Int. J. Advance Soft Compu. Appl, Vol. 11, No. 3, November 2019
ISSN 2074-8523

Biometric Identification based Scores Discretization on Multi Forms of Image Structures

Wong Yee Leng and Siti Mariyam Shamsuddin

School of Computing,
Universiti Teknologi Malaysia, Johor, Malaysia
Email: nureiliyah@utm.my, sitimariyams@gmail.com

Abstract

Multi-Biometric Identification is one of a well-known biometric in the area of pattern recognition and has always been under study through its important role in forensic science that could help government criminal justice community. In this paper, a novel multi-biometric identification of individuals by means of both physiological and behavioural biometrics is introduced. Different from the most conventional biometric identification, the extracted physiological and behavioural biometrics will go through a proposed Multi-Biometric Feature Discretization. The intention of Discretization in this study is to prevail over the deficiencies caused by poor features mining and to attain individual unique features that could reflect the individual varianceness in order to discriminate one person from another. The experiments was conducted on real-world multiple form of images which is adopted from INF/ENIT, HITMW and FVC databases. The experimental results supported our analysis by demonstrating a remarkable potential of the new structure of multi-biometric identification adaptability on multi forms of image structures and has excellent potency to conserve the distinctiveness of individual during identification.

Keywords: *Pattern Recognition, Machine learning, Data Mining, Identification, Multi-Biometrics, Feature extraction, Discretization*

1 Introduction

Data mining and machine learning both are an established field of extracting potentially useful knowledge from huge database including the process of identifying the comprehensible patterns in data [1,2] and has become a research focus recently, especially in industrial applications [3-5]. In this paper, our main work is on biometric datasets, thus our discussion will be more on biometric datasets of data mining tasks which often involve continuous huge attributes. Thus, a discretization is introduced in this study to discretize those continuous input features. Discrete features compared to continuous features are easier to describe, understand because the datasets are well represented and simplified. This quality discretization of discrete attributes is an important criterion that ensures high performance on speed, accuracy and understandability of the classification models instead of continuous attributes [6]. In addition to this, discrete representation more to knowledge level concept which is practically useful for most of the pattern recognition applications. We believe that data mining algorithm using the discretized datasets can significantly impact the performance of individual identification used in the analysis of high dimensional based biometrical data.

A biometric identification is one of the established system that uses the concept of pattern recognition technology. Pattern recognition is a broad term which cover three common types of application fields; identification [7], verification [8] and recognition [9]. In order to serve for both academia and industry including potential applications, from security, forensics to archaeology activities, the scope of pattern recognition systems has been broaden up to a biometric based system. This includes identification and verification based on signature [10,11], iris [12,13], facial features [14,15], fingerprints [16] and handwriting [7,8].

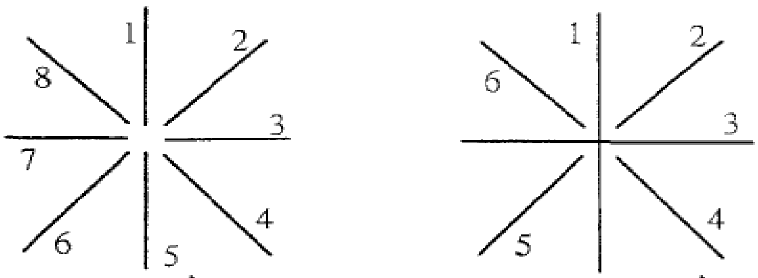
2 Analysis of Behavioral and Physiological Biometrics

Biometric term can be categorized into two parts: behavioral biometrics that require individual characteristics of a person's behaviour for identification process (e.g. handwriting, signature, speech, gait) and physiological biometrics that gauge a physical features of the human body to perform identification (e.g. fingerprint, iris, face, DNA). In this paper, the proposed multi-biometric identification pertains to both categories of biometric properties; physiological and behavioural biometrics. The behavioral identification is based on Arabic and Chinese handwriting. They are discussed in sub section 4.1. As for physiological identification, our focus is on fingerprint; an overview of fingerprint structure can be read in sub section 4.2.

2.1 Arabic and Chinese handwriting

Handwriting is an everyday form of communication for human being and commonly treated as one of the behavioral biometric which play a significant role in biometric system, a fundamentally pattern recognition system that acquire biometric data from an individual, extracting a feature set from the data or images, comparing the feature set against the template set in database and finally make decision on the pattern of feature set. In the present, most of research challenges and techniques in handwriting analysis are based on soft computing methods as described in [21] which idea on the fact that handwriting identification starts with discovery the basic features that are gathered together in order to identify the individuality of writer. Arabic, the second Semitic alphabetic writing system most widely used in the world besides Latin alphabet (the most widespread). Unlike Latin writing system, the writing direction of Arabic is performed from right to left. The most noticeable characteristics of the Arabic writing is that Arabic scripts are intrinsically cursive and Arabic alphabet composes of 28 basic letters which consists of strokes and dots (one, two, or three). Thirteen of Arabic characters don't have dots, ten of them have one dot, three have two dots, and two have three dots. Dots can be above (خ,ض), below (.ب, ي), or in the middle (ج) of the letter [31]. These dots and other diacritical marks intentionally for character vowelisation create ligatures that are difficult to segment during pre-processing [31].

Besides Arabic, complication structure also applied in Chinese character. 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 [32,33]. According to Fang, the complication structure in Chinese character mostly affected by multi stoke of each character. The characters may consist of one to thirty or more distinct strokes due to the variety of handwriting style. Fig. 1 shows the same Chinese character with different writing styles which has created a different number of strokes called multi strokes. Fang's proposed algorithms successfully unite some sub strokes into the proper and complete stroke to determine a Chinese character.



(a) 米 (b) 米

Fig.1 Same Chinese character with different number of strokes

2.2 Fingerprint

A fingerprint pattern is formed of ridges and valleys over the surface of finger (Fig. 2). Different shape and structure of ridge and valley in each finger of an individual contributes to different global and local analysis. A global analysis is use to extract the common singular points namely loop, delta, whorl and etc. Fig. 3 shows examples of singular points such as core point, delta, loop and whorl. The core points or small circles shown in Fig. 3 is the center point of the highest loop in the singular region and generally use as a features to be pre aligned with other fingerprint pattern during matching process. These singular regions may be classified into five classes: left loop, right loop, whorl, arch and tented arch as illustrated in Fig. 4.

