Secure and Efficient Biometric-Data Binarization using Multi-Objective Optimization

Eslam Hamouda [1] **Xiaohui Yuan** [2,3] **Osama Ouda** [1] **Taher Hamza** [1]

[1] *Faculty of Computer and Information Sciences, Mansoura University,*
Mansoura 35516, Egypt

[2] *College of Information Engineering, Chinese University of Geosciences,*
Wuhan,430074,China

[3] *Department of Computer Science and Engineering, University of North Texas,*
Denton 76203, U.S.A.

Received 8 December 2014

Accepted 5 October 2015

Abstract

Biometric system databases are vulnerable to many types of attacks. To address this issue, several biometric template protection systems have been proposed to protect biometric data against unauthorized use. Many of biometric protection systems require the biometric templates to be represented in a binary form. Therefore, extracting binary templates from real-valued biometric data is a key step in such biometric data protection systems. In addition, binary representation of biometric data can speed-up the matching process and reduce the storage capacity required to store the enrolled templates. The main challenge of existing biometric data binarization approaches is to retain the discrimination power of the original real-valued templates after binarization. In this paper, we propose a secure and efficient biometric data binarization scheme that employs multi-objective optimization using Nondominated Sorting Genetic Algorithm (NSGA-II). The goal of the proposed method is to find optimal quantization and encoding parameters that are employed in the binarization process. Results obtained from the experiments conducted on the ORL face and MCYT fingerprint databases show a promising recognition accuracy without sacrificing the security of the system.

Keywords: Biometrics, Binarization, Quantization, Encoding, Multi-objective Optimization

1. Introduction

The growing need of biometrics in access control and verification applications makes security of biometric data a pressing and important issue. Biometric template protection schemes [1,2] have been developed to ensure biometrics privacy and security. The idea behind these schemes is to store an encoded version for the biometric template (the distinct traits extracted from biometric data) rather than the original one by applying a transformation function. Unfortunately, several biometric template protection schemes, such fuzzy commitment [3] and BioEncoding [4] schemes, require the input biometric data to be in a binary form. As a consequence of this limitation, direct application of these schemes is restricted to binary-valued biometric data such as iris-codes [5].

In order to employ such biometric template protection schemes to secure other biometric traits which are usually represented as real-valued templates, such as face and fingerprint, these templates have to be transformed first into binary form. That is, creating such transformation, also known as biometric binarization, is a fundamental and crucial step in many biometric protection systems [6]. In addition, the binarization process produces data representation that usually takes less storage capacity as well as reduced time for matching templates [7], which is vital in situations where less computing resources are available.

A typical biometric template binarization technique involves two stages, namely, quantization and encoding. The goal of quantization is to divide the original feature domain into intervals whereas the goal of encoding is to assign a binary code to each interval. The sheer volume of possible ways of mapping is daunting. Because preserving the discrimination power of the generated binary string is important, both quantization and encoding should provide optimal performance when using hamming distance classifier for comparing the resulting binary strings.

The quantization could be user-specific process, also known as supervised quantization, in which the user feature distribution is used to determine the genuine interval and the remaining intervals are then constructed either using equal width or equal probability approach. Alternatively, the quantization could be user-independent process, in which all intervals are blindly constructed based on specific number of intervals. Using user information in the quantization process enhances the discrimination power of the generated binary template which is likely improves the recognition accuracy. However, the user-specific information used in the quantization must be stored in the system storage in order to construct the same quantization intervals in the verification stages which introduces privacy and security threats [6]. On the other hand, the user-independent approach is more preferable from the privacy and security perspective. However, the recognition accuracy is degraded.

In view of this trade off between performance and security, we propose a new unsupervised binarization method based on an optimization strategy to search for the optimum quantization levels and encoding functions for each feature dimension to achieve balance between the security and the recognition accuracy for the biometric system. The goal of this optimization strategy is to maximize the inter-class distance and minimize the intra-class distance in order to retain the discrimination power for the binary templates. Additionally, the proposed strategy aims at maximizing the entropy of the transformed binary template in order to satisfy the security needs of the biometric systems.

In this work, the proposed binarization strategy is formulated as a multi-objective optimization problem. Classical single objective optimization algorithms do not generate proper optimal solutions in the presence of non-convex search spaces. Besides, determining the appropriate weights that can used to incorporate the conflicting objectives into single objective function is not trivial [8].

In this paper, we make use of the Nondominated Sorting Genetic Algorithm (NSGA-II) [9] which is one of the most popular multi-objective optimization algorithms. NSGA-II is an efficient multi-objective evolutionary strategy that is able to deal with many conflicting objectives and to handle both maximization and minimization problems [8]. The recognition accuracy of the proposed binarization method is evaluated and compared with conventional schemes using the AT&T ORL face [10] and MCYT-fingerprint public databases [11].

The rest of this paper is organized as follows. In Section 2, we briefly describe the related works in biometric binarization, also the fundamental concepts of multi-objective optimization algorithms are presented. The proposed optimization based biometric template binarization method is introduced in Section 3. Experimental results are presented in Section 4. In Section 5, we analyze the security perspective of the proposed binarization method when it is applied in a biometric template protection scheme. Finally, Section 6 concludes this paper and summarizes our results.

