97%.

Another feature extraction technique based on moment function has been successfully carried out by Azah *et. al* [34]. The authors have conducted experiment with various words representation to show the effectiveness of the proposed moment function. The result successfully illustrated that the proposed method is capable to extract the feature's shape with the validation on similarity measurements in terms of Mean Absolute Error (MAE) function. According to the authors, the uniqueness of each individual are still maintained. Thus, this advantage can be used to identify writer's individuality according to the class of features accordingly. Table 4 depicts related studies on feature extraction for handwriting identification.

Table 4: Feature Extraction method for Writer Identification

Writers	**Features method**	**Identification text level**	**Advantages**	**Identification Result**
YuChen Yan et al. [35]	Fast Fourier Transformation (FFT)	Paragraph text	-Eliminates randomness of spectral feature. -Applicable to large handwriting sample	100 writers-98% 500 writers-95%
Anuraj Bhardwaj et al. [54]	Geometrical Moments (GM)	Word level	-Faster than texture based methods. -Applicable for multilingual document	Accuracy >98% over three different languages (Arabic, Chinese and Hindi.

Zhengyu He et al. [55]	Wavelet based GGD method	Paragraph text	-Applicable to multilanguage document. -Higher identification rate and lower computation time	-Rate of 85.71% at top of 10 matches. -Lower elapsed time (0.53) compared to 2D Gabor model.
Azah Kamilah, Siti Mariyam and Maslina Darus [34]	Integrated Aspect Scaling Invariant (ASI) into United Moment Invariant (UMI)	Word level	-Good handwritten shape features descriptor through basic transformation of rotation, translation, scaling. -Preserve the character's shape after scaling.	-Result shows that the proposed method gain the smallest MAE compared to other moment techniques such as GMI, ASI-GMI, and UMI.
Anoop Namboodiri and Saphin Gupta [28]	Shape based Curve Extraction	Sub Character level	-Required only small data testing. -Fast classification process.	Acuracy>85%
Shahabi and Rahmati [56]	Multi Gabor Filtering Channel and co-occurrence matrix feature (comparison)	Document level	-Reflect large variability between handwriting.	Gabor Filter achieves better accuracy than Co-occurrence matrix features.
KunYu et al. [57]	Gaussian Mixture Model (GMM)	Paragraph text	-Decrease computational cost with improved accuracy.	Accuracy of 62.5% (without feature extraction). Accuracy of 80% (feature extraction).

Bin Chang and Sargur [58]	GSC base on Word Feature algorithm	Word level	-Contains more individuality information. -Efficient for identification and verification.	Successfully identify correctly 80% of the writer with the combination of words (been, Cohen, Medical and referred).
Sreela Sasi et al. [38]	Wavelet Packet Transform (WPT)	Character level	-Lower cost. -Lowest mean square energy value across all compression ratios.	Result shows multi resolution level of each character better than using only fuzzy logic.

3.4 Discretization

Most of the classification problem focuses on a set of training instances. Usually each training instances will be categorized into one of the classes and is described by a number of distinct features. By undergoing a discretization process, the ranges of each continuous features is transformed into a discrete partition, represented by a number of intervals. Upper and lower boundary in each interval represents the interval range.

However, to represent continuous features, there are thousands of ways can be explored. Thus, there are two important points that need to be implemented. First, we need to decide the number of discrete intervals. In most cases, the intervals are selected by the users randomly. Secondly, determine the locations of the interval boundaries. Commonly, there are several discretization methods being used, and these include Equal Information Gain (EIG), Maximum Entropy (ME), Equal Interval Width (EIW) and others. Recently, Invariants Discretization has been proposed and successfully given higher identification rates [34]. It is treated as a supervised method that searches for a set of appropriate interval which is entitled to represent writer information's. Each interval is approximated with upper and lower values. The number of interval is the same as the number of feature vector from each word image. According to the authors, feature vectors that are group into a similar interval will have the same characteristics, which represent the owner of the handwriting.

The beneficial of Discretization included easier human interpretation through a set of intervals [59], faster computation process and higher accuracy because the amount of data is reduced [60,61], and non linear representation [62]. Therefore, it can be said that it is necessary to have post-discretized data instead of pre-discretized data for a better classification. This is proven in the studies conducted by [34]. Their result shows that the implementation of dicretization on their proposed Integrated Moment Invariants successfully increases the accuracy of writer identification.

