International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:6, No:2, 2012

Fingerprint Identification using Discretization Technique

W. Y. Leng and S. M. Shamsuddin

***Abstract*—**Fingerprint based identification system; one of a well known biometric system 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, we proposed an identification framework of individuals by means of fingerprint. Different from the most conventional fingerprint identification frameworks the extracted Geometrical element features (GEFs) will go through a Discretization process. The intention of Discretization in this study is to attain individual unique features that could reflect the individual varianceness in order to discriminate one person from another. Previously, Discretization has been shown a particularly efficient identification on English handwriting with accuracy of 99.9% and on discrimination of twins' handwriting with accuracy of 98%. Due to its high discriminative power, this method is adopted into this framework as an independent based method to seek for the accuracy of fingerprint identification. Finally the experimental result shows that the accuracy rate of identification of the proposed system using Discretization is 100% for FVC2000, 93% for FVC2002 and 89.7% for FVC2004 which is much better than the conventional or the existing fingerprint identification system (72% for FVC2000, 26% for FVC2002 and 32.8% for FVC2004). The result indicates that Discretization approach manages to boost up the classification effectively, and therefore prove to be suitable for other biometric features besides handwriting and fingerprint.

***Keywords*—**Discretization, fingerprint identification, geometrical features, pattern recognition

I. Introduction

A reliable personal authentication system with Biometric technology [1] that uses the physical or behavioral characteristics to identify actual persons or users has become a popular research during the past 30 years. Among various Biometric applications (e.g., face, handwriting, signature, voice, iris, retina, thermogram, hand vein, palm, gait, ear, keystroke dynamics, etc. [2-3], fingerprint identification is one of the most mature, well known and have highest record employ in government law enforcement applications. The popularity and importance of this research has successfully produces a great number of publications. The study shows that fingerprint techniques can be classified into two common groups:

The work of W. Y. Leng and S.M. Shamsuddin was supported by Ministry of Higher Education (MOHE) under MyBrain15 programme.

W. Y. Leng is with the Research Group of Soft Computing, University of Teknology Malaysia, 81300 Skudai, Johor, Malaysia (012-9050323; e-mail: yeeleng28@gmail.com).

S. M. Shamsuddin, is with Universiti Teknologi Malaysia, Johor, 81310 Malaysia. She is now the Head of Soft Computing Research Group (SCRG), Universiti Teknologi Malaysia, Johor 81310 Malaysia (e-mail: mariyam@utm.my).

Minutiae based and Correlation or Ridge based. The minutiae points are the anomalies features found within the finger patterns. Minutiae based approach first look for minutiae points then gather two samples of minutiae points and finally decides the number of matched minutiae.

Correlation based approach, in contrast, compare the overall ridges within the pattern to see if the ridges exist in the two fingerprints align. An approach use of minutiae based approach with Hough transform method has been proposed by Ratha *et al.* [1]. This aproach is based on the multilevel indexing which integrates a number of high level pattern class and ridge density features. This approach able to diminish the search space.

Nain *et al.* [2] proposed a novel minutiae based approach using mass centroid concept. In this method, the alignment within minutiae found enormously fast. The algorithm proposed by Umair *et al.* [3] is based on wavelet transformation fused with minutiae features. Meanwhile, the minutiae approach proposed by Abbad is based on centered round regions and overall position for fingerprint verification [4]. It finds the best position between both minutiaes using Hough Matching technique. Another approach proposed in paper of Roli [5] focuses on mathematical morphology. The author extracts the minutiae in thinned fingerprint image using binary Hit or Miss Transform (HMT). This approach provides consistent fingerprint ridge formation thus successfully reduces a lot of effort in the post-processing phase.

In this paper, the contents of this study are organized as follow. In the following section, the review of the fingerprint structures is introduced. Section 3 explains the related work regarding to the previous and issues of fingerprint identification system. Because of the values obtained through previous findings provide low inter-variability between the feature values and class labels, high intra-variability within the samples of fingerprint features of a person, identification based fingerprint still remain a challenging problem especially the issues of loss originality information of an individual. Thus, a concept of Discretization is introduced and briefly described in section 4. Meanwhile, section 5 reports the comparison result between the conventional system and proposed system. Section 6 summarizes the paper.

II. Fingerprint Structures

A fingerprint pattern is formed of ridges and valleys over the surface of finger (Fig. 1). 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 used to extract the common singular points namely loop, delta, whorl and etc. Meanwhile, Fig. 2 shows the examples of singular points such as core point, delta, loop and whorlon a fingerprint image. The core points or small circles shown in Fig. 2 is the center point of the highest loop in the singular region and generally used as the 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. 3.

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 global and local information including observes the regions of the ridge termination or ridge anomalies are known as minutiae points. Some of typical ridge discontinuous structures are shown in Fig. 4. 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. Figure 4 shows a closer observation of real fingerprint image.