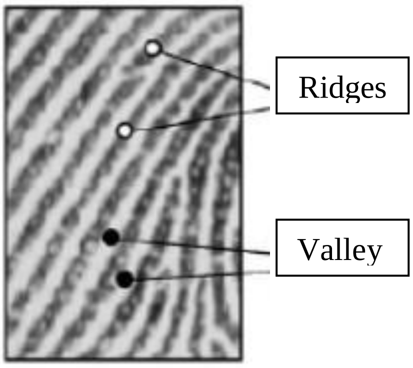


Fig.2 Ridges and valleys on a fingerprint image

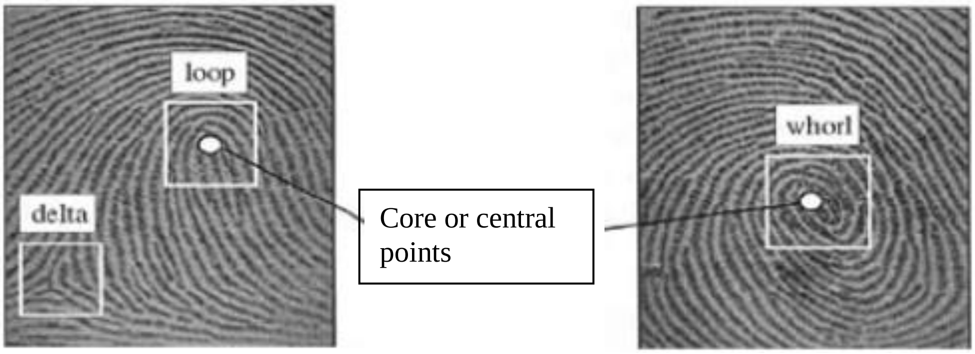


Fig.3 Singular regions and core points

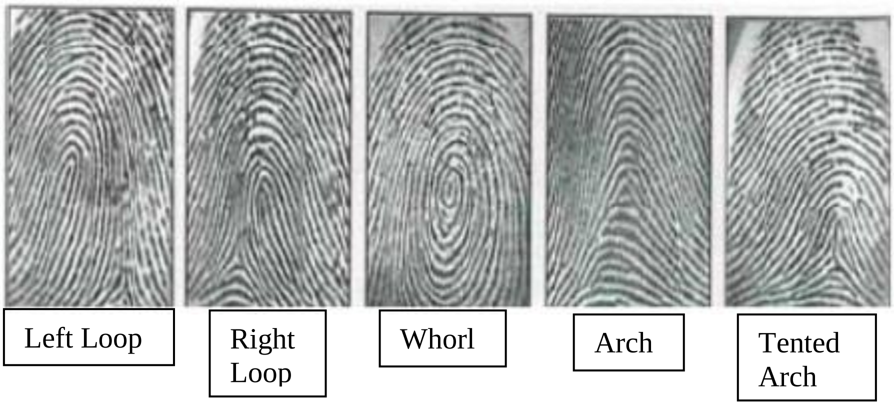
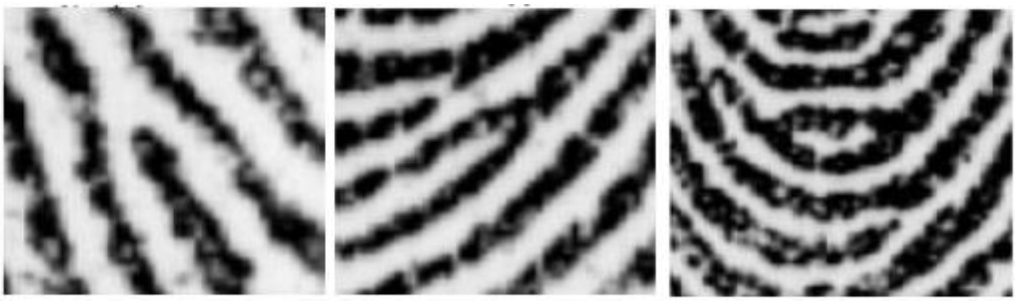


Fig.4 Fingerprint classes

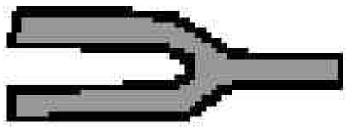
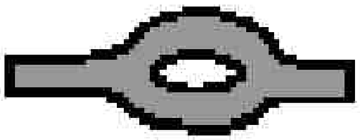
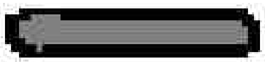
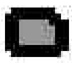
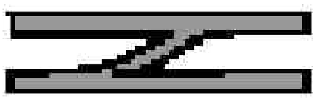
In summary, global analysis provides an overall picture of fingerprint classification. Local analysis, on the other hand, gives more detail information of the image. These significant information including observe the regions of the ridge termination or ridge anomalies termed minutiae points. Some of typical ridge discontinuous structures are shown in Fig. 5. The ridges that suddenly break or discontinuous at the end are called termination. A terminate ridge with two split ridges at the end is called bifurcations. Fig. 5 shows a closer observation of real fingerprint image. We can see a ridge ending in Fig. 5 (a), a ridge bifurcation in Fig. 5(b). Fig. 5 (c) shows a lake. More details on different types of minutiae can be seen in Table 1.



(a) (b) (c)

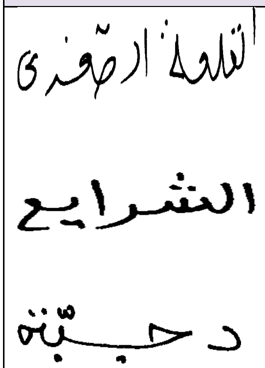
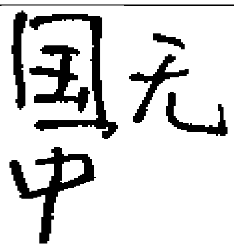
Fig.5 Types of Minutiae (a) ridge termination or ridge ending (b) ridge bifurcation and (c) a lake

Table 1. Types of Minutiae

Minutiae	**Types**
	**Termination**
	**Bifurcation**
	**Lake**
	**Independent Ridge**
	**Dot or Island**
	**Spur**
	**Crossover**

Overall, behavioural biometrics discussed here (Arabic and Chinese handwriting) can be expressed in variety of style, which is commonly based on the alignment and formation of multi directional, dots and stokes pattern. Naturally, the style of Arabic and Chinese character can be seen differently based on static, dynamic [34] 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 dots and stroke construction to write a certain character, the number of writing style of a character even by same person can be perpetuity. These features are those important points to be thought of when dealing with determining a writer of handwriting sample from a set of writers. The main challenge for behavioural identification is when dealing with different style of handwriting performed by the same author and how actually to acquire features that can reflect the authorship of the author. Same goes to the physiological biometrics identification as discussed in previous section. Table 2 summarizes the characteristics of behavioral and physiological biometric images.

Table 2. Characteristics of Behavioral and Physiological biometric images.

<table>
<tr><th>Biometric images</th><th colspan="2">Natural characteristic of Biometric images</th></tr>
<tr><td>
DB: IFN/ENIT</td><td>Dots and other diacritical marks intentionally for character vowelisation create ligatures that is difficult to segment during pre processing [31]</td><td rowspan="2">Cursive nature of characters itself leads to variability between curve angles, shape and size
(Those patterns vary greatly among different writers and even for same writer)
Higher similarities among features</td></tr>
<tr><td>
DB: HIT/MW</td><td>Multiple strokes structures [35]
Difficult to determine various unique features of individuality</td></tr>
</table>

 **DB: FVC 2000,2002,2004**	• **Formed of ridges and valleys over the surface of finger**	• **Two general approaches: minutiae based [36] and correlation based approach; encounter overall patterns of ridges and valleys including distort regions found in the ridges**

3 Problem in existing methods

3.1 Arabic and Chinese Identification

It is generally accepted that most of handwritten Arabic and Chinese identification is a difficult problem especially due inherently to the cursive nature of Arabic characters itself, which leads to a lot of variability between curve angles, shape and size though in this paper we are not focusing on this issue. However, those features still considered as one of the identification complications to identify the individuality for same writer. To the best of our knowledge and assumptions, we conclude a few common categories of difficulties in a writer identification for Arabic and Chinese handwriting.