Geneddy and Stanimir [63] addressed the issues in supervised and unsupervised discretization method on continuous data. They enhanced two supervised descritization method namely entropy based discretization and MVDM-based discretization. The result proves that both enhanced algorithm successfully boost the accuracy in classification. On the other hand, unsupervised discretization method, correlation preserving discretization was proposed by Sameep, Srinivasan and Hui Yang [64] has shown the efficiency of the proposed algorithm on multivariate dataset and it is sufficient to predict missing values.

3.4.1 Classes of Discretization methods

Discretization methods can be categorized into supervised versus unsupervised class, global versus local and disjoint versus non-disjoint category. Further details on categorization of discretization methods are well described in Table 5.

Table 5: Classes of Discretization method

Authors	**Discretization class**	**Features**
Gennady Agre and Stanimir Peev [63]; Sameep et al. [64]	Supervised	Method that depends on the class informations to determine the discretization's interval
	Unsupervised	Independent method, do not need class information.
Dougherty et al. [65]; Chmielewski and Grzymala [66]	Global	Method that uses a single set of interval to represent a single classification task. Discretization happens only once in this method.
	Local	Different set of interval represent a single attribute; each set can be applied to a different classification task.
Ying Yang and	Disjoint	Method that discretize range of interval

Geoffrey [67]		values into disjoint cuts without overlapping with each other.
	Non-disjoint	Method that discretize range of values into interval with overlapping among intervals.

3.5 Writer Identification

Three important parts in a handwriting identification system are pre-processing, feature extraction and classifier. Feature vectors or feature set that gained through feature extraction techniques used to represent all characters of handwriting. The features are then used as input to the character classifier. The common issues found in identification system commonly involved preserving dissimilarity character features of each writer, finding the best discriminating features, selecting the best way to compare them and creating the classification rules that suits with our handwriting language. These issues are highly dependent on each other. Template matching is one of the most popular classification methods. This is how its work. First, individual character image pixels in template matching are used as features. Then, classification is performed by comparing an input character image with a set of templates from each character that belong to specific set of writer class. Each comparison will produce distance weight measurement between input characters with the template. If the pixels in the observed character image are identical or near to the pixels in template, then the probability of distance measurement will be lower. Otherwise, the character is not belonging to that class. After gaining the distance weight measurement of the character position, the character's identity will be assigned to the most similar template belong to the writer class.

Details description on the classification techniques that are commonly chosen for writer identification can be read in this section. Classification techniques can be group into two components which are *classifier* known as supervised classification and the other one is *clustering*, namely unsupervised classification. Detail descriptions on supervised and unsupervised data classification can be read in paper written by Bagirov et al. [68]. The authors inspect various approaches that best suit on clustering and classification data. Numerical experimental is carried out with k-Nearest Neighbour, Clustering Connectionist techniques, Evolutionary Approaches, Support Vector Machines and etc. The result shows that non-smooth optimisation approach best for clustering problem. As for classification, it is based on non-smooth and global optimisation approaches. It is proven that not all approaches can best fit for all data classification techniques. This is because different classification techniques will have different approaches, rules features to be followed. The next section gives the description on classifier and clustering of the classification.

3.5.1 Classifier

This classification technique is often implemented in writer identification phase where the handwritten text has been classified into a few classes according to their similarity features. The general classifiers that would be discussed here are K Nearest Neighbours Classifier and Weighted Euclidean Distance.

3.5.1.1 Weighted Euclidean Distance Classifier (WED)

The main goal of this classifier is to find out a feature vector from training data that is the closest one to the test data feature vector. According to Yong Zhu, Tieniu Tan and Yunhong Wang [69], features representation for each writer was the feature extracted during the training of handwritten text. Each newly input handwritten text from unidentified writer will undergo feature extraction process too. Then, the newly extracted features from unidentified writer were compared with the existed features of a set of identified writers in a database. They tested this classifier on the features that have been extracted by using Multichannel Gabor Filtering and Grey Scale Co-occurrence Matrices (GSCM) on 40 different writers and successfully obtained the identification accuracy of 96 %.

