Discriminative and Non-User Specific Binary Biometric Representation via Linearly-Separable SubCode Encoding-based Discretization

Meng-Hui Lim and Andrew Beng Jin Teoh
School of Electrical and Electronic Engineering, College of Engineering, Yonsei University, Seoul, South Korea.
[e-mail: menghui.lim@gmail.com, bjteoh@yonsei.ac.kr]
*Corresponding author: Andrew Beng Jin Teoh

Received November 13, 2010; revised December 15, 2010; accepted January 9, 2011; published February 28, 2011

Abstract

Biometric discretization is a process of transforming continuous biometric features of an identity into a binary bit string. This paper mainly focuses on improving the global discretization method – a discretization method that does not base on information specific to each user in bitstring extraction, which appears to be important in applications that prioritize strong security provision and strong privacy protection. In particular, we demonstrate how the actual performance of a global discretization could further be improved by embedding a global discriminative feature selection method and a Linearly Separable Subcode-based encoding technique. In addition, we examine a number of discriminative feature selection measures that can reliably be used for such discretization. Lastly, encouraging empirical results vindicate the feasibility of our approach.

Keywords: Biometric, discretization, quantization, encoding, linearly separable subcode

DOI: 10.3837/tiis.2011.02.008

1. Introduction

Biometric has emerged as a promising surrogate for conventional identity representation mechanisms such as password (i.e. pin number) and token (i.e. ID card) due to its merits of being representative and convenient. Nonetheless, being inextricably linked to an user, biometric once compromised can never be reissued or replaced and thereby must be placed under careful protection when it is deployed in a recognition system.

Many verification systems integrate biometrics into their applications to provide a greater level of security and conveniency. However, many applications such as biometric-based cryptographic key generation schemes [2][5][8][14][15][17][22] and biometric template protection schemes [9][10][12][20][21] usually only work with binary secret. While most biometric modalities represent each identity using a set of continuous features intrinsically upon a preliminary feature extraction, these features need to be subsequently converted to a binary string through a transformation process called *biometric discretization*. **Fig. 1** depicts the block diagram of a binary string generation in a biometric verification system that uses a biometric discretization scheme as a basis.

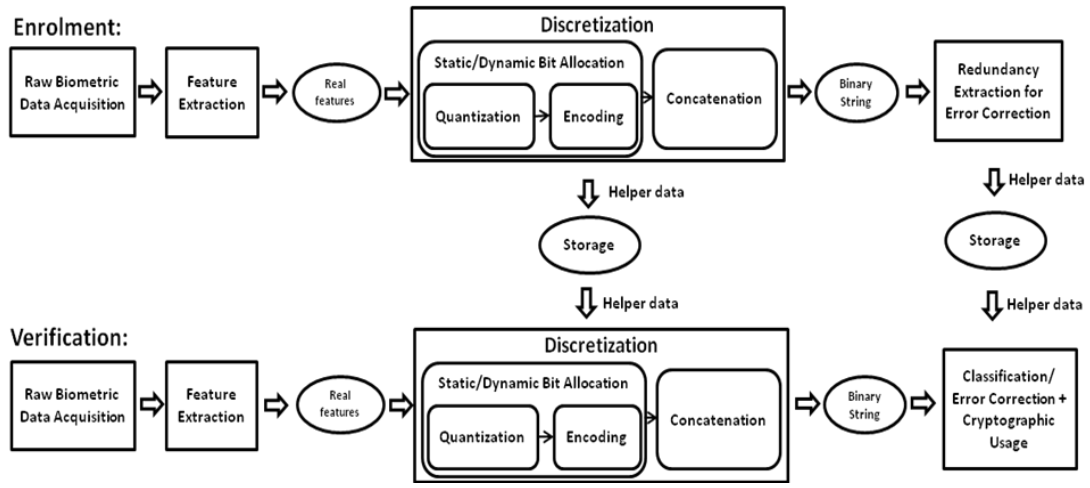


Fig. 1. Binary string generation in a biometric verification system.

In general, biometric discretization can be decomposed into two essential components, which can be alternatively described as a two-stage mapping process:

- Quantization: This first component can be seen as a continuous-to-discrete mapping process. Given a set of feature elements per identity, every one-dimensional feature space is initially constructed and segmented into a number of non-overlapping intervals which are associated to a set of decimal indices. Popular quantization techniques comprise equal-width quantization and equal-probable quantization. The former partitions each one-dimensional feature space into multiple non-overlapping equal-width intervals; while the latter partitions each one-dimensional feature space into multiple intervals according to the background probability distribution so that every interval encompasses the same amount of background probability mass. This paper focuses on the former due to stronger security provision.

- Encoding: The second component can be regarded as an discrete-to-binary mapping process, where the resultant index of each dimension is mapped to a unique *n*-bit binary codeword of an encoding scheme. The codeword output of every feature dimension is then concatenated to form the final bit string of an identity. The discretization performance is then evaluated in the Hamming domain. The existing encoding schemes for biometric discretization include Direct Binary Representation [2][8], Binary Reflected Gray Code [6] and Linearly Separable SubCode [11].