2. Background

2.1. Biometric Template Binarization

Biometric binarization can be decomposed into two fundamental components: biometric quantization and intervals encoding. Fig. 1 shows the stages of the biometric binarization process.

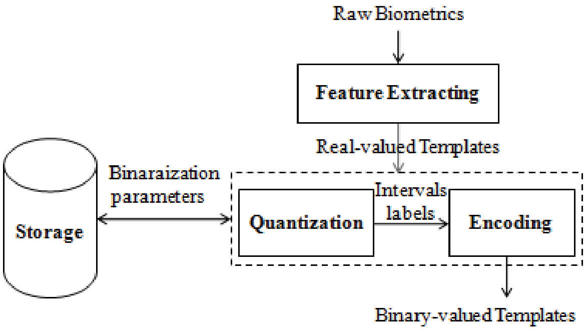


Fig. 1. Biometric binarization process.

As illustrated in Fig. 1, real-valued templates are extracted from the raw biometrics through feature extraction stage. Then, the real-valued feature space is quantized into a set of intervals according to the used quantization design. Each feature element inside interval is then mapped to a short binary string using the encoding function. Eventually, the final binary-valued template is generated by concatenating the binary code for each feature dimension. In order to produce the same binary string for a user in the verification stage, the binarization parameters (quantization and encoding) are stored as helper data in the biometric system.

Linnartz and Tuyls [12] proposed one-bit binarization approach, in which the feature space is modeled by a background probability density function and every feature dimension of the genuine user is modeled by a user probability density function. The value range of features was quantized into fixed intervals each one assigned ‘1’ or ‘0’. The binarization methods proposed in [13,1] adapt multi-bit approaches, in which multiple quantization levels are used for every feature dimension to produce larger size binary string.

To enhance the discriminative power of the generated binary templates, Teoh et al. [14,15] proposed dynamic bit allocation approaches. The fundamental idea was based on assigning different number of bits to feature dimensions according to their discriminative power. Features with lower standard deviation in its probability density function were encoded with more bits. Alternatively, Chen et al. [7] proposed another dynamic bit-allocation approach by considering the detection rate (DR) as their feature discriminative measure. Chen et al. [16] developed a similar dynamic bit-allocation approach based on the area under the false rejection rate (FRR) curve as a discriminative measure.

Discrete binary representation (DBR) has been used in [17,2,7,13,1] to encode the label of quantization intervals using binary values. However, DBR is unstable; it requires at least one bit changes at a time to generate different sequences, which affects the binary template [6]. As an improvement, binary reflected gray code (BRGC) [18,15,16] was used for intervals encoding. BRGC is more stable and suitable than binary encoding. However, the non definite intervals labels encoding in the hamming domain is likely affect the classification performance [6]. Novel encoding functions have been proposed to address the problem of BRGC, known as Partial Linear Separable Sub Code (PLSSC) and Linear Separable Sub Code (LSSC) [19,6,14,20]. PLSSC and LSSC are quite similar to the unary encoding. Although LSSC produces high redundancy bits in the generated binary code compared to PLSSC, it achieves the optimum intervals labels encoding in the hamming domain [19].

Although there are many biometric template binarization methods, achieving balance between security and recognition accuracy for the biometric system is not completely addressed. Therefore, novel methods need to be proposed to convert the real valued templates into robustness binary representation without ignoring the security concerns.

2.2. Multi-Objective Optimization

Real world optimization problems often contain multiple conflicting objectives. To solve a single-objective optimization problem, one attempt is applied to obtain the best single global minimum or maximum solution. Whereas, in multi-objective optimization, there are a set of solutions superior than the other solutions when all objectives are considered and inferior to the other solutions when only subset of the objectives are considered. These solutions are well known as pareto-optimal solutions or nondominated solutions [21]. In multi-objective optimization problems, it is useful to have knowledge about the alternative pareto-optimal solutions to select one of these solutions [22]. Classical optimization methods suggest utilizing the multi-objective problem as single-objective optimization by selecting one particular pareto solution at a time. However, such methods have to be repeatedly applied in order to find a different solution at each simulation run [9]. A number of multi-objective optimization algorithms have been reported [22,9,23,24,25]. The traditional evolutionary algorithms are extended to produce a diverse set of solutions with the ability to move towards the pareto-optimal region in the search space [9]. The primary motivation for these algorithms is their ability to find multiple pareto-optimal solutions in a single run.

GA is one of the evolutionary computation based algorithms that follows Darwin's theory of survival of the fittest. The idea of GA is based on representing the solution space of a given problem as a population of chromosomes (individuals) that reproduce with each other. Over time, the best individuals survive and eventually evolve to do well in the given environment [26]. The standard version of GA require the human user to specify [27]:

(i) The chromosomes (individuals) representation.

(ii) The fitness measure for measuring the fitness of individuals in the population.

(iii) Certain parameters for controlling the run.

(iv) The termination criterion and method for designating the result of the run.