Common mathematical rule for Weighted Euclidean Distance Classifier is described in equation (1).

$$WED(k) = \sum_{i=1}^{N} \left[\left(f_i - f_i^{(k)} \right)^2 \Big/ \left(\delta_i^{(k)} \right)^2 \right] \qquad (1)$$

Where,

f_i ; i^{th} features of the testing data,

$f_i^{(k)}$; mean value of training data from writer *i*,

$\delta_i^{(k)}$; variance value of training data from writer *I*,

N ; total number of features extracted from a single writer.

3.5.1.2 Nearest Neighbour Classification

Nearest Neighbour classification is usually performed in a 'leave one out' strategy and popularly used in writer identification by classifying the handwritten character [70]. Handwritten text will be compared with all existed sample in database. In this rule, distances measurement between two character images is the important criteria to be considered.

The detail explanation of Nearest Neighbour classifier (K-NN) used by Said et al. [72], in his work, is well illustrated in equation (2).

$$d_k = \left[\sum_{j=1}^{N} \left(U_j - f_{kj} \right)^2 \right]^{\frac{1}{2}}, \tag{2}$$

where,

d_k ; class distance of unknown writer,

U ; feature of unknown writer,

f_{kj} ; best feature vector, *j* of class *k*,

$j = 1 \ldots N$; number of features,

$k = 1 \ldots$; number of classes.

According to the authors, each feature vector of a class, K in a training set is represented as f_k . The authors resolved the class, K of writer by measuring the similarity of the distance computed. According to them, by using K-NN approach, they successfully achieved identification accuracy of 90.0%.

3.5.2 Clustering

Clustering or popularly known as unsupervised classification, commonly used to cluster the training samples, for instances generating graphemes codebook. The common classifiers that would be discussed here are K-means Clustering and Kohonen Self-Organizing Feature Map. Literature on others clustering approaches on handwriting can be read from paper written by Strehl and Ghosh [73], Fern [74], Fei Yin and Cheng [75].

3.5.2.1 Support Vector Machine

Support Vector Machine (SVM) was first introduced by Boser et al. [76]. The purpose of SVM is to map feature vectors into a higher dimensional feature space, and then creating a separating hyperplane with maximum margin to group the features. SVM approach often used by researches to solve fault in pattern recognition [77]. Michael and Joshua [77] briefly evaluate the theory of SVM approach. According to them, the SVM is a learning apparatus that can easily represent problem with yes and no decision, which is 1 and -1 as illustrated in equation (3).

$$x \mapsto \begin{cases} 1 & \text{if } w^T x - \beta \geq 0; \\ -1 & \text{otherwise.} \end{cases} \tag{3}$$

Where;

$w^T x - \beta$; used to determine classifier

w and β ; represent features of training set

The most challenging when dealing with Chinese character classification is the large number of class type in Chinese character [78]. As such, Fu Chang [78] combined three approaches in pre-processing stage to show the efficiency of SVM classification, which could enhance the accuracy of their recognition task.

Another paper on SVM written by Fan and Fang [79], indicate that the combination of SVM and other method may carry out a better task in classification. Their integrated approaches, SVM and Rough set successfully perform faster and faultlessly compared to SVM alone when dealing with Chinese character.

3.5.2.1 K-means Clustering

K-means clustering algorithm very depends on the quality of training data. Accuracy of this algorithm might be affected if there is much noise founded during data training. Fred and Jain [80] performed K-means clustering algorithm in their work to gain a number of partitions of a data set.

3.5.2.2 Kohonen Self-Organizing Feature Map

This type of algorithm is popularly used as a tool to understand high and huge set of data. Kohenen Self Organizing Map consists of a map of a standard grid of neuron. Each neuron is associated with vector values that represented by features. The map tries to represent all existed observations with the maximum exactness using classified set of models. The models are ordered on the grid so that the similar models close to each other and dissimilar models disclose from each other.

In the paper written by Lambert, Marius and Katrin [81], Kohonen's algorithm was implemented in three steps to seek for the accuracy of automatic writer identification. First, the fragmented connected component that represents the handwritten ink is extracted from the images page. Then, the fragmented connected component contour training set was presented to a Kohonen self organizing feature map and was trained with Beowolf high-performance Linux cluster of 128 nodes. Secondly, the writer specific feature vectors were computed from a database of 150 writers and represented by histogram. Finally, writer identification performance was evaluated by a sorted hit list of writer samples.