Apart from that, information regarding the constructed intervals in each dimension, also known as *interval information*, is stored as the *helper data* along with the relevant parameters during enrolment. In the verification phase, this helper data is used to reproduce the same binary string of every genuine user. Note that it is important to prevent such helper data from leaking any helpful information regarding the output binary string (security concern) and the biometric feature itself (privacy concern) so that such leakage would not happen under the worst case scenario where a biometric system is compromised.

- **Security**: Binary output of every user should contain an adequately high amount of entropy and this entropy should not be undermined by analyzing the helper data. In the context of biometric discretization, the entropy of each user's binary output is usually formed by concatenating entropy offered by every feature dimension of that user. Depending on the probability of occurrence p_i of every binary output i, the entropy of a feature dimension is calculated by $l = -\sum_{i=1}^{S} p_i \log_2 p_i$. If the probability of the binary outputs is equal, then the entropy can be simplified as $l = -\log_2 p_i$ for any $i \in \{1, ..., S\}$. Thus, the more the quantization intervals are constructed in each dimension when the more the binary outputs/the longer the informative bits need to be derived from a dimension. Proportionally, the probability p_i for every binary output gets smaller and the higher the entropy would be against any adversarial brute force attack (more secure).
- **Privacy**: To avoid devastated consequence upon compromise of the irreplaceable biometric features of every user, the helper data must not be correlated to the raw or projected features. For example, the interval information must not be derived in accordance with the exact location of the features to avoid such vital sensitive information being leaked via mere observation. Thus, the non-invertibility of the helper data should be guaranteed in order to impede any adversarial reverse engineering attempt in obtaining the raw features. Otherwise, it has no difference from storing the biometric features in the clear in the system database.

1.1 Related Work

In general, biometric discretization can be dichotomized into global discretization and user-specific discretization. A global discretization uses a common set of helper data to derive a bit string for every user in a system; while a user-specific discretization extracts each bitstring based on information that is specific to the corresponding user.

In fact, each of these discretization schemes offers several advantages which could be very useful for different applications. For instance, since distinctive local information of each user is utilized in bitstring generation, user-specific discretization schemes are typically able to outperform global schemes and thus are appropriate for applications that desire a discriminative performance.

On the other hand, global discretization does not contain any user-specific information. Therefore, it could be a promising option when privacy protection is a priority in an

application. Additionally, in some user-specific discretization schemes [2][3], boundary intervals are left unused with the purpose of best fitting the genuine user pdf in an interval (called the *genuine interval*) in each dimension. Since a common codeword associated with the boundary intervals is excluded from the guessing range, these schemes experience a non-trivial entropy loss in each dimension. Unlike such schemes, global discretization completely utilizes the elements of an encoding scheme, giving an adversary no chance to eliminate any codeword from his guess. Thus, global discretization does not trade security for performance, making it a better choice of discretization for cryptographic application.

Numerous user-specific biometric discretization schemes have been proposed over the past decade but only a little attention was paid to the global schemes. A global 1-bit discretization scheme fundamentally employed by Monrose et al. [14][15], Teoh et al. [17] and Verbitsky et al. [21] partitions each feature space into two intervals which are labelled using a simple set of '0' and '1' based on a threshold. As the scheme is global, every user employs the same quantization setting, and thus do not require any user-specific helper data to be stored.

Han et al. [7] extracts a 9-bit pin from users' fingerprint impressions based on the global discretization setting. The binary pin extraction process can be divided into two parts:

(a) The derivation of the first 6 bits comes from the 6 pre-identified reliable/stable minutiae: if a minutia belongs to bifurcation, a bit '0' is assigned; if it is a ridge ending, a bit'1' is assigned.

(b) The derivation of the last 3 bits is constituted by a one-bit discretization on length of the maximal side, a one-bit discretization on the median angle and a one-bit discretization on the minimal angle of the triangular feature. Note that these discretizations are applied to the tenth digit (in decimal representation) of the feature value. For instance, for any non-negative integer x and y, the feature range of [xy.0, xy.49] is mapped to a bit '0' while the feature range of [xy.5, xy.99] is mapped to a bit '1'.

Tuyls et al. [20] and Kevenaar et al. [10] used a user-specific 1-bit discretization technique by selecting the feature mean of the entire user set as the threshold for each dimension in their biometric discretization scheme. This scheme identifies a distinct set of reliable components from either the training bit statistics [20] or a reliability function [10] for every user so that such local information can be utilized to enhance the discretization performance.

However, the incapability of these schemes in meeting the increasing entropy security requirement has driven the researchers to consider the need of generating informative binary stream through extracting multiple bits from each feature space. Hao-Chan [8] employed a user-specific multi-bit discretization scheme, of which the genuine interval is determined as $\left[\mu - k\sigma, \mu + k\sigma\right]$, where μ and σ denote the mean and the standard deviation of the user distribution, and k is a free parameter. The remaining intervals are constructed based on a constant width of $2k\sigma$. Chang et al. [2] introduced a similar scheme to Hao-Chan's scheme [8]. This scheme extended the real feature space of every dimension to account for the extra equal-width intervals in order to form a total of 2^n intervals which are labelled by n-bit direct binary representation (DBR) encoding elements (i.e. $3_{10} \rightarrow 011_2$, $4_{10} \rightarrow 100_2$, $5_{10} \rightarrow 101_2$).