First, the Arabic writing styles include the shapes of the descenders and even on Arabic calligraphic styles can be performed more than a dozen. Table 3 summarizes the Arabic characters in isolated, initial, medical and final form. Isolated Arabic characters could be performed into other words when those characters are joined from right with other characters [37]. Those patterns vary greatly among different writers and even for same writer. Second, Arabic characters composes of many dots and other diacritical marks intentionally for character vowelisation create ligatures that is difficult to segment during pre-processing [38-41]. Third, the Arabic can be splits into subwords, also known as Parts of Arabic Words (PAWs) and PAWs create many problems [42]. Finally the quality of the scanned handwritten images in particular for highly degraded historical images in some real world applications could also lead to specific problems.

Table 3. Four forms of Arabic characters [37]

No	Letter Name [a]	Isolated Form	Initial Form	Medial Form	Final Form	No	Letter Name	Isolated Form	Initial Form	Medial Form	Final Form
1	Alef [b, c]	ى ا	-	-	ـى ـا	16	Tah	ط	طـ	ـطـ	ـط
2	Beh	ب	بـ	ـبـ	ـب	17	Zah	ظ	ظـ	ـظـ	ـظ
3	Teh [d]	ة ت	تـ	ـتـ	ـة ـت	18	Ain	ع	عـ	ـعـ	ـع
4	Theh	ث	ثـ	ـثـ	ـث	19	Ghain	غ	غـ	ـغـ	ـغ
5	Jeem	ج	جـ	ـجـ	ـج	20	Feh	ف	فـ	ـفـ	ـف
6	Hah	ح	حـ	ـحـ	ـح	21	Qaf	ق	قـ	ـقـ	ـق
7	Khah	خ	خـ	ـخـ	ـخ	22	Kaf	ك	كـ	ـكـ	ـك
8	Dal [b]	د	-	-	ـد	23	Lam	ل	لـ	ـلـ	ـل
9	Thal [b]	ذ	-	-	ـذ	24	Meem	م	مـ	ـمـ	ـم
10	Reh [b]	ر	-	-	ـر	25	Noon	ن	نـ	ـنـ	ـن
11	Zain [b]	ز	-	-	ـز	26	Heh	ه	هـ	ـهـ	ـه
12	Seen	س	سـ	ـسـ	ـس	27	Waw [b]	و	-	-	ـو
13	Sheen	ش	شـ	ـشـ	ـش	28	Yeh	ي	يـ	ـيـ	ـي
14	Sad	ص	صـ	ـصـ	ـص	29	Hamza [e]	ء	ئـ	ـئـ	أ
15	Dad	ض	ضـ	ـضـ	ـض						

Besides Arabic, complication of Chinese character is due to the alignment and formation of multi stokes pattern. Chinese characters are well known as large alphabet language; it has a variety of categories. Numerous variation of stroke construction to write a certain Chinese character, the number of writing style of a Chinese character can be perpetuity even from same writer.

Natural characteristic of Arabic and Chinese writing itself, which surrounded by multifaceted mechanism and various complications leads to many problems unsolve and most of the identification system tends to be not effective enough. Identification based on cursive handwriting is very difficult and the overall issues can be summarized as follows. First, difficult to obtain discriminant handwriting features as handwritings are usually not stable enough and have vast variability, so this creates a large amount of features data consequently slow down the computation progression and misclassification often occurs during learning and classification. Misclassification often happen causes by overlapping feature space. Second, too much of pre-processing at early stage such as normalisation to discard mixes noise of the character in order to refine classification would almost certainly trade off important individualistic of the writer [43]. This creates a poor classification which leads to a low identification rate either in terms of its capability and competency. Third, real world classifications often involve with continuous features. Unfortunately most of machine learning and data mining method such as Bayesian Network could not perform well with continuous features unless the continuous features are first pre-processed. This process is called Discretization, which will be discussed in more detail in next section. Moreover, in real world writer identification problems the number of writers to be identified is often unlimited and each writer has only few reference handwriting samples, extremely, only one reference handwriting sample served as training samples.

3.2 Fingerprint Identification

From an extensive research of available literature, it was known that the popular fingerprint approaches are classified into two general classes: minutiae based approach and correlation based approach. Both approaches have some of its own shortcomings. The minutiae-based techniques first looks for the minutiae points and then match their chosen regions in a query finger with the stored patterns in database. Different fingerprint images have different number of minutiae points. The best total number of minutiae points for a good quality of fingerprint is around 60 to 80 minutiae points [44]. A disadvantage of minutiae-based technique is that it does not mend well to indexing mechanism.

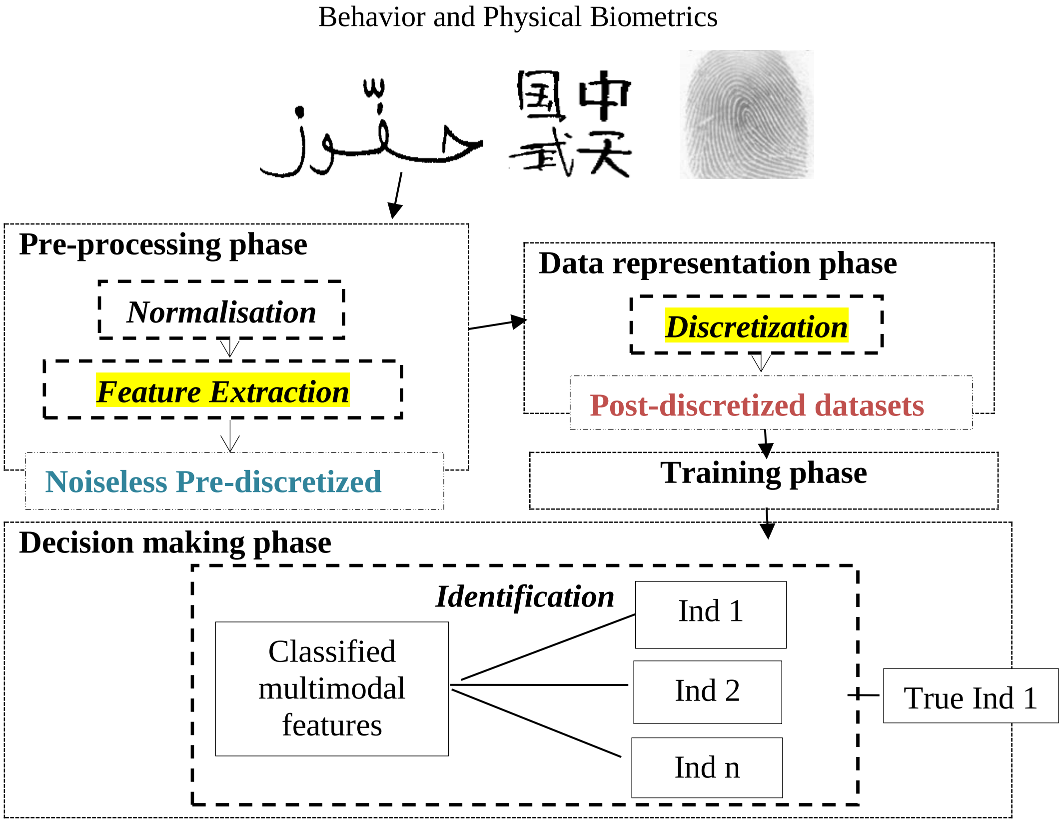
Second approach called Correlation-based techniques [45,46], which consider the overall patterns of ridges and valleys found in a finger. These approaches encounter all information including distort regions found in the ridges. Thus, these ideas inferred to a noise problem as encountered in signal or image processing applications.