Overall, Self-Organizing Map (SOM) is an unsupervised single layer artificial neural network (ANN). SOM learning process work just exactly like K-means, where it will keep iterates until each cluster centre converges to the centre of the possible texture pattern. The advantage of SOM is that the more number of neurons are placed in the grid, the higher classification result will be obtained. However, according to Marten et al. [82], if the numbers of neurons are too large, SOM may end up with over classification. On the other hand, the numbers of neurons required is unknown. Further detail on SOM, its comparison with other method and SOM's efficiency as a classifier for Chinese handwriting can found in [83-85].

4 Conclusion and Future Work

As we have attempted to demonstrate by providing a brief overview of the work in writer identification based on Chinese handwriting including the strength and weaknesses of related studies. The aim of the paper is to provide substantial review on producing efficient Chinese writer identification task, which has been done in the field since 1988. Chinese handwriting identification is a very active and multifaceted domain that continues to gain the interest of researchers. In this paper, we also summarized and compared the various on-going methods and techniques that commonly selected as best practice for Chinese feature extraction stage and writer identification classifiers.

From the review, we acknowledged that, classification phase is essential for identification task besides pre-processing and feature extraction. This phase normally required higher computation expenses and time consuming when a huge data is used. For instances, most of handwriting is performed in a variety of style. Each of the handwriting is unique. Thus, to deal with the distinctiveness of the features, a method needs a large training data samples to adapt and learn about the writer's handwritten text [33]. For this intention, some of the method consume time and cost during computation process. Another issues regarding to misclassification. Sometimes, abnormal written characters misguide the system to classify into inappropriate class. Or in other words, any letter of a class can be treated as a member of another class. Therefore, to overcome these circumstances, further studies need to be carried out to discover the best classifier in order to reduce the computation cost and maintain the efficiency of the classification phases.

Further research currently in progress includes exploring classification phase to deliver a capable writer identification accuracy. However, our hope is that these studies on the state of the art would be useful to researchers in the related field and will serve as a good introduction related to the field undertaken

ACKNOWLEDGEMENTS.

This work is supported in part by Universiti Teknologi Malaysia, Skudai Johor Bahru MALAYSIA. Authors would especially like to thank *Soft Computing Research Group* (SCRG) for their support in making this research a success. Authors would like to thank anonymous reviewers for their fruitful comments.

References

[1] G.R. Ball and S.N. Srihari: Semi-Supervised Learning for Handwriting Recognition. ICDAR 2009: 26-30

[2] G.X. Tan, C.V. Gaudin and A.C. Kot, "Individuality of Alphabet Knowledge in Online Writer Identification**",** IJDAR (2010) Springer Berlin / Heidelberg pp. 1433-2833

[3] C-Y. Low, A.B.J. Teoh and T. Connie: Fusion of LSB and DWT Biometric Watermarking Using Offline Handwritten Signature for Copyright Protection. ICB 2009: 786-795

[4] E. Marcu, "Method of Combining the Degrees of Similarity in Handwritten Signature Authentication Using Neural Networks", *Research and Development in Intelligent,* Springer-Verlag London, 2010, p. 481

[5] R. Narayanswamy, G.E. Johnson, P.E.X. Silveira, and H. B. Wach, "Extending the Imaging Volume for Biometric Iris Recognition,"*Applied Optics*, Feb. 2005

[6] K.W. Bowyer, S.E. Baker, A. Hentz, K. Hollingsworth, T. Peters and P.J. Flynn, "Factors That Degrade the Match Distribution In Iris Biometrics", *Identity in the Information Society,* December 18, 2009

[7] T. Bourlai and J. Kittler, "On Design and Optimisation of Face Verification Systems that are Smart –Card Based" *Machine Vision and Applications,* February 11, 2009

[8] E. Norouzi, M.N. Ahmadabadi and B.N. Araabi, "Attention control with reinforcement learning for face recognition under partial occlusion", *Machine Vision and Applications,* Springer-Verlag 2010

[9] D.Bayly, M. Castro, A. Arakala, J. Jeffers and K. Horadam, "Fractional Biometrics: Safeguarding Privacy in Biometric Applications", Int. J. Inf. Secur. (2010) 9:69–82, Springer, Melbourne, Australia