Yip et al. [22] employed a global multi-bit discretization scheme based on an equal-width intervals construction. In the scheme, every one-dimensional feature space is partitioned into 2^n equal-width intervals which are labelled using a *n*-bit Binary Reflected Gray Code (BRGC) [6] encoding scheme (i.e. $3_{10} \rightarrow 010_2$, $4_{10} \rightarrow 110_2$, $5_{10} \rightarrow 111_2$).

Chen et al. [3] demonstrated another user-specific multi-bit discretizer based on likelihood-ratio. The involved quantization scheme constructs intervals in an equal-probable manner where the background probability mass is equally distributed within each interval. The

leftmost and rightmost intervals with insufficient background probability mass are wrapped into a single interval which is tagged with a common label. Similar to Yip et al.'s scheme, BRGC is adopted for the encoding usage.

Other user-specific discretization schemes based on more complicated dynamic bit allocation methods include those which are illustrated in [4][16][18].

1.2 Motivations and Contributions

As we have seen in the literature, there is a lack of effort in developing a superior global discretization scheme that offers a good recognition performance with satisfactory security provision and privacy protection practically. In fact, only Yip et al. [22] has contributed in such relevant development. Moreover, it has also been reported that discretization performance of all the above schemes is susceptible to deterioration whenever a high entropy binary representation is needed [11].

In view of the vital importance of such discretization scheme in fulfilling strong security provision and privacy protection, this paper proposes a standard multi-bit global discretization strategy that could always offer an improved performance through adopting the use of a recently proposed encoding scheme, known as Linearly Separable SubCode (LSSC) [11] and a global feature selection method.

Note that the proposed methodology is unique in the sense that the adopted LSSC encoding scheme is not substitutable with any other encoding scheme in achieving an improved performance because a very important property known as the "ideal separability" attribute needs to be satisfied by the encoding scheme in order to allow the feature selection to be carried out in an effective manner.

Moreover, the global feature selection method which we are going to deal with is somewhat novel comparing to the conventional user-specific feature selection method. A user-specific method usually selects a different set of discriminative feature dimensions for each user and thus is unsuitable to be employed when a global discretization is desired. To overcome such inappropriateness, we alter the usual way which the dimensions are selected by picking a pre-specified number of dimensions with the highest quantity of discriminative features across all the users in a system to be the common discriminative dimensions.

On the whole, the contribution of this paper is two-fold:

- We demonstrate a novel global feature selection technique to select discriminative dimensions that contain most discriminative features among the users.
- We combine such a feature selection technique with Linearly Separable SubCode encoding to achieve a greater discretization performance.

The structure of this paper is organized as follows. In the next section, more details regarding our approach are given. In section 3, experimental results justifying the effectiveness of our approach are presented. Lastly, several concluding remarks are provided in section 4.

2. The Proposed Approach

Suppose that a total of J users are enrolled in a system with each of them represented by D ordered feature elements upon a preliminary feature extraction. In view of potential intra-class variation, the *d*-th feature element of the *j*-th user can be modeled by a user pdf, denoted by $f_j^d(v)$ where $d \in \{1,2,...,D\}$, $j \in \{1,2,...,J\}$ and $v \in$ feature space V^d. On the other hand,

due to inter-class variation, the d-th feature element of the entire population can be modeled by a background pdf, denoted by $f^d(v)$. Both distributions are assumed to be Gaussian. That is, the d-th dimensional background pdf has a mean μ^d and a standard deviation σ^d while the j-th user's d-th dimensional user pdf has a mean μ_j^d and a variance σ_j^d.

Our strategy of enhancing the performance of state-of-art global discretization encompasses two necessary criteria:

a) [Global Feature Selection] Select D_{fs} ($D_{fs} < D$) most discriminative dimensions which are common to all users in a system from the D initial dimensions produced by the preliminary feature extraction process in accordance with the discriminative measure in use. In particular, all discriminative measurements of all users in each dimension are summed up initially to give the final measurement value for actual discriminative evaluation. Then, D_{fs} highest final measurement values will be taken as the common discriminative dimensions for all users.
b) [Encoding Scheme] Employ LSSC encoding for discretization to preserve possibly superior performance led by the prior discriminative feature selection disregarding the entropy requirement imposed on the discretized output.

2.1 Discriminative Measures for Global Feature Selection

In this subsection, we identify and suggest two potentially reliable discriminative measures from the user-specific schemes in the literature, where any of such could be applied in our global feature selection process efficiently, namely reliability and detection rate. The descriptions of these measures are provided as follows.

2.1.1 Reliability (RL)

Reliability is an efficient discriminative measure proposed by Kevenaar et al. [10] to sort the discriminability of the feature components in their user-specific 1-bit discretization scheme. The definition of this measure is given by

$$RL_j^d = \frac{1}{2}\left(1+erf\left(\frac{\mu_j^d-\mu^d}{\sqrt{2\left(\sigma_j^d\right)^2}}\right)\right),\ j\in\{1,2,...,J\},\ d\in\{1,2,...,D\} \quad (1)$$

where *erf* denotes the error function. This reliability measure would produce a higher value when a feature element has a larger difference between μ_j^d and μ^d relative to σ_j^d. As a result, a high reliability measurement indicates a high discriminating power of a feature component.