Non-domination sorting genetic algorithm (NSGA) [22] is an extension for the standard GA for multi-objective optimization. Although it is very effective algorithm, it suffers from high computational cost, lack of elitism and the difficulty of choosing the optimal value for the sharing parameter. A modified version, NSGA-II [9] was developed to provide a better sorting algorithm and incorporate elitism. Moreover, NSGA-II does not require choosing a sharing parameter a priori [8]. The basic steps of NSGA-II are illustrated in Fig. 2.

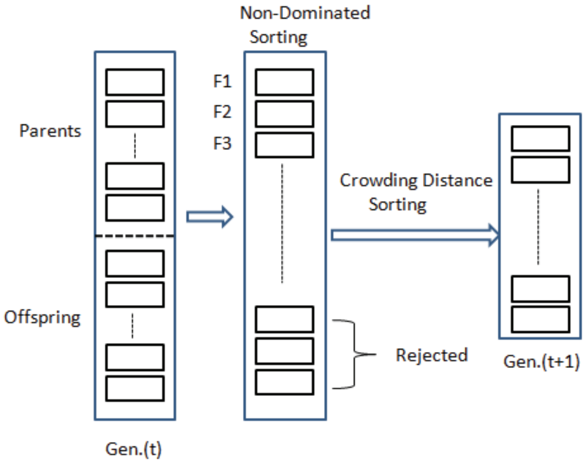


Figure 2: Diagram of NSGA-II Algorithm

After the population is randomly initialized, the individuals are ranked on the basis of nondomination into each front. Individuals in the first front F_1 are entirely non-dominated set in the current population. Individuals in the second front F_2 are only dominated by the individuals in F_1. The process continues until all individuals in the population are ranked. All individuals in F_1 are given rank (fitness) of 1 and individuals in F_2 are assigned rank of 2, and so on. To maintain the diversity of the population, crowding distance is calculated for each individual. The crowding distance measure the similarity between individual and its neighbors. Parents are selected from the population by using tournament selection. An individual is selected if the rank is less than the other or if the crowding distance is greater than the other. The selected population generates offspring using crossover and mutation operators. The current parents and the generated offspring are sorted again on the basis of nondomination and only the best N individuals are used for the new generation, where N is the population size.

3. Biometric Binarization using NSGA-II

The proposed method employs multi-objective optimization to search for the combination of both quantization levels and encoding function such that the discrimination among classes is maximized. Fig. 3 illustrates the stages of the proposed binaraization process. The original crossover and mutation process implemented in [9] are modified to maintain the chromosome structure for the binarization optimization problem, as will be discussed later.

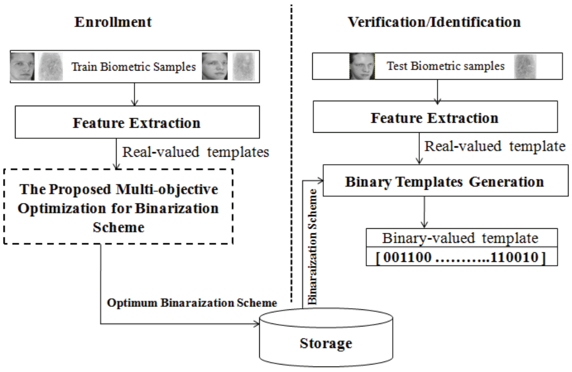


Fig. 3. Our proposed binarization Process.

As shown in Fig. 3, real-valued biometric templates are generated from the original biometric images through feature extracting stage for each enrolled user. Then, using a population of randomly generated binarization schemes, the proposed optimization process starts the evolving stage in order to eventually find the optimum quantization/encoding combination and save it in the biometric system database to be used later in the verification stage.

The proposed binarization scheme (chromosomes representation) consists of both the number of quantization intervals and the encoding function to be used to map the real-valued feature into a binary representation. The quantization intervals ($q_i \in \{2, \dots, Q\}$) and encoding function (e_i) are represented with discrete values. In this work, three encoding functions are implemented to perform the intervals label to binary mapping: {1:BRGC, 2:LSSC, 3:PLSSC}. As mentioned before, DBR is non-stable encoding since it requires at least one bit change at a time to generate different sequences which in turn affects the classification accuracy represented in the generated binary template. Fig. 4 shows an example of a binarization scheme applied to a real-valued biometric template of size 4. Depending on the number of quantization intervals, a sequence of thresholds is randomly selected within the dynamic range $[L_i, H_i]$, where L_i and H_i are the minimum and maximum values for the feature dimension i, respectively. The three thresholds are represented with t_1, t_2, t_3 where $t_j \in [L_i, H_i]$.

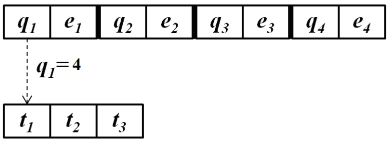
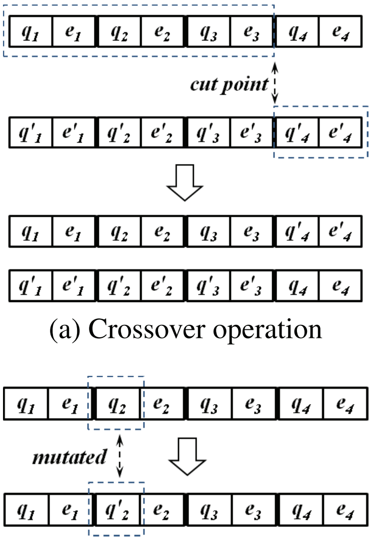


Fig. 4. A binarization scheme with four dimension.