[10] A.K. Jain, Fellow, IEEE, R.P.W. Duin, and J. Mao, Senior Member, IEEE," Statistical Pattern Recognition: A Review", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22 (1), January 2000

[11] S. Aksoy, 2010. Introduction to Pattern Recognition. Lecture Notes Comput. Sci., CS 551, Spring 2010

[12] M. Bashir and J. Kempf, "Person Authentication with RDTW based on Handwritten PIN and Signature with a Novel Biometric Smart Pen Device", IEEE Workshop, pp: 63-68 (2009)

[13] B. Helli and M. E. Moghaddam, "A text-independent Persian Writer Identification based on Feature Relation Graph (FRG)", Pattern Recognition 43: 2199–2209 (2010).

[14] T-H. Su, T. Zhang, D.J. Guan and H.J. Huang: Off-Line Recognition of Realistic Chinese Handwriting using Segmentation-Free Strategy. *Pattern Recognition* 42(1): 167-182 (2009).

[15] K. Berkner and L.L. Sulem, Special issue on Document Recognition and Retrieval 2009, *International Journal on Document Analysis and Recognition*, 2010.

[16] J. Feng and X. Gao, "Study on Chinese Handwriting Identification Based On Texture Analysis", *Proceedings of the 2009 International Conference on Wavelet Analysis and Pattern Recognition, Baoding*, 12-15 July 2009.

[17] T. Davis, "The Practice of Handwriting Identification", *The Bibliographical Society (typography) and the contributors (content) The Library*, 7th series, vol. 8, no. 3 (2007).

[18] S. Gupta, 2008. Automatic Person Identification and Verification using Online Handwriting. Master Thesis**.** International Institute of Information Technology Hyderabad, India.

[19] R. Plamondon and G. Lorette, “Automatic Signature Identification and Writer Verification – the state of the art”, *Pattern Recognition* 22(2) (1989) 107–131

[20] V. Bouletreau, N. Vincent, R. Sabourin, and Emptoz,”Handwriting and Signature:One or Two Personality Identifier?” *Proc.14th Int’l Conf, Pattern Recognition*, pp. 1,758-1,760, Brisbane, Australia, Aug. 1998.

[21] C-H. Tung and E-Y. Jean, “A modified phoneme-based Chinese input method for minimizing conflict code rate”, Computer Standards & Interfaces 31, Elsevier Vol. 31, Issue 2, February 2009, pp. 292-299

[22] F.H. Cheng, “Multi-Stroke Relaxation Matching Method for Handwritten Chinese Character Recognition”, *Pattern Recognit.* 31(4), pp. 401–410 (1998)

[23] G. Nagy, "Chinese character recognition, A twenty five years retrospective", *Proceedings of ICPR*, pp. 109-114, 1988.

[24] A.D. Cheok, J. Zhang and C.E. Siong: Efficient mobile phone Chinese optical character recognition systems by use of heuristic fuzzy rules and bigram Markov language models. Appl. Soft Comput. 8(2): 1005-1017 (2008)

[25] P.H. Wu, 2003. Handwritten Character Recognition. Degree of Bachelor Thesis. The University Of Queensland, School of Information Technology and Electrical Engineering.

[26] B.Li, Z.Sun, and T.N.Tan. “Hierarchical Shape Primitive Features for Online Text-independent Writer Identification”, *Proc. of 2th ICB*, pages 201–210, 2007.

[27] G. Zhu, X. Yu, Y, Li, and D. Doermann, Language identification for handwritten document images using a shape codebook. *Pattern Recogn.* 42, 12 (Dec. 2009), 3184-3191.

[28] S. Gupta and A.M. Namboodiri, “Repudiation Detection in Handwritten Documents “, *Proc of The 2nd International Conference on Biometics (ICB'07),* PP. 356-365 Seoul, Korea, 27-29 August, 2007.

[29] J. Hochberg, K. Bowers, M. Cannon and P. Kelly, Script and language identification for handwritten document images, Int. J. Document Analysis and Recognition 2 (2-3) (1999) 45–52.

[30] K. Wang, J. Jin and Q. Wang. High Performance Chinese/English Mixed OCR with Character Level Language Identification. *The 10th International Conference on Document Analysis and Recognition*, July 26-29, Barcelona, Spain, pp. 406-410.