2.1.2 Detection Rate (DR)

Detection rate is a discriminative measure which considers an additional factor in discriminativity evaluation of each feature component – the position of the user pdf with reference to the constructed genuine interval, on top of statistical information of the pdfs in each dimension. This measure was adopted by Chen et al. in their dynamic bit allocation scheme [4] and is defined as the area under curve of the user pdf enclosed by the genuine interval in a dimension. It can be described mathematically by

$$\delta_j^d(S^d) = \int_{\text{int}_j^d} f_j^d(v)dv \tag{2}$$

where δ_j^d denotes the *j*-th user's detection rate of the *d*-th dimension, int_j^d denotes the the *j*-th user's genuine interval of the *d*-th dimension and S^d denotes the number of constructed intervals in the *d*-th dimension.

2.2 Linearly Separable SubCode (LSSC) Encoding Scheme

Generally speaking, Linearly Separable SubCode (LSSC) [11] is an encoding scheme that was introduced to replace Direct Binary Representation (DBR) and Binary Reflected Gray Code (BRGC) schemes in discretization for efficient classification purpose.

By viewing a discretization as a two stage mapping process – the continuous-to-discrete map (quantization) and the discrete-to-binary map (encoding), LSSC has an exclusive ability to preserve completely the separation among the feature points during the discrete-to-binary map as the ultimate distance evaluation is performed in the Hamming domain. On the other hand, LSSC might need to utilize a large amount of bits redundancy to expand such separability in the Hamming space to enable a one-to-one correspondence between every non-reference codeword and the Hamming distance incurred with respect to every possible reference codeword. It is reported in [11] that the more such bit redundancy is used, the higher the improvement of discretization performance could result as compared to that of DBR and BRGC. (See [11] for a more detailed explanation about the advantage of LSSC over DBR and BRGC in biometric discretization).

LSSC has a code length of n_{LSSC} and it consists of $S = (n_{LSSC} + 1)$ codewords which happens to be a subset of $2^{n_{LSSC}}$ codewords. The bit redundancy *r* can simply be quantified by $r = n_{LSSC} - \log_2(n_{LSSC} + 1)$ bits. Here, we provide a brief construction of LSSC as follows: Beginning with an arbitrary n_{LSSC}-bit codeword, say an all zero codeword, the next n_{LSSC} codewords can be sequentially derived by complementing a bit at a time from the lowest (rightmost) to the highest order (leftmost) bit position. The resultant n_{LSSC}-bit LSSCs for the specified *S* = 4, 8 and 16 are shown in **Table 1**.

Table 1. A collection of n_{LSSC}-bit LSSCs for *S* = 4, 8 and 16 with $[\tau]$ indicating the codeword index.

$n_{LSSC} = 3$ $S = 4$	$n_{LSSC} = 7$ $S = 8$	$n_{LSSC} = 15$ $S = 16$	
[0] 000	[0] 0000000	[0] 000000000000000	[8] 000000011111111
[1] 001	[1] 0000001	[1] 000000000000001	[9] 000000111111111
[2] 011	[2] 0000011	[2] 000000000000011	[10] 000001111111111
[3] 111	[3] 0000111	[3] 000000000000111	[11] 000011111111111
	[4] 0001111	[4] 000000000001111	[12] 000111111111111
	[5] 0011111	[5] 000000000011111	[13] 001111111111111
	[6] 0111111	[6] 000000000111111	[14] 011111111111111
	[7] 1111111	[7] 000000001111111	[15] 111111111111111

2.3 Some Discussions and a Summary of our Approach

In a typical biometric cryptosystem, an entropy requirement *L* is usually imposed on the

binary output of the discretization scheme. For equal-probable quantization, L is equally divided by D dimensions for typical discretization schemes and by D_{fs} dimensions for our feature selection approach when fixed-bit allocation principle is based upon. Since the entropy

Algorithm 1: The Proposed Global Discretization Approach

For a user $j \in \{1, \dots, J\}$,

Input:

R-dimensional raw features of I measurements: $\boldsymbol{U}_j = \{u_{ij}^d | i = 1, \dots, I, d = 1, \dots, R\}$,
Number of extracted dimensions: D,
Number of globally selected discriminative dimensions: D_{fs},
Entropy per dimension: l_{fs},
Feature extraction function: $\mathfrak{F}(\cdot)$, e.g. PCA [19], FDA [1]
Discriminative measure: $\chi(\cdot)$, e.g. RL [10], DR [4]
Continuous-to-discrete mapping function: $\mathcal{Q}(\cdot)$,
LSSC encoding-based discrete-to-binary mapping function: $\mathcal{E}_{LSSC}(\cdot)$.

Initialize:

$$\mathfrak{D} = \{\emptyset\}.$$

Feature Extraction:

$$\boldsymbol{V}_j = \{v_{ij}^d | i = 1, \dots, I, d = 1, \dots, D\} = \mathfrak{F}(\boldsymbol{U}_j).$$

Global Discriminative Feature Selection:

$$\zeta = [\zeta^1, \dots, \zeta^D]^T = \left[\sum_{j^*=1}^{J} \chi(v_{1j^*}^1, v_{2j^*}^1, \dots, v_{Ij^*}^1), \dots, \sum_{j^*=1}^{J} \chi(v_{1j^*}^D, v_{2j^*}^D, \dots, v_{Ij^*}^D)\right]^T,$$

for $\partial = 1 : D_{fs}$

$$\partial^* = \arg\max_{\partial^* \in \mathfrak{D}} [\zeta]$$
$$\mathfrak{D} = \{\mathfrak{D}, \partial^*\}.$$

end for

Equal Probable Quantization & Encoding:

Number of equal-probable outputs in each dimension: $S_{fs} = 2^{l_{fs}}$,
Number of bits assigned to each dimension: $n_{LSSC(fs)} = 2^{l_{fs}} - 1$,

for $\partial = 1 : D_{fs}$

$$i_j^{\partial} = \mathcal{Q}\left(x_j^{\mathfrak{D}(\partial)}, S_{fs}\right),$$
$$b_j^{\partial} = \mathcal{E}_{LSSC}\left(i_j^{\partial}, n_{LSSC(fs)}\right).$$

end for

Output:

Helper data: $help_j = \{D_{fs}, \mathfrak{D}, l_{fs}, \text{interval information}\}$,
Final bitstring: $B_j = \{b_j^1 || \, b_j^2 || \, \dots || \, b_j^{D_{fs}}\}$.

per dimension l is logarithmically proportional to the number of outputs S or l_{fs} & S_{fs} for our approach) constructed in each dimension, this can be written as

$$l = L / D = \log_2 S \text{ for typical discretization scheme; or} \quad (3)$$

$$l_{fs} = L / D_{fs} = \log_2 S_{fs} \text{ for our feature selection approach} \quad (4)$$

By denoting n as the bit length of each one-dimensional binary output, the actual bit length N of the final bitstring can be described by

$$N = Dn \quad (5)$$

For LSSC encoding-based schemes where $n_{LSSC} = (2^l - 1)$ bits and $n_{LSSC(fs)} = (2^{l_{fs}} - 1)$ bits,

we have

$$N_{LSSC} = Dn_{LSSC} = D(2^{l} - 1) \tag{6}$$

or

$$N_{LSSC(fs)} = D_{fs} n_{LSSC(fs)} = D_{fs} (2^{l_{fs}} - 1) \tag{7}$$

With these, the full algorithmic description of our approach is illustrated in Algorithm 1. Note that ∂ and ∂^{*} are dimensional variables and || denotes a binary concatenation operator.

3. Experiments and Discussions

3.1 Data Set

To evaluate and to justify the performance superiority of our approach with reference to the existing global discretization schemes in particular, our experiments were conducted based on a popular face data set. The data set which we have adopted is a subset of the AR face data set [13]. This data set contains a total of 684 images which belong to 114 identities with 6 images per person (3 images for training and the other 3 for testing). The images were taken under strictly controlled conditions and particularly, those images which feature frontal view faces with different facial expressions and moderate illumination variations were selected for the experiments. The images are aligned according to standard landmarks, such as eyes, nose and mouth. After preprocessing, each extracted raw feature vector consists of 46 x 56 grey pixel elements.

Fig. 2. Some sample images in AR face data set.

3.2 Experimental Settings

In order to evaluate the False Acceptance Rate (FAR) of a system, each identity's image is matched against every other identity's image according to the corresponding image index. As for the False Rejection Rate (FRR) evaluation, each image is matched against every other images of the same identity for all identities. In the subsequent experiments, the equal error rate (EER) (error rate where FAR = FRR) are used for comparing the discretization performance among different encoding schemes, since it is a quick and convenient way to compare the performance accuracy of the discretizations. That is, the lower the EER is, the better the performance is considered to be and vice versa.

The experiment was carried out based on 2 different dimensionality reduction techniques:

Principal Component Analysis (PCA) [19] and Fisher Linear Discriminant Analysis (FDA) [1]. In the experiment, 2576 raw dimensions of AR images were reduced to D = 100 dimensions. For the feature selection schemes in our approach, D_{fs} is fixed as 50. Equal probable quantization is adopted throughout the experiment.

3.3 Experimental Subjects

The global discretization schemes which were involved in the performance evaluation are:

(A) Equal Probable (EP) quantization and Direct Binary Representation (DBR) encoding-based discretization (EP+DBR)
(B) Equal Probable (EP) quantization and Binary Reflected Gray Code (BRGC) encoding-based discretization (EP+BRGC) [3][10][20]
(C) Equal Probable (EP) quantization and Linearly Separable SubCode (LSSC) encoding-based discretization (EP+LSSC) [11]
(D) Maximum Reliability-based Global Feature Selection (GFS)-incorporated Equal Probable quantization with Linearly Separable SubCode encoding-based discretization (Max RL-based GFS+EP+LSSC)
(E) Maximum Detection Rate-based Global Feature Selection (GFS)-incorporated Equal Probable with Linearly Separable SubCode encoding-based discretization (Max DR-based GFS+EP+LSSC)

Note that the first three schemes are the conventional global discretization schemes while the last two schemes are the ones which adopt our approach.

Recall that LSSC encoding utilizes bit redundancy to achieve full preservation of the separation of feature points during the discrete-to-binary map and therefore a much larger bit length of the binary output might be resulted when a system-specified (per-dimensional) entropy is imposed. This is different from the case of DBR and BRGC encoding where the resultant length of the binary output of each dimension will be equivalent to the actual bit entropy offered by each dimension. Thus, it will be inappropriate to compare the performance through equalizing the bit length of the binary strings generated by different encoding schemes, since the dimensions utilized by LSSC will be much lesser than that by DBR and BRGC at common lengths.