3.1. *Crossover and Mutation Operations*

The crossover operation, $\mathscr{C}(c_a, c_b|\alpha)$, is performed with two randomly selected chromosomes, c_a and c_b. A crossover probability α is used to regulate crossover operations. When crossover is determined not to be conducted, the parent chromosomes are duplicated to the offspring without change. Varying the crossover probability α alters the evolution speed of the search process. In practice, the value of α is close to 1. Although conceptually a crossover operation is performed between two genes, a cutting point that separates the quantization and encoding compartments does not affect the integrity of the chromosome. When the quantization compartments are switched, the respective arrays of thresholds are traded as well. An example of crossover is shown in Fig. 5(a).

The mutation operation, $\mathscr{M}(c_a|\beta)$, involves altering the value at a randomly selected component within the chromosome. If the component gives an encoding scheme, its value is replaced with a different encoding schedule out of the list of options. If the component specifies the number of quantization intervals, randomly change the number of intervals within a finite set of choices, that is, 2 through Q. In addition, a set of thresholds is generated accordingly to ensure the integrity of the chromosome. A mutation probability β is also used to regulate the occurrence of mutation. Different from the crossover probability, the mutation probability is usually fairly small, i.e., $\beta \ll 1$. Essentially mutation operation could create completely new species, that is, an arbitrary locus in the fitness landscape. Hence, it is a means to get out of a local optimum. Fig. 5(b) shows an example of mutation operation.



(a) Crossover operation

(b) Mutation operation

Figure 5: Crossover and mutation operations.

3.2. *Objective Functions*

In the process of searching the appropriate quantization and encoding functions, our main objective aims to maximize the discrimination power among classes without scarifying the security concerns for the biometric system, therefor, the biometric binarization process is formulated as multiobjective optimization problem. The first objective is to minimize the intra-class distance, while the second objective is to maximize the inter-class distance. Finally, maximizing the entropy for the generated binary template to maintain the security concerns.

given K classes and L examples in each class, the intra class distance, denoted by $D_\eta(x_i^{(k)}, x_j^{(k)})$, is the hamming distance of two binary templates $x_i^{(k)}$ and $x_j^{(k)}$, where $i \neq j$, and $i, j \in \{1, \ldots, L\}$, and $k \in \{1, \ldots, K\}$. The inter class distance, denoted by $D_\tau(x_i^{(k)}, x_j^{(k')})$, is the hamming distance of two binary templates $x_i^{(k)}$ and $x_j^{(k')}$ from different classes. By computing the intra class distances between pairs of binary templates, we can construct normalized distance distribution. The problem then becomes finding the optimal separation between the two distributions by maximizing f_1 which is defined by equation(1) and minimizing f_2 which is defined by equation(2)

$$f_1 = |\mu_\eta - \mu_\tau| \tag{1}$$

$$f_2 = \sqrt{\sigma_\eta^2 + \sigma_\tau^2} \tag{2}$$

where μ_η and μ_τ represent the means and σ_η^2 and σ_τ^2 represent the variances of intra and inter distributions, respectively.

It is well known that that the more bit redundancy exist in bit string, the lower entropy for this bit string. Although LSSC achieve perfect encoding capability from the discrimination prospective compared to PLSSC, LSSC has higher bit redundancy compared to PLSSC [19]. An important observation reported in [20]: whenever the genuine interval located in the middle intervals, PLSSC encoding would be able to produce ideal separability quite equivalent to LSSC encoding. The third objective function f_3 is designed to maintain maximum entropy for the generated binary template. The function compute the randomness for the generated binary template which is vital for the security concerns. f_3 is defined by(3):

$$f_3 = -Hlog_2(H) - [(1-H)log_2(1-H)] \tag{3}$$

where H is the ratio of 1's in the binary template. The maximum entropy when $H = 0.5$ (same amount of 1's as 0's).

The objective functions for each chromosome are computed using the concatenated binary code generated by each feature dimension to avoid evaluating the performance of the binarization scheme for each independent feature dimension (local view) which is likely affect the overall performance.

4. Experimental Results and Discussion

This section summarizes the results obtained using the proposed binaraization method. Two experiments are conducted to evaluate the binarization scheme on face and fingerprint biometric data respectively.

4.1. Data set and Experiment Settings

In the first experiment, we used AT&T (ORL) face database [10], which consists of 40 individuals and 10 facial images per individual. Each image has a resolution of 92 by 112 pixels. Sample images of three different human subjects in the ORL face database are shown in Fig. 6. The real-valued biometric templates is extracted following the principal component analysis (PCA) [28].

Fig. 6. Sample images of three human subjects in ORL face database.