[31] B. Li and T. Tan, "Online Text-independent Writer Identification Based on Temporal Sequence and Shape Codes", *Document Analysis and Recognition, 2009. ICDAR '09. 10th International Conference*, pp.931 – 935, 02 October 2009

[32] W. Jin, Y. Wang, and T. Tan, "Text-independent writer identification based on fusion of dynamic and static features," in International Workshop Biometric Recognition Systems, p. 197, 2005.

[33] G. Kim and V. Govindaraju, "A Lexicon Driven to Handwritten Word Recognition for Real Time Applications", *IEEE Trans. Pattern Anal. Machine Intell.* 19 (4), pp. 366-379 (1997)

[34] A.K. Muda, S.M. Shamsuddin. and M. Darus, "Invariants Discretization for Individuality Representation in Handwritten Authorship," *International Workshop on Computational Forensic (IWCF 2008)*, LNCS 5158, Springer Verlag, pp. 218- 228.

[35] Y. Yan, Q. Chen, W. Deng and F. Yuan, "Chinese Handwriting Identification Based on Stable Spectral Feature of Texture Images" *International Journal of Intelligent Engineering and Systems*, Vol.2, No.1, 2009

[36] W. Jin, Y. Wang, and T. Tan, "Text-Independent Writer Identification based on Fusion of Dynamic and Static Features", in *International Workshop Biometric Recognition Systems*, pp. 197, 2005.

[37] Z.Y. He and Y.Y. Tang, "Chinese Handwriting-based Writer Identification by Texture Analysis," Proceedings of International Conference of Machine Learning and Cybernetics, Vol. 6. 2004, pp. 3488 – 3491.

[38] S. Sasi, L. Schwiebert and J. S. Bedi, " Wavelet Packet Transform and Neuro-Fuzzy Approach to Handwritten Character Recognition ", From internet. 1997.

[39] X. Li and X. Ding: Writer Identification of Chinese Handwriting Using Grid Microstructure Feature. ICB 2009: 1230-1239.

[40] A. Schlapbach and H. Bunke, Off-line Handwriting Identification Using HMM based Recognizers, in: *Proceedings of the 17th International Conference on Pattern Recognition*, vol. 2, 2004, pp. 654–658.

[41] C.-L. Liu, "Handwritten Chinese character recognition: effects of shape normalization and feature extraction", in *Arabic and Chinese Handwriting Recognition*, pp. 104-128, 2008.

[42] X. Gao, L. Jin, J. Yin, J. Huang: A New Stroke-Based Directional Feature Extraction Approach for Handwritten Chinese Character Recognition. ICDAR 2001: 635-639

[43] C.-L. Liu, "Normalization-Cooperated Gradient Feature Extraction for Handwritten Character Recognition", *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol.29, No. 8, August 2007.

[44] A. Kothari and A. Keskar, “Rough Set Approaches to Unsupervised Neural Network Based Pattern Classifier”, Lecture Notes Advances in Machine Learning and Data Analysis in Electrical Engineering, Springer Netherlands, Vol. 48. October 27, 2010, pp.151-163.

[45] K. Buza and L. Schmidt-Thieme, “Motif-Based Classification of Time Series with Bayesian Networks and SVMs”(eds.), *Advances in Data Analysis, Data Handling and Business Intelligence*, Studies in Classification, Data Analysis, and Knowledge Organization, Springer-Verlag Berlin Heidelberg 2010.

[46] B. Huang, Y.B. Zhang, M.T. Kechadi: Preprocessing Techniques for Online Handwriting Recognition. Intelligent Text Categorization and Clustering 2009: 25-45.

[47] G.S. Peake and T.N. Tan, “Script and Language Identification from Document Images”, *Proc. BMVC ‘97*, Essex, UK, Sept. 97, Vol. 2, pp.610-619.

[48] N. Arica and F.T. Yarman-Vural, "An Overview of character recognition focused on off-line handwriting", IEEE Transactions on System.Man.Cybernetics-Part C: Applications and Reviews, vol. 31, no. 2, pp. 216-233, 2001.

[49] V-C. Juan, A-C. Carlos, “Font Recognition by Invariant Moments of Global Textures”, In Proceedings of International Workshop VLBV05 (Very Low Bit-Rate Video-Coding 2005). 15-16 September 2005. Sardinia, Italy.