Perhaps the best way to compare these discretization schemes with equal probable quantization would be in terms of the entropy L of the final bit string. Recall that the entropy of a bitstring

$$L = \sum_{d=1}^{D} l = \sum_{d=1}^{D_{fs}} l_{fs} \tag{8}$$

and the actual bit length

$$N = Dn = D_{fs} n_{fs} . \tag{9}$$

Therefore, given a specific entropy requirement L, for (A) BRGC or (B) DBR encoding-based discretization where $n = l$ bits, we have

$$N = Dn = Dl = L. \tag{10}$$

For (C) LSSC encoding-based discretization where $n_{LSSC} = (2^l - 1)$ bits, we have

$$N_{LSSC} = Dn_{LSSC} = D(2^l - 1) . \tag{11}$$

For our approaches in (D) & (E) where $L = D_{fs} l_{fs}$, $D_{fs} = \frac{D}{2}$, $l_{fs} = 2l$ and $n_{LSSC(fs)} = 2^{l_{fs}} - 1$ $= 2^{2l} - 1$ bits, the actual bit length can be quantified by

$$N_{LSSC(fs)} = D_{fs} n_{LSSC(fs)} = \left(\frac{D}{2}\right)(2^{2l} - 1) \tag{12}$$

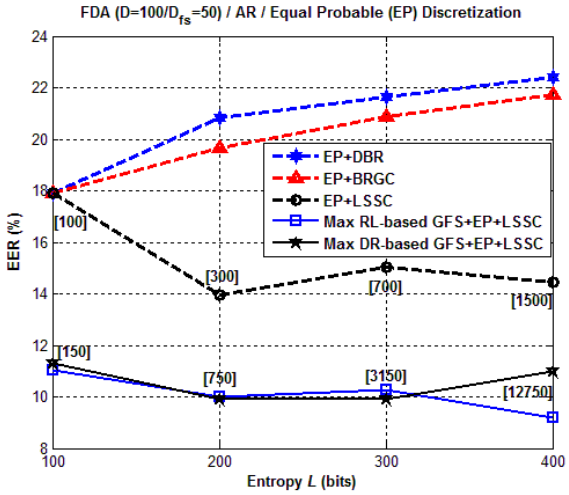
3.4 Performance Evaluation

Fig. 3 depicts the EER performance of several global schemes discretizing PCA- and FDA-extracted features from the AR face data set based on equal probable quantization. As depicted in **Fig. 3**, it is clear that all DBR and BRGC encoding-based discretization schemes exhibit a gradually deteriorating EER performance as the entropy requirement increases, due to their inability in preserving the separation among the feature points during the discrete-to-binary map. By being able to overcome such drawback, LSSC encoding-based discretization schemes, on the other hand, achieve the lowest EER among the schemes without feature selection capability when $l > 1$ or $L > 100$. It is worth a note that in **Fig. 3-(a)**, EP+LSSC outperforms EP+DBR and EP+BRGC schemes by 8% EER at $L = 300$; and approximately 9.5% EER at $L = 400$. In **Fig. 3-(b)**, the outperformance is averagely 6% EER at $L = 300$; and averagely 7.5% EER at $L = 400$. Due to perfect separation-preservation ability, the performance of EP+LSSC is almost unaffected by the increasing entropy beyond $L = 300$. In other words, as the performance of DBR and BRGC encoding-based schemes deteriorate more as entropy requirement increases, the difference in performance compared to that of EP+LSSC will become more significant along with the increase of entropy.

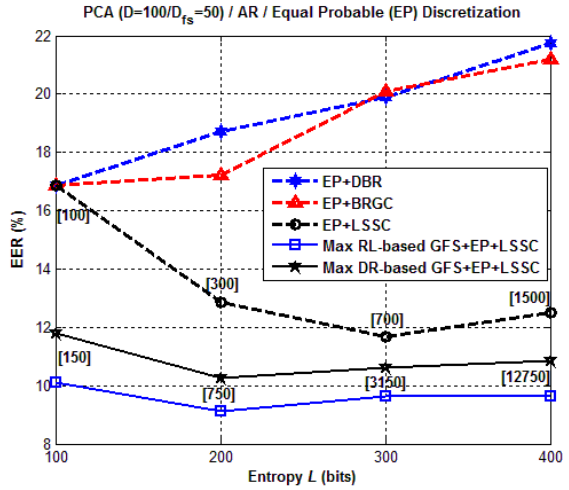
When a global feature selection mechanism is embedded into EP+LSSC, the performance, being independent of the discriminative measure, stabilizes earlier at $L = 200$ compared to that of EP+LSSC correspondingly. This could be explained by that $L = 200$ corresponds to $l = 2$ for EP+LSSC while it corresponds to $l_{fs} = 4$ for GFS+EP+LSSC respectively. This implies that the segmentation in each selected dimension is increased much faster than that of EP+LSSC along with the augmentation of entropy requirement, eventually leading to the increase in the number of codewords needed as well as the length of the binary output.