To determine the optimum length for PCA component, an evaluation experiment was done. In this experiment, six images of each subject were randomly selected and used as the training set. The rest images were used as the testing set. The experiments were repeated five times using random partitioning for the training and testing samples. Fig. 7 illustrates the average EERs for various number of principal components used in Eigenface method for face recognition [28]. The error bars on top depict the standard deviation. It is clear that the best dimension for PCA components, which yields the lowest EER, is 20.

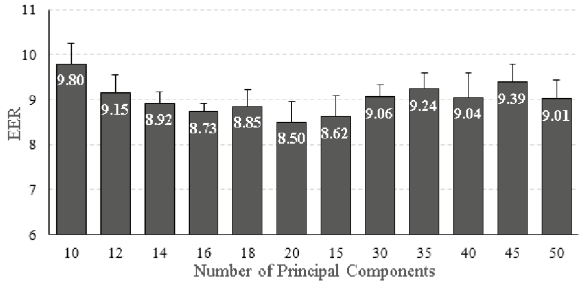


Fig. 7. Average EERs(%) of Eigenface method using different number of principal components.

In the second experiment, we used MCYT-Fingerprint-100 database [11]. MCYT fingerprint database contains 12 different samples of each fingerprint taken under different levels of control (high, medium and low). In each capture session each individual provides a total number of 480 fingerprint images to the database using two different devices/sensors. The file sizes and the image resolutions are 89 kbyte and 300 by 300 pixels when using capacitive device, and 102 kbyte and 256 by 400 pixels in case of using the optical device. For the experiment, the fingerprint region in the raw image are cropped and binarized to produce an image with resolution 96 by 96 pixels, then, real-valued templates are extracted using the following three steps [29]:

(1) Convolution of the cropped fingerprint image with four Gabor filters using orientations (0,p/4,p/2, 3p/4).
(2) Tessellation of the filtered images into equal-sized square disjoint cells with cell size 12 by 12 pixels.
(3) Computation of standard deviation of the pixel intensities for each cell to generate the FingerCode (real-valued template) for each filtered image leading to a total of 256 elements.

To study the effect of our proposed binarization method on MCYT Fingerprint, a subset of 20 subjects (240 fingerprints images) has been randomly selected. The fingerprint matching for the original templates is based on the Euclidean distance between the two corresponding templates. Sample images of different fingers in the MCYT database along with the corresponding cropped binarized images are shown in Fig. 8.

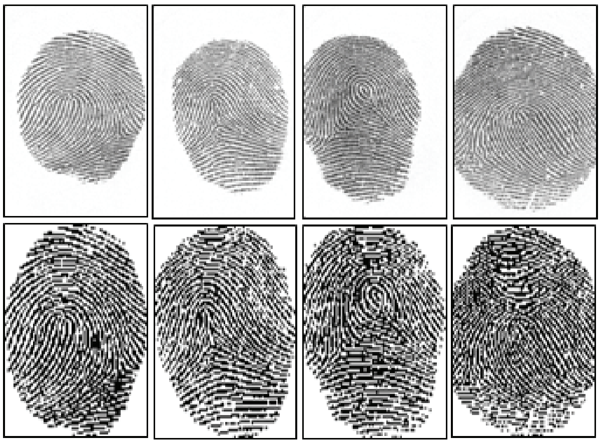


Fig. 8. Sample images of Fingers in MCYT database with the corresponding cropped binarized images.

After the real-valued biometric template is presented, the evolution stage is initiated to search for the optimum binaraization scheme for the real-valued biometric template, five templates for each subject are used for the evolving phase for the optimization process. Table 1 list the parameters used to control the evolving phase. The selection policy for selecting the individuals for the mating pool is based on tournament selection of size five.

Table 1. Parameters used in NSGA-II run.

Parameters	Values
Number of generations	30
Population size	100
Crossover probability	0.8
Mutation probability	0.05
Termination condition	Number of generation=30

4.2. Binary Template Performance Analysis

This subsection evaluates the recognition performance of the generated binary templates using our proposed binarization scheme. We adopted the equal error rate (EER) as a recognition performance measure. EER is the rate where a recognition system exhibits equal false acceptance rate and false rejection rate. The lower EER indicates greater performance of a biometric system.

Table 2 list the EER of our proposed method in comparison with the original real-valued templates. Note that the EER reported in this table is the average EER across all subjects. Given the maximum number of quantization intervals, biometric data was mapped to binary templates. In each repetition, binary templates were created for each biometric image using the maximum number of quantization intervals of 10 and 20 , that is, $q = 10$ and $q = 20$. As shown in Table 2, It is evident that regardless the real-valued template size, the binary templates generated with our proposed method on average achieved better recognition performance compared to the original for both face and fingerprint experiments. For ORL face experiment, The ranges of EERs for binary template are below 7 and in the best case the EERs is 2.87. While, for MCYT fingerprint experiment, The EER for binary template are 4.51 and 4.70 using maximum number of quantization intervals of 10 and 20, respectively.

Table 2. The average EER(%) for face and fingerprint data. q is the maximum number of quantization intervals , N is the real-valued template size.