With regard to the fingerprint matching disadvantages (minutiae-based ones and correlation based approach), a great number of solutions have been proposed such as employing normalized cross correlation technique which can reduces the computational effort and decreases the error rate better than the minutiae based approach [47]. Meanwhile, in [48], they proposed a method that uses ridges joint with minutia. With this technique, the template cand query fingerprints can be easily aligned.

4 Related Work

Currently, many researchers have done great work enhancing feature extraction and classification methods as mentioned involves handwritten processing including normalization [59-60], feature extraction [61-62], and classification techniques based on soft computing methods including neural network [63] and rough set approaches. Only a few works focuses on data analysis and data representation. Besides understand and able to extract useful object’s property at pre-processing phase, filter and intelligently discriminate unique features to represent data in a comprehensible way is important because this may intensify the individuality performance of identification process.

As such, we enlarge the scope of the possible use of the common biometric identification which has been, up to now, mainly evaluated on handwritings into biometrics identification coupled with the proposed Multi-Scores based Discretization which can work on any type of biometrics property in order to produce a wide variety of potential applications, from security, forensics and financial activities to archaeology. Framework of the Multi-Biometric Identification based Discretization as illustrated in Fig.5.



4.1 **Fig.5 Multi-Biometrics Identification based Discretization**

It is generally accepted that most of handwritten Arabic and Chinese identification is a difficult problem especially due inherently to the cursive nature of Arabic characters itself, which leads to a lot of variability between curve angles, shape and size though in this paper we are not focusing

To achieve the objective of this study, previous feature extraction and discretization methods which is not restricted to any particular types of biometric behaviour and physiological property are enhanced. The concept of Discretization for data representation is adopted and the algorithm is based on Invariant Discretization originally proposed by Azah [64]. This technique has been applied to variety of handwriting types including our previous work on English handwriting

[64] and Chinese handwriting [21]. For this research, we improve Azah's Discretization to fit for multi-biometric images so-called Multi-Scores Discretization (MSD). The improved algorithm is able to deal with large and high dimensional feature vectors as well as to produce significant individual features which directly unique to those individual. Again, the improved Discretization here does not need the use of Normalization process.

In biometric identification, the proposed Multi-Biometric Feature Discretization (MBFD) transforms the multiple continuous feature set of individual closer to discrete feature sets (knowledge level representation). The description in two main steps below explains how the actual feature sets of an individual are discretized in multi-biometric for personal identification.

Step 1: Compute the size of interval, I_{width}. Given a set of features, the discretization algorithm first computes the size of interval, I_{width} as defined in Equation (4), i.e., it determines its upper and lower bounds. The range is divided by the number of features which then gives each interval upper approximation, AV_{upper} and lower approximation, AV_{lower} as demonstrated in Fig.6.

$$I_{width} = (f_{\max} - f_{\min}) / No\_feature \quad (1)$$

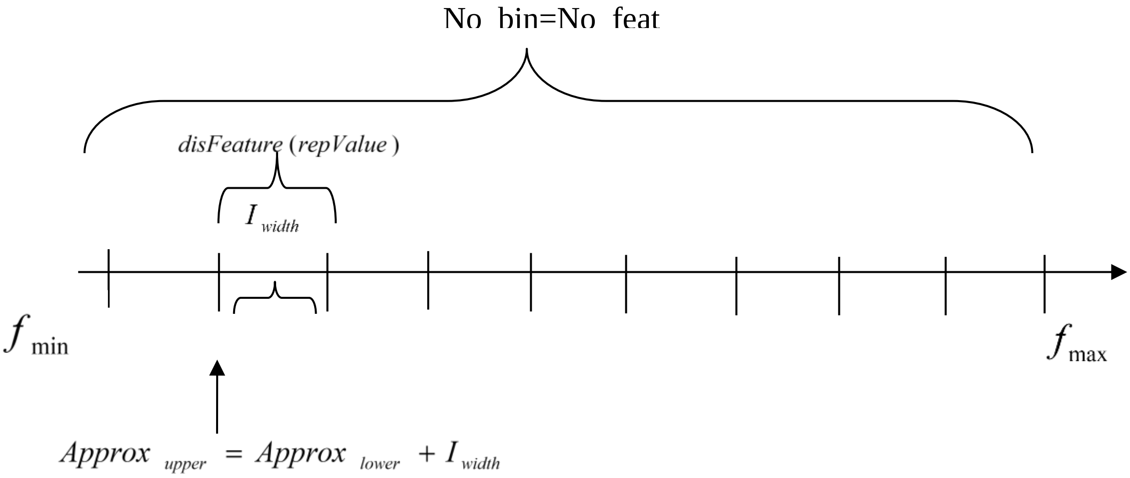


Fig.6 The improved Discretization; Multi-Biometric Feature Discretization.

The number of intervals generated is equal to the dimensionality of the feature vectors. This is to maintain the original number of extracted features from multiple extraction methods.

Step 2: Compute a single discriminatory value, *disFeature* of each interval. Instead of taking the range between the interval, the *disFeature* here is considered by taking the midpoint of the AV_{upper} and AV_{lower} interval as defined in Equation (2) (improvised from Azah's Invariant Discretization).

$$disFeature = \left(\frac{AV_{lower} + AV_{upper}}{2} \right) \tag{2}$$

With this improved procedure in computing intervals, the estimated representation feature values are more close to the actual multi-biometric features distribution (true feature values). This preserves the discriminative power of the original multi-biometric features and enhances the statistical distinctiveness between individuals. Therefore, here, the discretization is said to be robust and efficient enough to handle the issues in feature distribution of Multi-biometric identification. The graphical representation of the improved Discretization compare to normalization and previous Discretization is shown in Fig.7.