In terms of performance superiority, RL-based GFS+EP+LSSC achieves the best performance enhancement overall compared to EP+LSSC respectively, that is, by {7%, 4%, 4.5%, 4.5%} in **Fig. 3-(a)** and {7%, 4%, 2%, 2.5%} in **Fig. 3-(b)** corresponding to $L = \{100, 200, 300, 400\}$. It is noticeable that the performance improvement achieved by DR-based GFS+EP+LSSC is also very close to that of RL-based GFS+EP+LSSC, justifying the feasibility of our approach.

It is observed that improvement by RL and DR discriminative feature selections in PCA-based discretization in **Fig. 3-(b)** is slightly less significant compared to those in FDA-based discretization in **Fig. 3-(a)**. This could be influenced by that decision made by a feature selecting process on a given set of features may not be ideal due to indefinite pdf estimation from a limited number of training samples. Some indiscriminative feature dimensions may be mistakenly selected. Vice versa, some moderately discriminative dimensions may be excluded by mistake due to the similar reason. Therefore, to what extent the influence of a feature selection on a certain baseline performance would greatly depend on the accuracy of the pdf estimation which could range distinctively in accordance with different extracted set of features.



(a)



(b)

Fig. 3. EER performance of several global discretization schemes using equal-probable quantization based on PCA- and FDA-extracted features from AR dataset in satisfying several different entropy requirements. $[\alpha]$ associated with each measurement of LSSC-based discretization scheme indicates the corresponding length α of the binary output. For DBR and BRGC-based discretization schemes, the length of the binary output is equivalent to the length of the entropy.

An alternative way of saying this is that the quality of the unselected feature dimensions decides the amount of improvement with respect to the baseline. If the excluded feature dimensions are truly the least discriminative dimensions, the improvement will be the greatest. Otherwise, if the excluded feature dimensions are somehow discriminative, the improvement will be minor, or even worse, performance deterioration could occur. This signifies that substantial deviation between the estimated user pdf from the training samples and the true user pdf should be avoided in order to circumvent such trivial improvement or deterioration scenarios. To achieve this, utilizing as many training samples as possible with adequate preprocessing steps for the estimation process is a good approach of how the estimation error can be minimized. Our observation in Fig. 3 infers that most FDA-extracted features have

extreme discriminativity (both highly discriminative and less disciminative features) while PCA-extracted features are of mostly moderately discriminative features. Thus, precisely retaining highly discriminitave and excluding less discriminative features enables FDA-based feature selection scheme to achieve a better discretization performance and attain a larger improvement (compared to that without feature selection) than PCA-based feature selection scheme.

Perhaps the only limitation arose in achieving such improvement is the inevitable derivation of large length binary string per user when a high entropy strength is desired by a system. As shown in **Fig. 3**, a bitstring with at least 4 times longer than the entropy is needed to fulfil a 300-bit entropy strength, while bit string that is at least 8 times longer is required to fulfil a 400-bit system-specific entropy strength. Indeed, these amounts of binary bits indeed would pose a high processing challenge to the system capability. However, with the current state of technology advancement, it is expected that processing this challenge would not raise so much of a critical threat to the current authentication systems.

3.5 Discussion on Computational Efficiency

Table 2 depicts the computational efficiency of the global discretization schemes in extracting bitstrings with 300 bits entropy based on PCA feature extraction and equal probable quantization. This experiment was carried out using a Pentium Dual-Core CPU E5200 (2.50 GHz) with 3.50 Gb RAM.

In **Table 2**, the time spent on generating a bitstring does not vary much among the schemes, where the preliminary feature extraction is a common component for all discretization schemes. The time difference in quantization and encoding between our approach and the conventional approaches is quite trivial, that is, not more than 8ms. As we can see, most of the computational time spent appears to be on the global feature selection stage, which is in fact strongly dependent on how complicated the discriminative measurements are taken. As the feature selection is only involved in the enrolment stage, it can be concluded that generating a bitstring during verification stage will not incur much degradation in computational efficiency than that of the conventional approach.

Table 2. Computational efficiency of the global discretization schemes.

	Time spent on PCA feature extraction (D = 100)	Time spent on global feature selection (D_{fs} = 50) (Training only)	Time spent on quantization & encoding	Total time elapsed from generating a bitstring with L = 300 bits (feature extraction + quantization + encoding)
EP+BRGC & EP+DBR	5.6ms	-	70.9ms	76.5ms
EP+LSSC	5.6ms	-	76.4ms	82ms
RL-based GFS+EP+LSSC	5.6ms	0.9630s	77.5ms	83.1ms
DR-based GFS+EP+LSSC	5.6ms	5.9573s	77.5ms	83.1ms

4. Conclusions

In a nutshell, we have proposed a global discretization approach to achieve an enhanced performance compared to that of the global schemes of the current state of art. This approach generally consists of two essential components: a global feature selection method and a Linearly Separable SubCode encoding technique. The way our approach works is based on the following rationale: The former element selects a pre-specified number of reliable discriminative dimensions for discretization while the latter preserves such discriminative performance by introducing an amount of redundant bits to satisfy the system-specified entropy requirement. Experimentally, we have shown that any of the following discriminative measures: reliability or detection rate is appropriate to be employed for global feature selection. Lastly, promising performance improvement is guaranteed when the discriminative measurements can be reliably depended upon in feature selection, thus justifying the feasibility of our approach.