Data	Size	Original	Binary Template	
			q=10	q=20
ORL [10]	N=10	10.56	2.87	4.13
	N=20	9.2	3.34	3.88
	N=50	9.88	6.31	5.92
MCYT [11]	N=256	6.78	4.51	4.70

To compare the proposed method with well known unsupervised quantization methods which are used in [20], Equal-width quantization with LSSC encoding (EW+LSSC) and Equal-probability quantization with LSSC encoding (EP+LSSC) are selected. Both methods adapt static bit allocation strategy which assigns fixed number of bits for each feature dimension. The same experimental settings are used for these schemes and the results are shown in Tables 3 and 4 for face and fingerprint data, respectively.

Table 3. The average EER(%) using ORL face database. q is the maximum number of quantization intervals , N is the real-valued template size.

Method	N=10		N=20		N=50	
	q=10	q=20	q=10	q=20	q=10	q=20
EW+LSSC	12.89	12.54	12.63	11.4	10.76	10.2
EP+LSSC	11.34	12.16	12.86	11.34	10.86	9.86
Proposed	2.87	4.13	3.34	3.88	6.31	5.92

Table 4. The average EER(%) using MCYT fingerprint database. q is the maximum number of quantization intervals.

Method	q=10	q=20
EW+LSSC	6.86	6.93
EP+LSSC	6.02	5.91
Proposed	4.51	4.70

It can be seen from the previous comparison, the proposed method improves the EER performance of the generated binary templates when compared with the unsupervised: Equal-width quantization and Equal-probability quantization based on static bit allocation strategy and LSSC encoding for both face and fingerprint biometric data.

Another advantage of binary template is the reduced size. Recall the original face image is of 92 by 112 pixels and each pixel is represented with an unsigned integer, that is, 8 bits. Thus, the minimum size on storage or memory is 82,432 bits. As shown in Table 5, When the maximum number of quantization intervals is 10, the average binary template length are 39, 79, and 198 bits in case of real-valued feature size 10, 20, and 50 respectively. If we double the allowed number of quantization intervals, the binary template length increases and the average length binary template length are 71, 145, and 330 bits in case of real-valued feature size 10, 20, and 50 respectively. The improvement is apparent in face data and also for fingerprint experiment.

Additionally, we evaluate the performance of the proposed method in terms of computational time. In Table 5 the average time used to produce a binarization schemes is illustrated.

Table 5. The average used time in minutes and the average binary template length in bits. q is the maximum number of quantization intervals , N is the real-valued template size.

Data	ORL Face					
	N=10		N=20		N=50	
	q=10	q=20	q=10	q=20	q=10	q=20
Length	39	71	79	145	198	330
Time	9.9	11.8	18.4	25.2	45.8	60
Data	MCYT Fingerprint					
	q=10			q=20		
Length	971			1684		
Time	135			178		

Although the overall time used when $q = 20$ is greater than that of $q = 10$, it is interesting to note that in all experiments, the time cost is not proportional to the number of quantization intervals. Additional time was actually used to evaluate individuals fitness. While the population remains the same and so is the number of generations, the average time varies slightly. This could be an advantage of binary template. There is no doubt that generating binary template requires extra resources. However, this process is usually performed off-line and, hence, is manageable.

5. Security Analysis

To analyze the security of our proposed binarization method, a well-known template protection scheme known as fuzzy commitment scheme is selected [3]. Fig. 9 shows a fuzzy commitment scheme with our proposed binaraization method. The input is the real-valued feature vector R, which is extracted from the raw biometric. Using the proposed binarization method, a binary vector R_B is generated. The binaraization scheme is stored as helper data H_1 for each enrolled user. Instead of directly storing the binary vector as a reference template for the enrolled user, an encoded version H_2 is generated and stored as a helper data. The codeword C corresponding to a randomly generated secret K is XOR-ed with the binary vector R_B to obtain H_2, where C is generated using error correction code to deal with the bit errors in the binary vector which is come from the normal measurement noise in the biometric data. The enrolled user identity $h(K)$ is generated by hashing K. In the verification phase this process is reversed using the stored helper data resulting into a candidate user identity $h^*(K)$. Only when the hamming distance difference $d_H(R^e_B, R^v_B) \leqslant \theta$ then $h(K) \equiv h^*(K)$, thus the input user is verified.

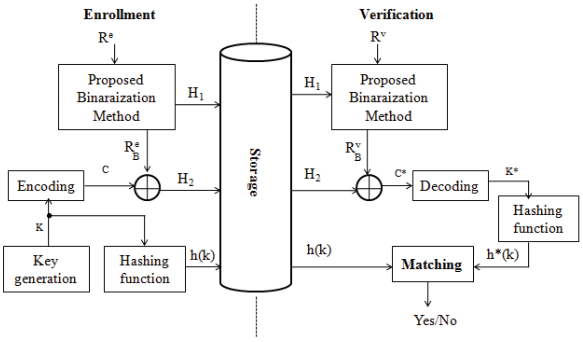


Fig. 9. The Proposed Binarization method in fuzzy commitment Template Protection Scheme.