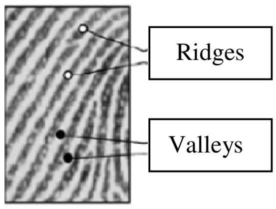


Fig. 1 Example of Ridges and Valleys on a fingerprint image

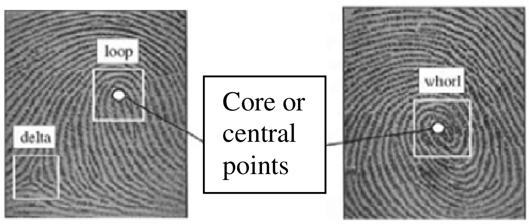


Fig. 2 Example of Singular regions and core points on a fingerprint image

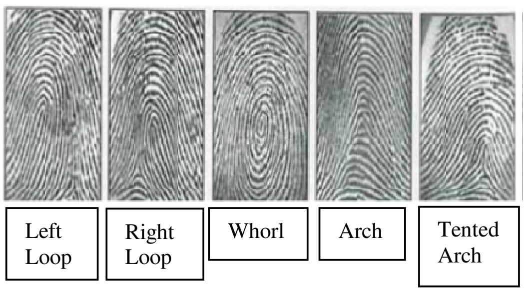
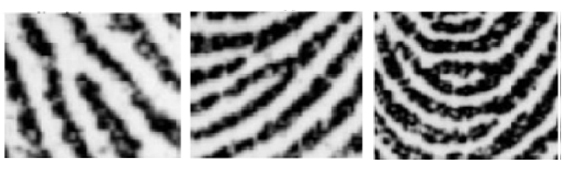


Fig. 3 Classes of fingerprint



(a) (b) (c)

Fig. 4 Type of Minutiaes (a) ridge termination or ridge ending (b) ridge bifurcation and (c) a lake

We can see a ridge ending in Fig 4(a), a ridge bifurcation in Fig 4(b). Fig. 4(c) shows a lake.

III. Related Work

It is not easy to develop an accurate fingerprint identification system. Several authors have suggested different systems that suit for the identification of fingerprints. Figure 5 shows a common and traditional fingerprint identification system. The system contains three main components; Pre-processing, Feature extraction and Classification.

After image pre-processing module (Fig. 5), input image is noiseless and binarized. Segmentation [6], filtering [7] and normalization [8] are the common image processing techniques used to remove spurious noise of fingerprint images. Noiseless images are then fed into feature extraction module for minuitiae extraction.

In feature extraction module, a fingerprint's pattern of ridges is the important properties to represent the relationship and characteristic of an individual. From an extensive research of available literature, fingerprint approaches are classified into two general classes: minutiae based approach and correlation based approach. Both approaches have own of its advantageous and some of its shortcomings. In minutiae-based, it looks for some region of the appropriate minutiae types that form a particular pattern of a query finger.

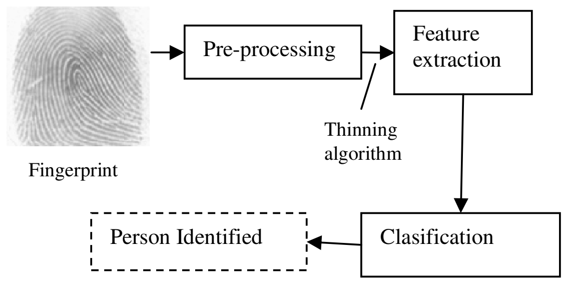


Fig. 5 Conventional Fingerprint Identification System

The common types are ridge ending and ridge bifurcation minutiae. In fingerprint identification system [9], chain code contour is used for minutiae extraction. Another system focussed on core point extraction [10]. Angle values are calculated and indexed technique is applied for fingerprint matching. The matching is compared with those stored in database. However, system that based on core point extraction has some disadvantages that when the different orientation of finger from same person is input to the system, the angle values change. These changes direct to the false matching or rejection.

Meanwhile in fingerprint systems based on correlation approach [11]-[13], it considers the overall patterns of ridges and valleys found in a finger image. Correlation based approach also provide some disadvantages to the system where it only present accurate identification for good quality images. Low quality image often comes with spurious features caused by fingerprint imperfections such as ridge gaps due to skinfolds, contiguous ridges due to finger pressure. Since this type of extraction encounters all region of the finger pattern, insufficient information including distort and false features are also consider as important part for matching process thus decreasing the accuracy of identification. According to Jain *et al.* [7], the best total number of minutiae points extraction for a fingerprint is around 60 to 80 minutiae points.

Recently, a great number of solutions have been proposed to overcome the issues in fingerprint system. Karna *et al.* [14] presented a normalized cross correlation technique to decrease the computational effort and error rate for fingerprint matching. Meanwhile, Jain *et al.* [15] proposed a method that uses ridges joint with minutia.

The final module (refer to Fig. 5) deals with the classification of extracted features into appropriate class. However this module often fed with large continuous features, close similarity features between fingerprint samples and inappropriate feature spaces which effect from the issues of feature distributions and representation after extraction. The following section gives the description of the steps in the proposed framework for the identification of fingerprint.

IV. Architecture of Propose Framework

In order to produce an efficient identification system, a question is always on *how to acquire features that can reflect the identity of an individual*? Three interests of the proposed framework are stated below;

1. Uniqueness: The feature amongst a wide set of individuals; sets must have a very high individual predictive power;
2. Inter-individual variability: The feature sets within Fingerprint images writing of an individual must e extent amongst a wide set of individuals; and
3. Intra-individual variability: The feature sets within the samples of handwriting must differ as little as possible for each individual.

The common issues of an identification system are often regarding to the complexity of the extracted features, its distribution and how to represent those features systematically in order to differentiate a person among other based on fingerprint samples.