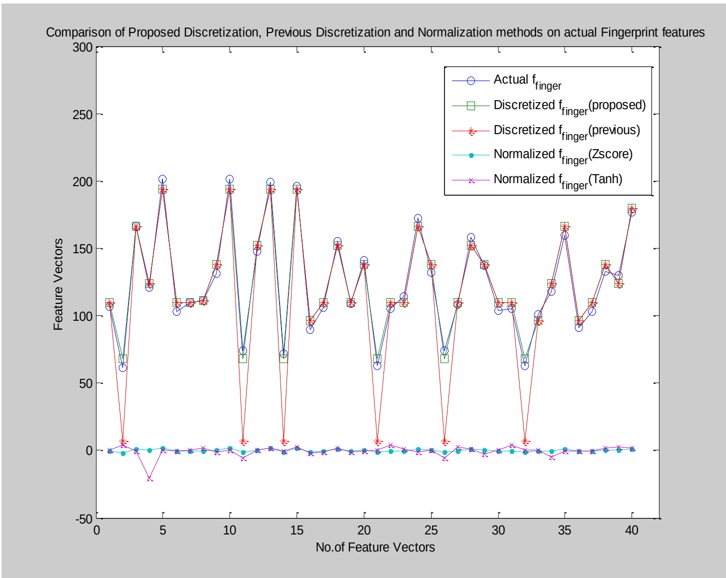


Fig.7 Comparison of the actual, discretized multi-biometric features (proposed), previous discretized and normalized feature distributions on a fingerprint dataset.

Unlike other discretization approaches, the main motivation behind this algorithm is to maximize inter-class distances for all biometric samples that do not belong to the same individual class. By representing the features into a set of intervals, the issues of dimensionality caused by overlapping features can be avoided. This makes the identification process easier and faster due to the easier clarification by the classification and decision tasks. Fig.8 and Fig.9 each presented the samples of pre-discretized datasets for Arabic and Fingerprint, which is the actual extracted datasets before executed with the proposed Multi-Scores based Discretization algorithm. There are nine extracted vectors in each row and the last column represents the participant's class.

0.0027	0.1198	0.0306	0.4324	0.6136	0.9662	0.8006	0.0796	0.1145	1
0.0148	0.0099	0.0142	0.8289	0.0562	0.0314	0.2361	0.0233	0.1537	3
0.0075	0.0837	0.0062	0.5652	0.7277	0.5844	0.0149	0.0142	0.8720	2
0.0029	0.0367	0.0082	0.5854	0.0838	0.2044	0.0371	0.2278	0.0659	1
0.0106	0.0530	0.0075	0.8783	0.1322	0.0292	1.7893	0.1524	0.0629	3
0.0030	0.1742	0.5573	0.0438	0.0373	0.5001	0.1037	0.2453	0.1610	1
0.0135	0.0693	0.0538	0.1839	0.0404	1.4319	0.0190	0.1452	0.0153	3
0.0172	0.0670	-1.0000	0.0353	0.1841	0.0205	0.1077	0.0416	0.3497	2
0.0017	0.0766	0.0562	0.2755	0.0328	0.0850	0.1933	0.2309	0.0028	1
0.0020	0.0120	-1.0000	0.0080	0.0305	0.0099	0.0187	2.0052	0.0939	2
0.0317	0.2651	0.0122	0.7570	0.6109	0.7127	2.3325	0.1340	0.0580	3
0.0045	0.0111	0.0063	0.0402	0.0142	0.3099	0.4702	0.8670	0.0121	2

Fig.8 Samples of Pre-discretized Arabic handwriting written by 3 writers

3.9500	7.0000	6.0000	9.1500	12.7500	16.9500	1.8500	5.8500	10.6000	1
1.9500	7.2000	5.6500	8.0000	14.4000	8.6500	3.4500	7.0000	3.5500	2
11.7000	15.8500	7.5500	14.9000	21.7500	13.8000	8.4500	15.0500	14.7000	3
1.6500	2.5500	2.2500	6.3500	9.0000	5.8000	3.5000	3.5500	9.1000	1
4.9500	7.6000	5.4500	12.1000	17.6000	8.6500	3.8000	7.1000	4.3500	2
3.5500	2.5500	1.2500	8.2000	11.5500	6.7500	4.3000	5.6000	7.3500	1
5.8000	6.7500	6.2000	16.2000	22.3000	15.4500	6.5000	9.7000	3.5000	2
9.8500	16.0500	6.7000	8.5500	7.9000	8.5500	3.9500	5.7000	5.2500	3
4.9500	8.8500	5.5500	13.0000	17.8500	11.5500	4.3500	6.2500	6.7000	2
13.5000	34.9500	22.3500	19.2500	31.7000	27.9000	9.7500	21.1500	17.3000	3
6.1000	8.0000	9.5000	14.0000	18.8000	17.0000	4.5500	7.5000	10.4500	1
4.1000	10.6500	8.1500	19.4500	31.9000	21.8000	10.9000	19.5500	23.5500	3

Fig.9 Samples of Pre-discretized Fingerprint from 3 participants

Fig.10 and Fig.21 show the discretized datasets for Arabic handwriting and fingerprint after process. Discretized datasets are also known as Post-Discretized scores. The circled vectors illustrate the *unique values* obtained from each individual. From Fig. 20, the different colours of circled vectors indicate that each writer has it own style of handwriting and this is demonstrated by its *unique values* even though one writer performs variety styles of handwriting. *Unique values* here used as a discriminative role to portray the uniqueness features of individuality from others.

0.0535833	0.16245	0.0535833	0.48395	0.591117	0.80545	0.0535833	0.16245	1	
0.0535833	0.0535833	0.0535833	0.591117	0.0535833	0.16245	0.0535833	0.269617	0.0535833	1
0.0535833	0.16245	0.591117	0.0535833	0.0535833	0.48395	0.0535833	0.269617	0.16245	1
0.0535833	0.0535833	0.0535833	0.269617	0.0535833	0.0535833	0.16245	0.269617	0.0535833	1
0.129167	0.129167	0.129167	0.911667	0.129167	0.129167	0.129167	0.129167	0.129167	3
0.129167	0.129167	0.129167	0.911667	0.129167	0.129167	1.68667	0.129167	0.129167	3
0.129167	0.129167	0.129167	0.129167	0.129167	1.42833	0.129167	0.129167	0.129167	3
0.129167	0.129167	0.129167	0.653333	0.653333	0.653333	2.20333	0.129167	0.129167	3
0.168689	0.168689	0.168689	0.5026	0.836511	0.5026	0.168689	0.168689	0.836511	2
0.168689	0.168689	0.166956	0.168689	0.168689	0.168689	0.168689	0.168689	0.5026	2
0.168689	0.168689	0.166956	0.168689	0.168689	0.168689	0.168689	1.83824	0.168689	2
0.168689	0.168689	0.168689	0.168689	0.168689	0.168689	0.5026	0.836511	0.168689	2
…………									
…………									

Fig.10 Samples of Post-discretized Arabic handwriting scores by 3 writers

Meanwhile, for fingerprint, different shape and structure of ridge and valley in each finger are the useful information to determine an individual from others. After going through the proposed Multi-Biometric Feature Discretization, the unique values of an individual are obtained as presented in Fig.11 below.