From the security perspective, our proposed method is vulnerable to the threat of the unauthorized use for the binarization schema which is stored as helper data in the system storage. If the stored binarization scheme is compromised by an attacker, the stored parameters should do not reveal useful information about the original real-valued biometric template R. Also, if the binary vector R_B is determined by a brute force attack, XOR operation with the stored helper data H_2 will reveal the original key K which in turn will be used to compromised the user identity $h(k)$. This section analyzes the efforts needed by an attacker to retrieve the output binary template or the input real-valued biometric template using the compromised data.

Recall that a given NSGA chromosome represents the binaraization scheme. It consists of N genes, N is the real-valued biometric template size (number of feature dimensions). Each gene is considered as independent binaraization scheme for a single feature dimension. The scheme consists of the number of quantization intervals (Q) and the corresponding threshold values (T) to determine the intervals boundaries. Also, the scheme contains the used encoding function (E) which maps each interval index into a binary representation. The number of quantization intervals and encoding function are represented with discrete integer values and the threshold values are represented with real numbers (Fig. 4 shows an example of a binaraization scheme, N=4).

We have number of quantization intervals (Q) where: $Q \in \{q_1, q_2, \ldots, q_N\}$ and $q_i \in \{2, \ldots, \text{max}\}$. The minimum number of quantization intervals is 2, two different experiments on face dataset with max =10 and 20 are explored. Each experiment applied on real-valued biometric template size N=10, 20,50.

5.1. *Analysis of Brute Force Attack*

The more binary template length the more effort needed by an attacker to guess the binary template. The attacker needs maximum number of trails= 2^M to guess a binary template with size= M. In our proposed method, the binary code length for each feature dimension (m_i) depends on the used encoding function (e_i) in addition to the number of quantization intervals (q_i) for feature dimension i. Since we have three possible encoding functions $(e_i) \in \{1 : BRGC; 2 : LSSC; 3 : PLSSC\}$, the binary code length is computed by [20]:

$$m_i = \begin{cases} log_2 q_i & e_i = 1 \\ q_i - 1 & e_i = 2 \\ q_i/2 & e_i = 3 \end{cases} \tag{4}$$

The best case is to maximize the value of m_i in order to maximize the total binary code length (M). According to equation 4, the maximum quantization intervals and the encoding function which yield the largest bit length allow maximum number of trails = 2^{max*N} to guess the binary template. While in the worst case, the minimum quantization intervals and the encoding function which yield the least bit length allow maximum number of trails = 2^N to guess the binary template. Empirically, the least binary template size in our experiment was 30 bit in case of N =10 and the largest binary template size was 375 bit in case of N =50.

5.2. *Security Analysis of Genuine User Intervals*

Given a real-valued template R_1 for a genuine user, a fake real-valued template R_2 could be used by the binarization method such that: B(R_2)= B(R_1), where B is the binarization method, this is a type of the preimage attack [30]. The attacker can guess the location (index) for the genuine quantization interval for each feature dimension. As mentioned before, the threshold values (T) which is used to determine the intervals boundaries are stored in the binarization scheme for each feature dimension. Knowing the intervals boundaries, allow the attacker to generate a fake real-valued template which could go through the compromised binarization scheme to generate a genuine binary template.

Each feature dimension i has q_i different quantization intervals, therefore, the attacker needs maximum number of trails is q_i to guess the genuine interval index for this feature dimension. By multiplying trails numbers for all dimensions, a maximum number of trails is $\prod_{i=1}^{N} q_i$ is needed to guess all genuine interval indexes. In the best case, maximum number of trails is $\prod_{i=1}^{N} max$ is needed to guess genuine interval indexes. While in the worst case, maximum number of trails is $\prod_{i=1}^{N} 2$ is required to guess genuine interval indexes. We con conclude that the larger real-valued biometric template size (N), the more efforts needed to guess the the genuine interval locations.

From the privacy perspective, if the same binarization scheme is stored in different applications for each user, an adversary can track users through applications which introduce privacy threat. To overcome this threat, different binarization schemes should be generated for the user in order to prevent any trial to track users through applications using the stored binarization schemes. Reinitiating NSGA run ensure different binarization schemes for the same user.

6. Conclusion

Biometrics has been widely adopted in various applications, therefore, its security and privacy can not be ignored. Many template protection methods have been developed to ensure biometrics privacy and security using the transformed biometric templates rather than the original ones. Extracting binary biometric strings is a fundamental step in biometric compression and template protection. In this paper, we proposed a novel method to transform the original templates into corresponding binary representation using multiobjective optimization. NSGA-II is used to search for the optimal configuration for the binarization scheme. The objective function is designed to maximize the interclass distance while minimizing the intra-class distance to enhance the discrimination power for the new generated binary templates. Experimental results with the generated binary encoded templates and a hamming distance classifier show superior performance in terms of equal error rate comparing to the original real-valued templates for both face and fingerprint biometric data. Another advantage of binary template is the significantly reduced size compared to the original image.

Acknowledgments

This work was supported by Egyptian Ministry of Higher Education.