Therefore in this study, a method called Discretization which is a promising independent technique is adopted into this identification framework as presented in Fig. 6 to handle the issues as mentioned above. Discretization here is based on the work done by Azah *et al.* [16]. Azah's previous work has successfully achieved identification accuracy rate of 99.9% on English handwriting and 98% on discrimination of twins handwriting [17]. Beside this, the research of Discretization have been researched from the beginning of the fifties [18] nineties [19], twenties [20] and recently numerous scientific papers and inventions in Discretization have been widely expanded in the field of pattern recognition [16], [21], [22], [23] and data mining [24], [25], [26]. Due to its efficiency to cope with the issues of data complication and misclassification in the mentioned area, as such in this paper, Discretization is further explored into fingerprint identification to seek for its accuracy which has not been reported and done yet by other study.

In Fig. 6, the geometry based approach is used for our fingerprint extraction. This scheme is transformation independent and is based on local and global based technique. Geometrical element features (GEFs) are the output of the Pre-Discretization module and are extracted based on the singular and minutiae points in a fingerprint contour. GEFs have been proved as an effective expression of fingerprint images and have been successfully used in recognition [27] and verification [28] of fingerprint system. After extraction, the GEFs are then fed into Discretization module. In this study, the intention of Discretization is to represent the extracted features into systematic way in order to make classification task easier thereby increasing the identification accuracy. According to Azah *et al.* [16], Discretization is a process of discovering the unique features from contiguous set of an individual samples, then signifying them into a single standard value representation. This representation defines a clear

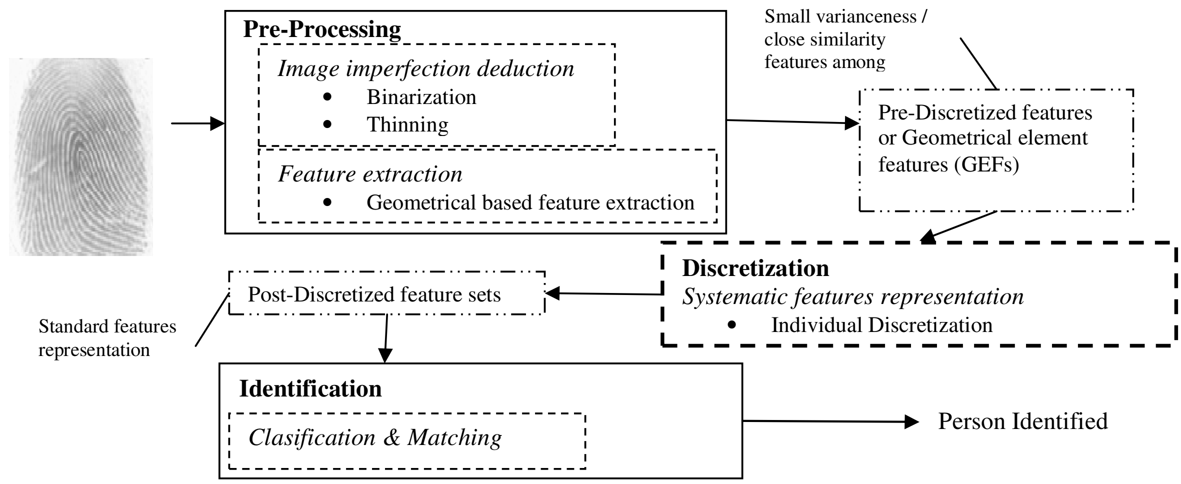


Fig. 6 Proposed Fingerprint Identification Framework

varianceness of similarity features between individuals which offer maximum discriminative power that is capable to discriminate one person from another. Meanwhile, Discretization defined by Lee and Shin [29] is a process that divides continuous numeric values into discrete categorical interval values. This practice is essential because machine learning only understands and execute entirely with discrete features. More details on each module are well described in the following section.

A. Pre-Processing

The basic objectives of pre-processing in this section are to overcome the distortion of input fingerprint image and to convert the input image into levels of pixels representation that could be interpreted by computer for further analysis. First, binarization is performed on the grayscale fingerprint image to convert it into a black and white image with an appropriate threshold value. Threshold value is the mean value of block gray value to provide an image with minimum noise. Meanwhile, to reduce the thickness of all ridge lines, a thinning process is carried out with parallel thinning algorithm. Detail explanation on pre-processing techniques can be found in next sub-section.

1. Binarization

Binarization is a process to convert the gray scale image pixel value into 0 or 255. The algorithm is presented below;

For gray value of each pixel gp in I
  If gp > M,
    set gp = 1
  Else,
    set gp = 0

Where;
I-Fingerprint image;
M- Mean value of block gray value;

Mean value, M of each block gray image value is used as an adaptive threshold to be compared with each gray pixel value, gp. If the gray pixel value, gp of fingerprint image, I is more than M value, set gp value to 1 (white) else set gp value to 0 (black).

2. Thinning Algorithm

Commonly, variation thickness of fingerprint ridges is due to the digitization process during data acquisition. This is also known as digitization noise. It is necessarily to reduce the ridges thickness into single-pixel width strokes. The thinning mathematical morphological operation used for our fingerprint image is a built-in morphological thinning function provided in Matlab. The process will end when the fingerprint image is reduced to a single pixel wide otherwise the process is repeated until all pixels are one pixel wide (obtaining the line diagram). The processes of Pre-processing from original image to the thinned image are shown in Fig. 7 below.