0.975	0.975	0.975	6.125	8.075	6.125	4.175	4.175	10.025	1
4.175	6.125	6.125	10.025	11.975	17.825	0.975	6.125	10.025	1
4.175	0.975	0.975	8.075	11.975	6.125	4.175	6.125	8.075	1
6.125	8.075	10.025	13.925	17.825	4.175	8.075	10.025	8.075	1
1.13056	7.60278	5.34167	7.60278	14.3861	7.60278	1.13056	7.60278	1.13056	2
5.34167	7.60278	5.34167	12.125	16.6472	7.60278	1.13056	7.60278	5.34167	2
5.34167	7.60278	5.34167	16.6472	21.1694	14.3861	7.60278	9.86389	1.13056	2
5.34167	9.86389	5.34167	12.125	18.9083	12.125	5.34167	5.34167	7.60278	2
12.5611	16.0056	16.0056	22.8944	12.5611	9.11667	9.11667	16.0056	16.0056	3
9.11667	16.0056	1.72222	9.11667	9.11667	9.11667	1.72222	1.72222	1.72222	3
12.5611	22.8944	19.45	33.2278	26.3389	9.11667	19.45	16.0056	16.0056	3
1.72222	9.11667	19.45	33.2278	22.8944	9.11667	12.5611	19.45	22.8944	3
…………									
…………									

Fig.11 Samples of Post-discretized fingerprint scores of 3 participants

5 Experimental Results and Discussion

Several experiments were carried out to seek for the efficiency of the generic identification on both types of biometric images using the proposed MBFD in Biometric Identification. The identification performance is compared with the traditional one. To illustrate this, the experiments are done with different classification methods which are based on Rough set theory [65]. Four reduction algorithms based on rough set were chosen from ROSETTA Toolkit namely Johnson’s algorithm, Holte IR algorithm, Genetic algorithm and Exhaustive algorithm. The experiments were conducted on different biometric properties of datasets: physiological and behavioural biometrics as summarized in Table 4 and Table 5 accordingly.

Table 4. Identification accuracy after reduction algorithm

Biometric	Images	ROSETTA Built-in Methods on Reductions	Data Types	70% Training Data 30% Testing Data	60% Training Data 40% Testing Data	50% Training Data 50% Testing Data	Average Acc (%)
Handwriting	Chinese datasets HIT/MW	Johnson's Algorithm	Pre_Dis	11.11	8.33	6.89	8.78
			Post_Dis	94.44	95.83	93.10	94.46
		Holte 1R Algorithm	Pre_Dis	11.11	12.50	10.34	11.32
			Post_Dis	100.00	100.00	100.00	100.00
		Genetic Algorithm	Pre_Dis	11.11	12.50	10.34	11.32
			Post_Dis	100.00	100.00	100.00	100.00
		Exhaustive Algorithm	Pre_Dis	11.11	12.50	10.34	11.32
			Post_Dis	100.00	100.00	100.00	100.00
	Arabic datasets IFN/ENIT	Johnson's Algorithm	Pre_Dis	66.67	75.00	60.00	67.22
			Post_Dis	100.00	100.00	100.00	100.00
		Holte 1R Algorithm	Pre_Dis	66.67	75.00	60.00	67.22
			Post_Dis	100.00	100.00	100.00	100.00
		Genetic Algorithm	Pre_Dis	66.67	75.00	60.00	67.22
			Post_Dis	100.00	100.00	100.00	100.00
		Exhaustive Algorithm	Pre_Dis	66.67	75.00	60.00	67.22
			Post_Dis	100.00	100.00	100.00	100.00
Fingerprint	Fingerprint datasets FVC	Johnson's Algorithm	Pre_Dis	67.22	26.11	32.78	42.04
			Post_Dis	100.00	100.00	58.89	86.30
		Holte 1R Algorithm	Pre_Dis	73.89	26.11	32.78	44.26
			Post_Dis	100.00	100.00	100.00	100.00
		Genetic Algorithm	Pre_Dis	73.89	26.11	32.78	44.26
			Post_Dis	100.00	100.00	100.00	100.00
		Exhaustive Algorithm	Pre_Dis	73.89	26.11	32.78	44.26
			Post_Dis	100.00	100.00	100.00	100.00

From the above-obtained result in Table 4, it clearly shows that the first approach based on Rough set theory, four reduction methods namely Johnson's algorithm, Holte 1R, Genetic algorithm and Exhaustive algorithm perform well with post-discretized datasets. Each method successfully achieved the overall average accuracy of more than 90.0%. From the same result table, it can be seen that the same four algorithms presented the worst performance with pre-discretized datasets, which provide low identification rates; below 75.0%.

Clearer illustration and discussion are summarized in Table 5 below. The comparison with the proposed Multi-Biometric Identification based Discretization is presented. The proposed identification successfully achieves more than 95.0% on an average when compared to the existing system which obtain an average accuracy lower than 50.0% on every biometric datasets. This illustrates that the proposed architecture is much competent than existing system; can extract useful features in much better way during pre-processing phase and able to bear a higher discriminative power of individuality during post-processing phase. This can be seen through the significant improvement in the identification rates.

Table 5.Comparisson of Identification Accuracy on each FVC databases

Classification	**Biometric Datasets**	**Conventional Biometric Identification (%)**	**Novel Multi-Biometric Identification (%) (Proposed Multi-Biometric Feature Discretization)**
Reduction algorithms based on Rough set Theory:	Chinese Handwriting	10.68	98.61
	Arabic Handwriting	67.22	100.00
Johnson's Algorithm Holte 1R Algorithm Genetic Algorithm Exhaustive Algorithm	Fingerprint	43.70	96.57
Overall Acc **(%)**		40.53	98.39

On the whole, to quantify the performance of the proposed algorithm on Multi-Biometric identification in brief, a *t-test* is conducted to illustrate the significance difference between the proposed and the existing one. For hypothesis testing, the scalar mean of an average accuracy for pre and post discretized scores after several training and testing procedure is used as a statistic and basis for one-to-one individual identification performance comparisons of the three biometric datasets.