7. References

1. Y. Chang, W. Zhang, and T. Chen. Biometric-based cryptographic key generation. In *Proc. the IEEE Intl. Conf. on Multimedia and Expo*, pages 2203 – 2206, 2004.
2. A. Teoh, A. Goh, and D. Ngo. Random multispace quantisation as an analytic mechanism for biohashing of biometric and random identity inputs. *Proc. IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(12), 2006.
3. A. Juels and M. Wattenberg. A fuzzy commitment scheme. In *Proc. the 6th ACM Conf. on Computer and Communication Security*, pages 28–36, 1999.
4. O. Ouda, N. Tsumura, and T. Nakaguchi. Tokenless cancelable biometrics scheme for protecting iris codes. In *Proc. Intl. Conf. on Pattern Recognition*, pages 882–885, 2010.
5. J. Daugman. How iris recognition works. *IEEE Trans. on Circuits and Systems for Video Technology*, 14(1), 2004.
6. M. Lim and A. Teoh. An effective biometric discretization approach to extract highly discriminative, informative, and privacy-protective binary representation. *EURASIP Journal on Advances in Signal Processing*, 2011.

7. C. Chen, R. Veldhuis, T. Kevenaar, and A. Akkermans. Biometric quantization through detection rate optimized bit allocation. *EURASIP Journal on Advances in Signal Processing*, 2009.
8. C.Coello. An updated survey of evolutionary multi-objective optimization techniques: state of the art and future trends. In *Proc. the Intl. Congress on Evolutionary Computations*, 1999.
9. K. Deb, A. Pratap, S.Agarwal, and T.Meyarivan. A fast and elitist multi objective genetic algorithm : Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2), 2000.
10. ORL face image database AT&T Laboratories www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.
11. MCYT-Fingerprint-100 ATVS Biometric Recogntion Group http://atvs.ii.uam.es/databases.jsp.
12. J. Linnartz and P. Tuyls. New shielding functions to enhance privacy and prevent misuse of biometric templates. In *Proc. the 4th Intl. Conf. on Audio and Video Based Person Authentication*, pages 238–250, 2003.
13. F. Hao and W. Chan. Private key generation from on-line handwritten signatures. *Information Management and Computer Security*, 10(4), 2002.
14. M. Lim, A. Teoh, and K. Toh. An analysis on equal width quantization and linearly separable subcode encoding-based discretization and its performance resemblances. *EURASIP Journal on Advances in Signal Processing*, 2011.
15. A. Teoh, W. Yip, and K. Toh. Cancellable biometrics and user-dependent multi-state discretization in biohash. *Intl. Journal of Pattern Analysis and Applications*, 13(3), 2010.
16. C. Chen and R. Veldhuis. Extracting biometric binary strings with minimal area under the FRR curve for the hamming distance classifier. *Signal Processing Journal*, 91(4), 2011.
17. P. Tuyls, A. Akkermans, T. Kevenaar, G. Schrijen, A. Bazen, and N. Veldhuis. Practical biometric authentication with template protection. In *Proc. the 5th Intl. Conf. on Audio- and Video-based Biometric Person Authentication*, pages 436–446, 2005.
18. A. Teoh, K. Toh, and W. Yip. 2n discretization of biophasor in cancellable biometrics. In *Proc. Intl. Conf. on Biometrics*, pages 435–444, 2007.
19. M. Lim and A. Teoh. A novel output label with high separability for biometric discretization. In *Proc. the 5th IEEE Conf. on Industrial Electronics and Applications*, pages 290–294, 2010.
20. M. Lim and A. Teoh. A novel encoding scheme for effective biometric discretization linearly separable subcode. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 35(2), 2013.
21. A. Hans. *Multicriteria optimization for highly accurate systems, Multicirteria optimization in engineering and sciences*. Plenum Press, 1988.
22. N. Srinivas and K. Deb. Multiobjective optimization using nondominated sorting in genetic algorithm. *The Journal of Evolutionary Computations*, 2(3), 1994.
23. M. Fonseca and J. Fleming. Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization. In *Proc. the 5th Conf. onGenetic Algorithms*, 1993.
24. J. Horn, N.Nafpliotis, and E. Goldberg. Multiobjective optimization using the niched pareto genetic algorithm. In *Proc. the 1st IEEE Intl. Conf. on Evolutionary Computations*, 1994.
25. E.Zitzler, K.Deb, and L.Thiele. Comparison of multiobjective evolutionary algorithms: empirical results. *Evolutionary Computation*, 8(2), 2000.
26. M. Mitchell. *An introduction to genetic algorithms*. MIT Press Cambridge, 1996.
27. J.Koza. *Genetic Programming on The Programming of Computers by Means of Natural Selection*. MIT Press, 1992.
28. A. Pentland M. Turk. Eigenfaces for recognition. *Journal of Cognitive Neurosicence*, 3(1):71–86, 1991.
29. F.Fernandez, J.Ortega, J. Fronthaler, H.Kollreider, and K.Bigun. Combining multiple matchers for fingerprint verification: A case study in biosecure network of excellence. *Annals of Telecommunications, Multimodal Biometrics*, 62(2), 2007.
30. P. Rogaway and T. Shrimpton. *Cryptographic Hash-Function Basics: Definitions, Implications, and Separations for Preimage Resistance, Second-Preimage Resistance, and Collision Resistance*. Springer-Verlag, 2004.