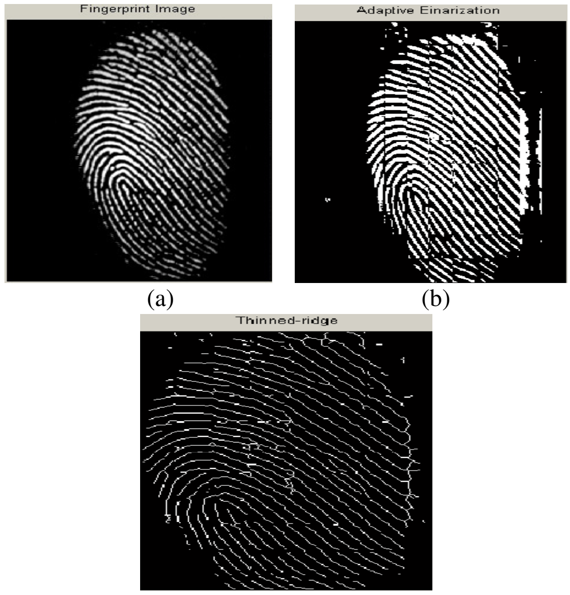


Fig. 7 Binarization of fingerprint image (a) Original image (b) Binarized image (c) Thinned fingerprint image

3. Geometrical based feature extraction approach on fingerprint contour

The basic idea of this approach is to extract the useful features based on geometry technique on the general shape of ridge flow and minutiae points including ridge terminations and ridge bifurcations after thinning process. The advantage of this technique provides a coarse level description of the whole fingerprint pattern includes the ridges flow of fingerprint. From ridges flow of fingerprint, minutiae points can be easily founded and extracted as shown in Fig. 8. Two mathematical morphological operation of *clean* and *spur* are applied in this algorithm. These algorithms remove noises such as spurs, spikes of the fingerprint image.

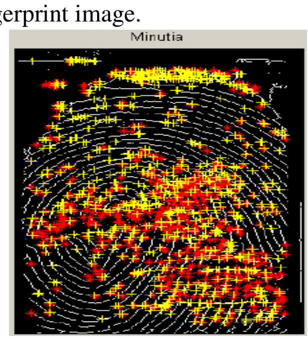


Fig. 8 Selected features for extraction*Shape of ridge flow and minutiae point detection*

Ridges termination and ridges bifurcation are extracted based on 9-pixel neighbourhood strategy. The determination of ridges in each image is defined as half the entirety of the differences between pairs of the next pixels in a 3x3 window. Generally, if the center point is 1 with 3 one points as neighbour in a 3x3 window, then it detected as a ridge bifurcation. If the center points is 1 and with only 1 one point as neighbour, then it is called as ridges termination. The mathematical equation (1) for the ridge type determination is as follows;

$$\text{Number of Pixels} = \frac{1}{2} \sum_{i=0 \ldots n}^{n=8} \left| P_i - P_i + 1 \right| \qquad (1)$$

Where;
If the number of pixels is 1, then it is termination minutiae;

Else if the number of pixels is 2, then is a bifurcation;

Else for more complex minutiae.

The extracted feature datasets are obtained from 3 different fingers with 4 impressions per finger, giving the total of 12 feature vectors for each FVC (Fingerprint Verification Competition) database. These feature datasets are known as pre-discretized datasets which are the original extracted features before execute with Discretization algorithm.

However after extraction process, it can be observed that some of the extracted sets of features are inconsistent (refer to Fig. 12) which consists of very close similarity features within the samples of fingerprint of another person (the occurrence issues of inter and intra-variability features). The inconsistent sets of features is caused by the ridge structure of fingerprint which provides similar ridge shapes and flows between individuals and variation in ridge shapes and flows through different orientation of fingerprint from the same individual. These weaknesses which often happen in traditional framework (refer to Fig. 5) causing *small variances* amongst fingerprint samples of various persons thereby minimizing the important information of individual. Addition to this, most of the previous studies illustrated the identification performance through the process of extraction techniques of individual features on fingerprint samples and not on signifying the individual characteristics of ridge pattern between individuals for systematic representation. Thus in this study, we proposed a new framework for fingerprint identification. Unlike to other identification system, the significance of this study is the introduction of Discretization which is added before classification module to represent the extracted features systematically in order to improve individual identification based on fingerprint samples efficiently (refer to Fig. 6).

B. Discretization

From a wide-ranging research of available literature, it was found that most of the enhanced identification system is based on feature extraction and classification phase. The distances among the central pixel of the ridge with the termination and bifurcation minutiae points are commonly adopted in those works [6], [30]. These techniques very depend on the position of the fingerprint. If the fingerprint position modify, the central point position, and the distance between central point with ridge point also will be adjusted. Overall, the conventional framework did not perform the system's objectives as expected; hence result in poor fingerprint identification. In this study, GEFs are performed with Discretization algorithm. It gives class information to each image to represent the identity of the individual. This algorithm has proven to be particularly efficient for individual identification as successfully performed in [16] and [17] for both of its low complexity and its ability to deal with large and high dimensional feature vectors. The sequences of steps involved in the algorithm are as follow:

Step1. Determine interval

Discretization first form a subset of features for each individual. The highest and lowest feature values (max and min) in each subset are picked to estimate the interval for each bin. Number of bins or cuts for each individual is created based on the number of features.

```
For an individual
        Min = min feature; Max = max feature;
        No_bin = no_geometric_feature;
        Interval = (Max - Min) / No_bin;
```

In this study, nine features are extracted from each fingerprint sample. Thus, there will be nine bins created for each individual class. Unlike to other schema, Discretization here keeps the original number of GEFs that has been extracted. The purpose of the algorithm is not to change the originality of the information but just to demonstrate the extracted features into a systematic features representation that could defines a clear *varianceness* of similarity features between individuals in order to determine the efficiency of Fingerprint Identification.