Given two paired sets; pre-discretize datasets, x_i and post discretized of each selected biometric datasets, y_i of n total datasets for each biometric databases. Examples, if the selected biometric datasets is fingerprint. Then, the given two paired sets would be pre-discretize of fingerprint datasets, x_i and post discretized of fingerprint datasets, y_i of *n* total fingerprint for each FVC database, the paired *t-test* determines if they differ from each other in a significant way. The t value is computed using the mathematical equation (1), (2) and (3) respectively.

$$\hat{x}_i = (x_i - \bar{x}) \tag{1}$$

$$\hat{y}_i = (y_i - \bar{y}) \tag{2}$$

$$t = (\bar{x} - \bar{y}) \sqrt{\frac{n(n-1)}{\sum_{i=1}^{n} (\hat{x}_i - \hat{y}_i)^2}} \tag{3}$$

Table 6. Comparison of Identification Accuracy on each biometric image

Biometric Databases			**Mean**	**Std. Dev.**	***T***	**P**	**Significance**
Handwriting	IFN/ENIT (Arabic)	Pre Post	67.22 100.00	6.41 0.00	9.77234E-10	0.05	Highly Significance
	HIT/MW (Chinese)	Pre Post	10.68 98.61	1.68 2.57	2.22631E-20	0.05	Highly Significance
Fingerprint	FVC2000	Pre Post	72.22 100.00	6.64 0.00	8.19455E-09	0.05	Highly Significance
	FVC2002	Pre Post	26.11 93.06	5.74 16.60	3.71138E-08	0.05	Highly Significance
	FVC2004	Pre Post	32.78 89.72	6.41 18.93	3.49287E-07	0.05	Highly Significance

As per Table 6, when *t* test analysis was done on post-discretized biometric scores and those of actual biometric datasets (pre- discretized datasets), the results were highly significant with *t* 8.19455E-09 and $P<0.05$ for fingerprint datasets. When *t* test analysis was done on IFN/ENIT Arabic datasets, the results were also highly significant with *t* 9.77234E-10 and $P<0.05$. Same analysis achievement on pre and post-discretized HIT/MW Chinese datasets which shown a highly significant with *t* 2.22631E-20 and $P<0.05$, stating that individual identification shows a better performance with discretized datasets as compared to the actual datasets. Thus, the analysis of the data reveals that there is a definite improvement in the individual identification performance after enhancing the previous biometric identification to overcome the discrepancy caused by poor generalization and able to extract useful information from any form structure of biometric property correctly. Moreover, the concept of discretization in biometric is able to achieve lower identification error than those one without the use of discretization.

6 Conclusion

In this paper we presented a study of a complex characteristics of biometric images adopted from three standard databases INF/ENIT, HITMW and FVC databases. The process of feature extraction is presented along with the results. Analysis of the usefulness features was conducted via searching and extraction of the best feature sets using the morphological operations and 3x3 zoned windows method. Beside this, in the most of the individuality analysis of handwriting studied, Azah's discretization method does possess good discriminative power for handwriting identification. This successfully justified their use in identification analysis. Thus, in this study, we enhance and propose a novel discretization to fit for Multi-Biometric identification so called Multi-Biometric Feature Discretization (MBFD). The algorithm handle multiple behaviour and physiological images by search, induce, organize and represent a list of intervals into a better individual informative representation. This algorithm was tested on a few classification methods to seek for its effectiveness on multiple biometric images. The biometric identification based on proposed MBFD extensively boosted up the individual identification accuracy of 99.0% on any form of biometric images and effectively prevails over the existing identification that had positional and oriental weaknesses.

ACKNOWLEDGEMENTS.
Authors would especially like to thank Universiti Teknologi Malaysia, Skudai Johor Bahru MALAYSIA for their excellent cooperation and contributions to improve this paper.

References

[1] U. Fayyad, S.G. Piatetsky and P. Smyth, From data mining to knowledge discovery: an overview.Adv Knowl Discov Data Min 1–34 (1996)
[2] X. Wu, V. Kumar, J.R. Quinlan, J. Ghosh , Q. Yang, H. Motoda , G.J. McLachlan, A. Ng, B. Liu, P.Y. Yu, Z. Zhou, M. Steinbach, D.J. Hand and D. Steinberg, Top 10 algorithms in data mining. Knowl Inf Syst 14(1) (2008)
[3] R. Ahmad, M. Darus, S.M. Shamsuddin and A.A. Bakar, Pendiskretan Set Kasar Menggunakan Ta'akulan Boolean Terhadap Pencaman Simbol Matematik, *Journal of Information Technology & Multimedia*, UKM, 2004, 1: 15-26
[4] H. Wang and L.F. Kong. Study on a new method of incremental association rule data mining. Chinese Journal of Scientific Instrument, 2009: 438~443
[5] Z. Yong, D. Qing, D.Y. Shi, S.P. Jeng, RBF neural network based on fuzzy evolution kalman filtering and application in mine safety monitoring. 2009 Ninth International Conference on Hybrid Intelligent Systems, 2009, 1: 467~470
[6] B. Marc, Khiops: A statistical discretization method of continuous attributes. Machine Learning, 55(1):53–69, 2004

[7] 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

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

[9] 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)

[10] 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

[11] 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

[12] 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

[13] 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

[14] 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

[15] 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

[16] 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

[17] S. Cost and S. Salzberg, A Weighted Nearest Neighbor Algorithm for Learning with
Symbolic Features. Machine Learning 10(1), pp. 57--78 (1993)

[18] K.A. Kaufman and R.S. Michalski, Learning from Inconsistent and Noisy Data: The AQ18 Approach. Proc. 11th Int'l Symp. Methodologies for Intelligent Systems (1999)

[19] K.J. Cios and L. Kurgan, Hybrid Inductive Machine Learning: An Overview of CLIP
Algorithms. New Learning Paradigms in Soft Computing, L.C. Jain and J. Kacprzyk, ed.,
pp. 276--322 (2001)

[20] P. Clark, T. Niblett, The CN2 Algorithm. Machine Learning, vol. 3, pp. 261--283 (1989)

[21] W.Y. Leng and S.M. Shamsuddin,"Writer Identification for Chinese Handwriting", *Int. J. Advance. Soft Comput. Appl.,* Vol. 2, No. 2, July 2010

[22] J. Catlett, On changing continuous attributes into ordered discrete attributes. In: EWSL-91: Proceedings of the European working session on learning on machine learning. Springer, New York, pp 164–178 (1991)

[23] J. Dougherty, R. Kohavi and M. Sahami, Supervised and unsupervised discretization of continuous features. In: International conference on machine learning, pp 194–202 (1995)

[24] R. Holte, Very simple classification rules perform well on most commonly used datasets. Mach Learn 11:63–91(1993)

[25] R. Kohavi, and M. Sahami, Error-based and entropy-based discretization of continuous features. In: Proceedings of the second international conference on knowledge discovery and data mining, pp 114–119 (1996)

[26] H . Liu, F. Hussain, C. Tan and M. Dash, Discretization: an enabling technique. Data Min Knowl Discov 6(4):393–423 (2002)

[27] H.S. Nguyen, S.H. Nguyen, Discretization Methods in Data Mining. In: Polkowski, L.,
Skowron, A. (Eds.), Rough Sets in Knowledge Discovery. Physica, pp. 451--482 (1998)

[28] A.K.C. Wong, D.K.Y. Chiu: Synthesizing Statistical Knowledge from Incomplete Mixed-Mode Data. IEEE Trans. Pattern Analysis and Machine Intelligence, pp. 796--805 (1987)

[29] Kerber, R.: Chimerge: Discretization of Numeric Attributes. In Proc. AAAI-92, Ninth
National Confrerence Articial Intelligence. AAAI Press/The MIT Press, pp. 123--128 (1992)

[30] Liu, H., Setiono, R.: Chi2: Feature Selection and Discretization of Numeric Attributes.
IEEE Computer Society, November 5–8, pp. 388--391 (1995)

[31] U.M. Fayyad, K.B. Irani: On the Handling of Continuous-Valued Attributes in Decision Tree Generation. Machine Learning, vol. 8, pp. 87--102 (1982)

[32] L. A. Kurgan, K. J. Cios: CAIM Discretization Algorithm. IEEE Trans. Knowledge And
Data Engineering, February vol. 16, no.2, pp. 145--153 (2004)

[31] A. Asiri, and M. S. Khorsheed. Automatic Processing of Handwritten Arabic Forms Using Neural Networks. Proceedings Of World Academy Of Science, Engineering And Technology Volume 7 August 2005 Issn 1307-6884.