[50] A. Rosenfeld, “Characterization of Parallel Thinning Algorithms”, *Inform and Control*, 29, pp. 286-291 (1975)

[51] M. Khaled Abdl and S. Zaiton Hashim, 2010, “Swarm-Based Feature Selection for Handwriting Identification”, *Journal of Computer Science* 6(1), pp. 80-86, 2010.

[52] R. Sarkar, N. Das, S. Basu, M. Kundu, M. Nasipuri and D. K. Basu, “Word level Script Identification from *Bangla* and *Devanagri* Handwritten Texts mixed with Roman Script”, Journal of Computing, Vol. 2, Issues 2, ISSN 2151-9617, 2010

[53] S. Abirami and D. Manjula: Feature string-based intelligent information retrieval from Tamil document images. IJCAT 35(2/3/4): 150-164 (2009)

[54] A. Bhardwaj, H. Cao and V. Govindaraju, "Script Identification of Handwritten Images", Document Recognition and Retrieval XVI, edited by Kathrin Berkner, Laurence Likforman-Sulem, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7247, 2009

[55] Z. He, X. You, Y.Y. Tang, Writer identification of Chinese handwriting documents using hidden Markov tree, *Pattern Recognition* 41 (2008) 1295–1307.

[56] F. Shahabi and M. Rahmati, "Comparison of Gabor-based Features for Writer Identification of Farsi / Arabic Handwriting", *Proc. of 10th Int'l Workshop on Frontiers in Handwriting Recognition (IWFHR 2006)*, 23-26 October, La Baule, France

[57] K. Yu, Y. Wang, and T. Tan, "Writer identification using dynamic features". *Pro.of ICBA*, pages 512–518, July 2004.

[58] B. Zhang and S. N. Srihari, "Analysis of Handwriting Individuality Using Word Features," *Proceedings of the Seventh International Conference of Document Analysis and Recognition, 2003*, pp. 1142 - 1146.

[59] H. Liu, Hussain, F., Tan, C.L., Dash, M.: Discretization: An enabling technique. Data Mining and Knowledge Discovery 6, 393–423 (2002)

[60] K. Waiyamai, T. Rakthanmanon, P. Pongaksorn, *" DCR: Discretization using Class Information to Reduce Number of Intervals "*,QIMIE'09: Quality issues, measures of interestingness and evaluation of data mining models,2009

[61] G.J. Hwang and F. Li, "A Dynamic Method for Discretization of Continuous Attributes". (Eds.): IDEAL 2002, LNCS 2412, Springer-Verlag Berlin Heidelberg , pp. 506-511.

[62] E. Frank, and I. Witten, "Making Better Use of Global Discretization", *Proceedings of the 16th International Conference on Machine Learning (ICML 1999)*, 115-123.

[63] G**.** Agre and S**.** Peev**.** On Supervised and Unsupervised Discretisation. CIT: Cybernetics and Information Technologies, Vol. 2, No. 2, Sofia, 2002, pp. 43-57.

[64] S. Mehta, S. Parthasarathy and H. Yang**, "**Correlation Preserving Discretization**",** *In Proceedings of ICDM'2004*, pp. 479-482.

[65] J. Dougherty, R. Kohavi and M. Sahami, "Supervised and Unsupervised Discretization of Continuous Features", *In: Proc of the 12th Int. Conf. on Machine Learning*, Tahoe City, CA, July 9–12, pp. 194–202 (1995).

[66] M.R. Chmielewski and J.W. Grzymala-Busse, "Global Discretization of Continuous Attributes as Preprocessing for Machine Learning", *Proceedings*

of the 3rd International Workshop on Rough Sets and Soft Computing, pp. 294-301, 1994.

[67] Y. Yang, and G.I. Webb, “Non-Disjoint Discretization for Naïve-Bayes Classifiers”, *Proceesings of the 19th International Conference on Machine Learning (ICML-2002)*, 666-673.

[68] A. Bagirov, A. Rubinov, N. Soukhoroukova and J. Yearwood, “Unsupervised and Supervised Data Classification Via Nonsmooth and Global Optimization”, Top, Vol. 11, Number 1, 1-93, Sociedad de Estadistica Operativa, Madrid, Spain, 2003.