Step2. Define representation value for each cut

Then it divides the range of max and min feature value with the number of bins or cuts to gives lower and upper approximation to each bin.

```
For each bin
        Find lower and upper value of interval;
        RepValue = (upper – lower) / 2;
```

Step3. Classify features into interval

Features that falls within the approximation will be replaced with a single standard value. This process of allocating features into appropriate interval began with the first feature and will be repeated until the end of the features.

```
For (1 to no_geometric_feature)
        For each bin
        If (feature in range of interval)
           Dis_Feature = RepValue;
```

The process and result after Discretization process are illustrated in Fig. 10 and Fig. 11 below. Discretization algorithm implemented in the matching phase involves the overall features gained during extraction phase of the fingerprint. All the other processes are done internally and the outcome of this phase is the discretized values (Fig. 11) that are not affected by the distance, positional and the angle difference. Fig. 9 presents the samples of original GEFs (before execute with Discretization algorithm). This dataset is also known as pre-discretized dataset. There are nine extracted vectors in each row and the last column represents the individual's class.

6.1000 8.0000 9.5000 14.0000 18.8000 17.0000 4.5500 7.5000 10.4500 **1**
3.9500 7.0000 6.0000 9.1500 12.7500 16.9500 1.8500 5.8500 10.6000 **1**
1.6500 2.5500 2.2500 6.3500 9.0000 5.8000 3.5000 4.5500 9.1000 **1**
3.5500 2.5500 1.2500 8.2000 11.5500 6.7500 4.3000 5.6000 7.3500 **1**

High intra variability features

Low inter/close similarity features

1.9500 7.2000 5.6500 8.0000 14.4000 8.6500 3.4500 7.0000 3.5500 **2**
4.9500 7.6000 5.4500 12.1000 17.6000 8.6500 3.8000 7.1000 4.3500 **2**
5.8000 6.7500 6.2000 16.2000 22.3000 15.4500 6.5000 9.7000 3.5000 **2**
4.9500 8.8500 5.5500 13.0000 17.8500 11.5500 4.3500 6.2500 6.7000 **2**

Fig. 9 Examples of Pre-discretized FVC2004 fingerprint datasets of 2 individual that has low inter-variability and close similarity features

Meanwhile, Fig. 11 shows the discretized feature datasets after process in Fig. 10. Features that are highlighted with different colors indicate the varianceness features obtained between that particular person from another. This varianceness of individual is clearly represented with standard values where in this study we define these values as domain values (DV) of individual (refer to Fig. 11). DV represent the uniqueness information of individuality where each person has it own finger's structures (ridges and valleys) even with different impressions or orientations per finger. This schema preserves the originality information of individual even with large number of candidates because Discretization here executes based on each individual and does not change the originality feature characteristics in a fingerprint of an individual.

LOW and UPPER BIN for Individual : 2
MIN Value 1.95 MAX Value 22.3
Bin 0: Low :1.95 Upper :4.21111 Rep Value for Bin 0: 1.13056
Bin 1: Low :4.21111 Upper :6.47222 Rep Value for Bin 1: 5.34167
Bin 2: Low :6.47222 Upper :8.73333 Rep Value for Bin 2: 7.60278
Bin 3: Low :8.73333 Upper :10.9944 Rep Value for Bin 3: 9.86389
Bin 4: Low :10.9944 Upper :13.2556 Rep Value for Bin 4: 12.125
Bin 5: Low :13.2556 Upper :15.5167 Rep Value for Bin 5: 14.3861
Bin 6: Low :15.5167 Upper :17.7778 Rep Value for Bin 6: 16.6472
Bin 7: Low :17.7778 Upper :20.0389 Rep Value for Bin 7: 18.9083
Bin 8: Low :20.0389 Upper :22.3 Rep Value for Bin 8: 21.1694
DISCRETIZE DATA
1.13056 7.60278 5.34167 7.60278 14.386 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

Fig. 10 Discretization process for FVC2004 fingerprint datasets of second individual

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**

DV for individual 1: 4.175

1.13056 7.60278 5.34167 7.60278 14.3861 7.602781.13056 7.60278 1.13056 **2**
5.34167 7.60278 5.34167 12.125 16.64727.602781.13056 7.60278 5.34167 **2**
5.34167 7.60278 5.34167 16.64721.169414.3861 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**

DV for individual 2: 5.34167

12.5611 16.005616.0056 22.894412.5611 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**

DV for individual 3: 9.11667

Fig. 11 Post-discretized fingerprint datasets of 3 individual from FVC2004

The final discretized feature data (Fig. 11) are then fed into the classification module using Rosetta Roughset Toolkit [31]. Four reduction algorithms are chosen namely Johnson's algorithm, Holte 1R, Genetic algorithm and Exhaustive algorithm. The results for individual identification performance using pre-discretized and post-discretized datasets with random number of training and testing data are reported in Table 1, Table 2 and Table 3. Meanwhile, Table 4 shows the comparison results between existing system (conventional system) and the proposed system to seek for the best individual identification accuracy based on fingerprint.