[32] 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

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

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

[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] *N. Nain, B.M. Deepak, D. Kumar, M. Baswal, and B.Gautham. “Optimized Minutiae–Based Fingerprint Matching”. Proceedings of the World Congress on Engineering 2008, Volume (I) , WCE 2008, July 2 - 4, 2008, London*

[37] G.A. Abandah and M.Z. Khedher, "Analysis of Handwritten Arabic Letters Using Selected Feature Extraction Techniques", presented at Int. J. Comput. Proc. Oriental Lang., 2009, pp.49-73

[38] K. Daifallah, N. Zarka and H. Jamous, ‘‘ Recognition-based Segmentation Algorithm for On-line Arabic Handwriting ‘‘ , ICDAR, Spain, pp. 886-890, 2009

[39] H. El Abed and V. Märgner, "Comparison of Different Preprocessing and Feature Extraction Methods for Offline Recognition of Handwritten ArabicWords", ICDAR, Brazil, pp. 974-978, 2007

[40] M.P. Schambach, J. Rottland and T. Alery, ‘‘How to convert a latin handwriting recognition system to Arabic’’, ICFHR, Canada, 2008

[41] J. Sternby, J. Morwing, J. Anderson and C. Friberg, ‘‘On-line Arabic handwriting recognition with templates’’, Pattern Recognition, vol 42, pp. 3278-3286, 2009

[42] H. Miled, C. Olivier, and M. Cheriet. Modlisation de la notion de pseudo-mot en reconnaissance de mots manuscripts arabes. In CIFED, 2000

[43] 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

[44] A. K. Jain, L. Hong, S. Pankanti, and R. Bolle, “An identity authentication system using fingerprints,” *Proc. IEEE*, vol. 85, pp. 1365–1388, Sept. 1997

[45] E. C. Driscoll, C. O. Martin, K. Ruby, J. J. Russel, and J. G. Watson, “Method and apparatus for verifying identity using image correlation,” U.S. Patent 5 067 162, 1991

[46] A. Sibbald, “Method and apparatus for fingerprint characterization and recognition using auto-correlation pattern,” U.S. Patent 5 633 947, 1994

[47] D. K. Karna, S. Agarwal, and S. Nikam, "Normalized Cross-correlation based Fingerprint Matching," in Fifth International Conference on Computer Graphics, Imaging and Visualization, 2008, pp. 229 - 232

[48] A. K. Jain, L. Hong, and R. Bolle, "On-line Fingerprint Verification," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, pp. 302 - 314, April, 1997, 1997

[49] I. A. Al-Sughaiyer and I. A. Al-Kharashi, “Arabic Morphological Analysis Techniques,” *J. of the Amer. Soc. For Inf. Sc. And Tech.*, vol 55, no. 3, pp. 189-213, 2004

[50] C.-L. Liu and K. Marukawa, Global shape nor-malization for handwritten Chinese character recognition: a new method, in *Proc. 9th Int'l Workshop on Frontiers of Handwriting Recog-nition*, Tokyo, Japan, 2004, 300-305

[51] Y. Yin, J. Tian, X.K. Yang," Ridge Distance Estimation in Fingerprint Images: Algorithm and Performance Evaluation" EURASIP Journal on Applied Signal Processing, 2004
[52] S. Chikkerur and V. Govindaraju. "K-plet and Coupled BFS: A Graph Based Fingerprint Representation and Matching Algorithm". Proceedings of the International Conference on Biometrics, Hong Kong, 2006
[53]N. Ansari, M.-H. Chen, and E.S.H. Hou, "A Genetic Algorithm for Point Pattern Matching," *Dynamic, Genetic, and Chaotic Programming*, B. Soucek, ed., pp.353-371. New York: John Wiley and Sons, 1992
[54] C.H. Wu, "Advance Feature Extraction Algorithms for Automatic Fingerprint Recognition Systems", Dissertation on The University of New York, April 2007
[55] B. Li, Z. Sun and T. Tan, Hierarchical Shape Primitive Features for Online Text-independent Writer Identification. *Proc. of 2th ICB*, 201–210 (2007)
[56] G. Zhu, X. Yu, Y. Li, and D. Doermann, Language identification for handwritten document images using a shape codebook. *Pattern Recogn.*, 42(12), 3184-319 (2009)
[57] A. Namboodiri and S. Gupta, Repudiation Detection in Handwritten Documents . *Proc of The 2nd International Conference on Biometics (ICB'07).* 27-29 August. Seoul, Korea, 356-365 (2007).
[58] H. Judith, K. Bowers, M. Cannon and P. Kelly, Script and language identification for handwritten document images. *Int. J. Document Analysis and Recognition*, 2 (2-3), 45–52 (1999)
[59] C.L. Liu, T.H. Su, T. Zhang and H.J. Huang, Segmentation-free recognizer based on enhanced four plane feature for realistic Chinese handwriting. ICPR 2008: 1-4 (2008)
[60] X. Bing and Z. Mei, A Classification Method for Building Detection Based on LiDAR Point Clouds. Urban Remote Sensing Joint Event, Shanghai. (2009)
[61] X. Gao, L.W. Jin, J.X. Yin, J.C. Huang, "A New Stroke-Based Directional Feature Extraction Approach for Handwritten Chinese Character Recognition," icdar, pp.0635, Sixth International Conference on Document Analysis and Recognition (ICDAR'01), 2001
[62] C.L. Liu, Normalization-Cooperated Gradient Feature Extraction for Handwritten Character Recognition. IEEE Trans. Pattern Anal. Mach. Intell. 29(8): 1465-1469 (2007)
[63] B. Krisztian and S.T. Lars, Motif-based Classification of Time Series with Bayesian Networks and SVMs, in Proceedings of 32nd Annual Conference of the Gesellschaft für Klassifikation (GfKl 2008), Hamburg, Germany.
[64] 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.
[65] J. Komorowski, A. Øhrn, and A. Skowron, The ROSETTA Rough Set Software System, In Handbook of Data Mining and Knowledge Discovery, W.

Klösgen and J. Zytkow (eds.), ch. D.2.3, Oxford University Press. ISBN 0-19-511831-6 (2002)
[66] Z. Pawlak, Rough sets. *International Journal of Computer and Information Sciences*, 341-356 (1982)
[67] R. Silipo, Neural Networks. In D. J. Hand (Ed.), *Intelligent Data Analysis: An Introduction* (Second ed.): Springer (2003)