[69] Y. Zhu, T. Tan and Y. Wang, “Biometric Personal Identification Based on Handwriting”, *Pattern Recognition, Proceedings. 15th International Conference*, Vol. 2, pp. 797 - 800, 2000.

[70] M. Bulacu and L.Schomaker, “Combining Multiple Features for Text-Independent Writer Identification and Verification”, *Proc. of 10th Int’l Workshop on Frontiers in Handwriting Recognition (IWFHR 2006)*, 23-26 October, La Baule, France.

[71] S. K. Chan, C. Viard-Gaudin, and Y. H. Tay, "Online Writer Identification using Character Prototypes Distributions," in Proceedings of SPIE - The International Society for Optical Engineering, 2008.

[72] H.E.S. Said, T. Tan and K. Baker, “Writer identification based on handwriting”, Pattern Recognition 33 (1), pp. 133–148 (2000).

[73] A. Strehl and J. Ghosh, “Clustering Ensembles – A Knowledge Reuse Framework for Combining Multiple Partitions”, *The Journal of Machine Learning Research* 3 583–617 (2002).

[74] X.Z. Fern and C.E. Brodley, “Clustering ensembles for high dimensional data clustering”, in: International Conference on Machine Learning, pp. 186– 193, 2003.

[75] F**.** Yin and C.-L. Liu**, “**Handwritten Text Line Segmentation by Clustering with Distance Metric Learning”**,** *Proc. 11th ICFHR*, Montreal, Canada, pp. 229 234, 2008.

[76] B.E. Boser, I.M, Guyon and V.N. Vapnik, “A Training Algorithm for Optimal Margin Classifiers”, *In*: *COLT ’92: Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. New York, NY, USA: ACM Press, pp. 144–152, 1992.

[77] E.M. Gertz and J.D. Griffin, “Using an Iterative Linear Solver in an Interior-Point Method for Generating Support Vector Machines”, Comput Optim Appl. 2009

[78] F. Chang, "Techniques for Solving the Large-Scale Classification Problem in Chinese Handwriting Recognition", (Eds.): Sach 2006, LNCS 4768, pp. 161–169, 2008.

[79] F. Jinsong and F. Tingjian, "Chinese Character Classification based on Rough Set and SVM Algorithm", *MVA2000 IAPR Workshop on Machine Vision Applications*, pp. 28-30, 2000.

[80] A. Fred and A.K. Jain, "Combining multiple clusterings using evidence accumulation", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (2005) 835–850.

[81] L. Schomaker, M. Bulacu and K. Franke, "Automatic Writer Identication Using Fragmented Connected-Component Contours", In: F. Kimura & H. Fujisawa, Proc. of 9th IWFHR, Japan, Los Alamitos: IEEE Computer Society, pp. 185-190, 2004.

[82] G. Martens**,** C. Poppe, P. Lambertand, R.V. Walle, "Unsupervised Texture Segmentation and Labelling using Biological Inspired Features"**,** *IEEE 10th Workshop on. Multimedia Signal Processing*, pp.159-164, 2008.

[83] J.X. Dong, A. Krzyzak and C.Y. Suen, "High Accuracy Handwritten Chinese Character Recognition using Support Vector Machine", *Proc. Int. Workshop on Artificial Neural Networks for Pattern Recognition*, Florence, Italy, 2003.

[84] C.-C. Chang, C.-J. Lin, LIBSVM: A Library for Support Vector Machines, Available: http://www.csie.ntu.edu.tw/_cjlin/libsvm (2001).

[85] D. Deng, K.P. Chan, and Y. Yu, "Handwritten Chinese Character Recognition using Spatial Gabor Filters and Self Organizing Feature Maps", *Proc. IEEE Inter. Confer. on Image Processing,* vol. 3, pp. 940-944, Austin TX, 1994.

[86] V. Roy and S. Madhvanath, "A Framework for Adaptation of the Active DTW Classifier for Online Handwritten Character Recognition", *10th International Conference on Document Analysis and Recognition,* IEEE ICDAR (2009).

[87] Amir Atapour Abarghouei, Afshin Ghanizadeh & Siti Mariyam Shamsuddin, Advances of Soft Computing in Edge Detection, *International Journal of Advances in Soft Computing and Its Applications,* Vol. 1, No. 2 (2009), 162 - 199 .