V. Experimental Result and Discussion

The proposed fingerprint identification system was executed on recognized Fingerprint database that are FVC2000, FVC 2002 and FVC 2004. Three databases were created for each FVC thus giving the total of 880 fingerprints from 110 different fingers with 8 impressions per finger. The fingerprint images are in TIFF format.

In this section, the experimental results of the study are presented. These datasets are classified using Roughset Toolkit (ROSETTA). The Reducer Algorithm is used to find reducts approximation in a decision table. It returns a set of reduct which contains a set of rules attached to it as a child. The minimal subsets will be taken to distinguish data from other datasets. Instead of fixing a single element *x,* it select subset *X* of *U*, and execute each $x \in X$ consecutively. It will first compute the minimal subsets that determine the first element in *X* from datasets in *U*, later discriminating the second element in *X* from datasets in *U;* the process will go on based on the set of rules until the fairly accurate reduct is produced.

As for identification performance, confusion matrix technique is adopted. Identification performance describes the accuracy percentage of identification cases where features make a correct fall into the right interval of an instance's class. It gives the overall accuracy measurement and sensitivity of each actual classified and misclassified object. Each image n (n = 1, . . . , m) of the FVC databases subset is matched against the other m-1 images of the same subset. We define an object was correctly classified into the instance class, when an image were predicted to the instance class from the same individual's finger. Moreover, if an image predicted to an instance class from a different individual's finger, we say that the image is misclassified into the instance class**.** The result for individual identification performance using pre-discretized and post-discretized datasets with random number of training and testing data is reported in Table I, II and III. The identification accuracy is computed using formula in (2).

Identification_rate=Predicted features in an instance class / total features of all instances classes (2)

From the above-obtained result, it clearly shows that four reduction methods namely Johnson's algorithm, Holte 1R, Genetic algorithm and Exhaustive algorithm perform well with post-discretized datasets. Each of the method in Table 1, 2 and 3 successfully achieved the overall average accuracy of more than 90.0%. On the other hand, same four reduction methods using pre-discretized datasets as presented in Table 1, 2 and 3 reports a worst performance, which provide low identification rates; below 75.0% on each method. As expected, fingerprint datasets executed with Discretization

TABLE I
FINGERPRINT IDENTIFICATION ACCURACY FROM FVC2000

ROSETTA Built-in Methods on Reductions	Data Types	70% Train Data 30% Test Data	60% Train Data 40% Test Data	50% Train Data 50% Test Data	Average Acc
Johnson's Algorithm	Pre_Dis	66.67	75.00	60.00%	67.22
	Post_Dis	100.0	100.0	100.00	100.0
Holte 1R Algorithm	Pre_Dis	66.67	75.00	80.00	73.89
	Post_Dis	100.0	100.0	100.00	100.0
Genetic Algorithm	Pre_Dis	66.67	75.00	80.00%	73.89
	Post_Dis	100.0	100.0	100.00	100.0
Exhaustive Algorithm	Pre_Dis	66.67	75.00	80.00%	73.89
	Post_Dis	100.0	100.0	100.00	100.0

TABLE II
FINGERPRINT IDENTIFICATION ACCURACY FROM FVC2002

ROSETTA Built-in Methods on Reductions	Data Types	70% Train Data 30% Test Data	60% Train Data 40% Test Data	50% Train Data 50% Test Data	Average Acc
Johnson's Algorithm	Pre_Dis	33.33	25.00	20.00	26.11
	Post_Dis	66.67	50.00	100.00	72.22
Holte 1R Algorithm	Pre_Dis	33.33	25.00	20.00	26.11
	Post_Dis	100.0	100.0	100.0	100.0
Genetic Algorithm	Pre_Dis	33.33	25.00	20.00	26.11
	Post_Dis	100.0	100.0	100.00	100.0
Exhaustive Algorithm	Pre_Dis	33.33	25.00	20.00	26.11
	Post_Dis	100.0	100.0	100.00	100.0

TABLE III
FINGERPRINT IDENTIFICATION ACCURACY FROM FVC2004

ROSETTA Built-in Methods on Reductions	Data Types	70% Train Data 30% Test Data	60% Train Data 40% Test Data	50% Train Data 50% Test Data	Average Acc (%)
Johnson's Algorithm	Pre_Dis	33.33	25.00	40.00	32.78
	Post_Dis	66.67	50.00	60.00	58.89
Holte 1R Algorithm	Pre_Dis	33.33	25.00	40.00	32.78
	Post_Dis	100.0	100.0	100.00	100.0
Genetic Algorithm	Pre_Dis	33.33	25.00	40.00	32.78
	Post_Dis	100.0	100.0	100.00	100.0
Exhaustive Algorithm	Pre_Dis	33.33	25.00	40.00	32.78
	Post_Dis	100.0	100.0	100.00	100.0

bear higher discriminative power of individuality compared to the actual datasets accordingly. Hence, it can be well concluded that Discretization method could enhance the performance of individual identification effectively on fingerprint besides on handwriting.

The comparison with the proposed framework is presented in Table 4. The proposed system successfully achieves more than 85.0% on an average when compared with every FVC databases. This illustrates that our proposed system is much competent than conventional system; can manage and represent real minutiae features in much better way during pre-processing phase and could provide higher discriminative power in post-processing phase to assist classification for individual identification.

For clearer picture, the comparison results reported in Table 4 are visualized as depicted in Fig. 12. To quantify the performance of the proposed system using Discretization in brief, a *t-test* to illustrate the significance difference between pre and post fingerprint datasets on individual identification performance, from three FVC databases, FVC2000, FVC2002 and FVC2004 is also presented in this section.

TABLE IV
COMPARISSON OF IDENTIFICATION ACCURACY ON EACH FVC DATABASES

FVC Databases	**Conventional or Existing Systems (%) (without using Discretization)**	**Proposed System (%) (using Discretization)**
FVC 2000	72.22	100.00
FVC 2002	26.11	93.06
FVC 2004	32.78	89.72

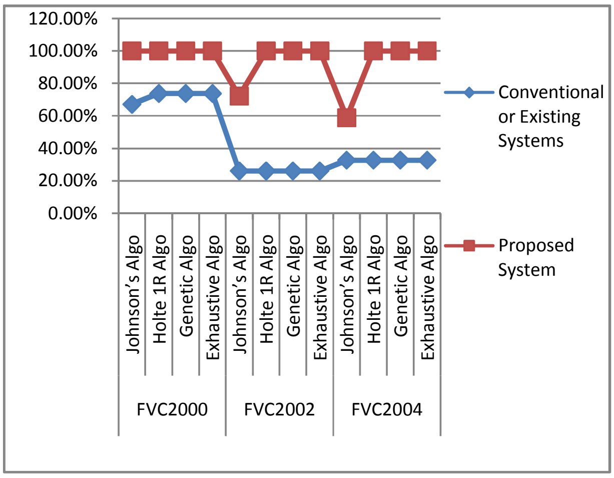


Fig. 12 Visualization of Post-discretized and Pre-discretized fingerprint datasets on each FVC databases

For hypothesis testing, the scalar mean of an average accuracy for pre and post discretized vectors after several training and testing procedure is used as a statistic and basis for one-to-one individual identification performance comparisons of the fingerprints. Given two paired sets; pre-discretize datasets, x_i and post discretized fingerprint datasets, y_i of n total fingerprint for each FVC database, the paired *t-test* determines if they differ from one another in a significant way. As per Table 5, when t test analysis was done on post-discretized fingerprint datasets and those of actual fingerprint datasets (pre-discretized datasets), the results were highly significant with t 8.19455E-09 and $P<0.05$ for FVC 2000. When t test analysis was done on FVC 2002 database, the results were also highly significant with t 3.71138E-08 and $P<0.05$. Same analysis achievement on pre and post-discretized FVC 2004 fingerprint datasets which shown a highly significant with t 3.49287E-07 and $P<0.05$, stating that individual identification shows a better performance with discretized fingerprint datasets as compared to actual extracted datasets. Thus, the analysis of the data reveals that there is distinct improvement in the individual identification performance after the process of Discretization on fingerprint datasets.

VI. Conclusion

Overall, the proposed framework for fingerprint identification system as illustrated in Fig. 6 which composes of three main components for collecting the important information, extracting the required features from the fingerprint image and classifying the features into actual individual class accordingly. The estimation and evaluation efficiency of Discretization on the proposed fingerprint framework can be seen through the comparison with the conventional fingerprint framework on identification performance results. The results shows that the proposed fingerprint identification system extensively boosted up the individual identification accuracy based on fingerprint and effectively prevail over the existing system that had positional and oriental weaknesses. Proposed system successfully achieves a maximum accuracy of 99.0%. As it can be seen from previous section in Fig. 10 and Table 4, the performance of Discretization is very good in terms of improving individual identification due to its capability to represent each individual features systematically and uniquely even though with different impressions or orientations per finger. This representation provides a high discriminative power to discriminate one person from another although we use fingerprint datasets as our comparing index instead of handwriting datasets.

TABLE V
Hypothesis Result of an Average Accuracy for Pre and Post Discretized Feature Datasets

FVC Databases	Mean	Std. Dev.	t	P	Significance
FVC2000 Fingerprint Pre Fingerprint Post	72.22 100.0	6.64 0.00	8.19455E-09	0.05	Highly Significance
FVC2002 Fingerprint Pre Fingerprint Post	26.11 93.06	5.74 16.60	3.71138E-08	0.05	Highly Significance
FVC2004 Fingerprint Pre Fingerprint Post	32.78 89.72	6.41 18.93	3.49287E-07	0.05	Highly Significance

Acknowledgment

Authors would especially like to thank Universiti Teknologi Malaysia, Skudai Johor Bahru MALAYSIA for the support and *the Soft Computing Research Group* (SCRG) for their excellent cooperation and contributions to improve this paper.

References

[1] N.K. Ratha, K. Karu, S. Chen, A.K. Jain. A Real-Time Matching System for Large Fingerprint Databases. IEEE Trans. Pattern Anal. Mach. Intell., 1996: 799-813.
[2] N. Nain, B.M. Deepak, D. Kumar, M. Baswal, and B. Gautham, Optimized Minutiae–Based Fingerprint Matching. Lecture Notes in Engineering and Computer Science, vol. 2170(1), pp. 682-687,2008.
[3] M.K. Umair, A.K. Shoab, N. Ejaz, and R. Riaz, "A Fingerprint Verification System using Minutiae and. Wavelet Based Features," International Conference on Emerging. Technologies, pp. 291-296, 2009.
[4] K. Abbad, N. Assem, H. Tairi and A. Aarab, "Fingerprint Matching Relying on Minutiae Hough Clusters", ICGST-International Journal on Graphics, Vision and Image Processing, vol. 10(1), 2010.
[5] B. Roli, S. Priti and B. Punam, "Effective Morphological Extraction of True Fingerprint Minutiae based on the Hit or Miss Transform" International Journal of Biometric and Bioinformatics, vol. 4(2), pp. 71-85, 2010.
[6] Y. Yin, J.Tian and X.K Yang," Ridge Distance Estimation in Fingerprint Images: Algorithm and Performance Evaluation" EURASIP Journal on Applied Signal Processing, 2004.
[7] A K Jain and F Farrokhnia. "Unsupervised texture segmentation using Gabor filters". Pattern Recognition. vol. 12, 1991, pp.238-241.
[8] S Greenberg, M Aladjem, D Kogan and I. Dimitrov. "Fingerprint image enhancement using filtering techniques". Proceeding 15th Internat. Conference on Pattern Recognition III. Barcelona, Spain. 2000, pp.326-329.
[9] B. G. Kim and D.J. Park, ''Adaptive image normalization based on block processing for enhancement of fingerprint image", Electronics Letters, IEE, Volume 38, Isuue:14, p.p 696-698.
[10] A. Mishra and M. Shandilya, "Fingerprint's Core Point Detection Using Gradient Field Mask" International Journal of Computsser Applications (0975 – 8887*),* Volume 2 – No.8, June 2010.
[11] 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.
[12] 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.
[13] A. Sibbald, "Method and apparatus for fingerprint characterization and recognition using auto-correlation pattern," U.S. Patent 5 633 947, 1994.
[14] 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.
[15] 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.
[16] 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.
[17] B. O. Mohammed and S. M. Shamsuddin, Feature Discretization for Individuality Representation in Twins Handwritten Identification, *Journal of Computer Science* 7(7) (2011), pp. 1080–1087.
[18] W. D. Fisher, On grouping for maximum homogeneity, *Journal of the American Statistical Association* 53(284) (1958), pp. 789–798.
[19] J. Dougherty, R. Kohavi and M. Sahami, Supervised and unsupervised discretization of continuous features, in A. Prieditis & S. Russell (Eds.), in *Int. Conf. on Machine Learning* (San Francisco, 1995), pp. 194-202.
[20] M. C. Ludl and G. Widmer, Relative Unsupervised Discretization for Association Rule Mining, in *European Conference on Principles of Data Mining and Knowledge Discovery* (European, 2000), pp. 148-158.

[21] R. Ahmad, M. Darus, S. M. Shamsuddin and A. A. Bakar, Pendiskretan Set Kasar Menggunakan Ta'akulan Boolean Terhadap Pencaman Simbol Matematik, *Information Technology & Multimedia* (2004) 15-26.
[22] A. Kumar and D. Zhang, Hand geometry recognition using entropy-based discretization, *IEEE Transaction on Information Forensics and Security* 2 (2) (2007), pp. 181–187.
[23] J. Zou and C. C. Liu, Discretized Gabor Statistical Models for Face Recognition, *Digital Content Technology and its Applications (JDCTA)* 5(5) (2011), pp. 175-181.
[24] K. Kianmehr, M. Alshalalfa and R. Alhajj, Fuzzy clustering-based discretization for gene expression classification, *Knowledge and Information Systems* (Published online, 2009).
[25] K. Sarojini, K. Thangavel and D. DXevakumari, Feature Subset Selection based on Modified Fuzzy Relative Information Measure for classifier, *Engineering Science and Technology* 2(5) (2010) 2456-2465.
[26] J. Gama and C. Pinto, Discretization from Data Streams: applications to Histograms and Data Mining, in *ACM Symposium on Applied Computing*, (ACM Press, New York, 2006), pp. 662–667.
[27] M.M. Min and Y. Thein, "Intelligent Fingerprint Recognition System by Using Geometry Approach" IEEE International Conference on Current Trends in Information Technology,pp.1-5,2009
[28] M. Poulos, E. Magkos, V. Chrissikopoulos, N. Alexandris, "Secure fingerprint verification based on image processing segmentation using computational geometry algorithms", 2003
[29] C. Lee and D. G. H. Shin, A context-sensitive discretization of numeric attributes for classification learning. In: Proceedings of the eleventh European conference on artificial intelligence. Amsterdam: Wiley; 1994.p. 428-32
[30] C.H. Wu, "Advance Feature Extraction Algorithms for Automatic Fingerprint Recognition Systems", Dissertation on The University of New York, April 2007.
[31] J. Komorowski, A. Øhrn and A. Skowron (2002). 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.

W. Y. Leng received B.S. degree in Computer Science from Universiti Teknologi Malaysia (UTM), Skudai Johor in 2008. Since then, from 2008 to 2009, she was a research assistant at the Department of Software Engineering, UTM. Her work is concerned with the Research and Development in the areas of Speech Recognition. In 2010, she received M.S. degree and she is currently working toward the Ph.D. degree under Soft Computing Research Group at the same University.Her current research interests lies in the domain of pattern recognition include identifying writer based on biometric images and image processing.

S. M. Shamsuddin received her Bachelor and Master degree in Mathematics from New Jersey USA, and Phd in Pattern Recognition & Artificial Intelligence from Universiti Putra Malaysia (UPM), MALAYSIA. Currently, she is a Head of Soft Computing Research Group, k-Economy Research Alliance, Universiti Teknologi Malaysia (UTM), Johor MALAYSIA. Her research interests include the Fundamental Aspects of Soft Computing and its Application, Pattern Recognition, Forensic Document Analysis, and Geometric Modeling. uy