Acknowledgements

This work was supported by the Korea Science and Engineering Foundation (KOSEF) through the Biometrics Engineering Research Center (BERC) at Yonsei University. (Grant Number: R112002105080020 (2011)); and by the MKE(The Ministry of Knowledge Economy), Korea, under IT/SW Creative research program supervised by the NIPA(National IT Industry Promotion Agency)" (NIPA-2010-(C1810-1002-0016)).

References

[1] P.N, Belhumeur, J.P. Kriegman and D.J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997. Article (CrossRef Link)

[2] Y. Chang, W. Zhang and T. Chen, “Biometric-based cryptographic key generation,” in *Proc. of IEEE International Conference on Multimedia and Expo (ICME 2004)*, 2004. Article (CrossRef Link)

[3] C. Chen, R. Veldhuis, T. Kevenaar and A. Akkermans, “Multi-bits biometric string generation based on the likelihood ratio,” in *Proc. of IEEE International Conference on Multimedia and Expo (ICME 2004)*, vol. 3, pp. 2203-2206, 2004. Article (CrossRef Link)

[4] C. Chen, R. Veldhuis, T. Kevenaar and A. Akkermans, “Biometric quantization through detection rate optimized bit allocation,” *EURASIP Journal on Advances in Signal Processing*, Article ID 784834, 16 pages, 2009. Article (CrossRef Link)

[5] Y. Dodis, R. Ostrovsky, L. Reyzin and A. Smith, “Fuzzy extractors: How to generate strong keys from biometrics and other noisy data,” *Eurocrypt 2004*, LNCS, vol. 3027, pp. 523-540, 2004. Article (CrossRef Link)

[6] F. Gray, “Pulse code communications,” *U.S. Patent* 2632058, 1953. Article (CrossRef Link)

[7] F. Han, J. Hu, L. He and Y. Wang, “Generation of reliable PINs from fingerprints,” in *Proc. of IEEE International Conference on Communications (ICC ’07)*, pp. 1191-1196, 2007. Article (CrossRef Link)

[8] F. Hao and C.W. Chan, “Private key generation from on-line handwritten signatures,” *Information Management & Computer Security*, vol. 10, no. 4, pp. 159-164, 2002. Article (CrossRef Link)

[9] A. Juels and M. Wattenberg, “A fuzzy commitment scheme,” in *Proc. of 6th ACM Conference in Computer and Communication Security (CCS'99)*, pp. 28-36, 1999. Article (CrossRef Link)

[10] T.A.M. Kevenaar, G.J. Schrijen, M. van der Veen, A.H.M. Akkermans and F. Zuo, “Face recognition with renewable and privacy preserving binary templates,” in *Proc. of 4th IEEE Workshop on Automatic Identification Advanced Technologies (AutoID '05)*, pp. 21-26, 2005. Article (CrossRef Link)

[11] M.-H. Lim and A.B.J. Teoh, "Linearly separable subcode: A novel output label with high separability for biometric discretization," in *Proc. of 5th IEEE Conference on Industrial Electronics and Applications (ICIEA'10)*, pp. 290 – 294, 2010. Article (CrossRef Link)
[12] J.-P. Linnartz and P. Tuyls, "New shielding functions to enhance privacy and prevent misuse of biometric templates," in *Proc. of 4th International Conference on Audio and Video Based Perso Authentication (AVBPA 2004)*, LNCS, vol. 2688, pp. 238-250, 2003. Article (CrossRef Link)
[13] A.M. Martinez and R. Benavente, "The AR Face Database," *CVC Technical Report* # 24, 1998. Article (CrossRef Link)
[14] F. Monrose, M.K. Reiter, Q. Li and S. Wetzel, "Cryptographic key generation from voice," in *Proc. of IEEE Symposium on Security and Privacy (S&P 2001)*, pp. 202-213, 2001. Article (CrossRef Link)
[15] F. Monrose, M.K. Reiter, Q. Li and S. Wetzel, "Using voice to generate cryptographic keys," *Odyssey 2001, The Speaker Verification Workshop*, 2001. Article (CrossRef Link)
[16] A.B.J. Teoh, K.-A. Toh and W.K. Yip, "2^N discretisation of biophasor in cancellable biometrics," in *Proc. of 2nd International Conference on Biometrics (ICB 2007)*, Lecture Notes in Computer Science, vol. 4642, pp. 435-444, 2007. Article (CrossRef Link)
[17] A.B.J. Teoh, D.C.L. Ngo and A. Goh, "Personalised cryptographic key generation based on FaceHashing," *Computers and Security*, vol. 23, no. 7, pp. 606-614, 2004. Article (CrossRef Link)
[18] A.B.J. Teoh, W.K. Yip and K.-A. Toh, "Cancellable biometrics and user-dependent multi-state discretization in BioHash," *Pattern Analysis & Applications*, vol. 13, no. 3, pp. 301-307, 2009. Article (CrossRef Link)
[19] M. Turk and A. Pentland, "Eigenfaces for Recognition", *Journal of Cognitive, Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991. Article (CrossRef Link)
[20] P. Tuyls, A.H.M. Akkermans, T.A.M. Kevenaar, G.-J. Schrijen, A.M. Bazen and N.J. Veldhuis, "Practical biometric authentication with template protection," in *Proc. of 5th International Conference on Audio- and Video-based Biometric Person Authentication*, LNCS, vol. 3546, pp. 436-446, 2005. Article (CrossRef Link)
[21] E. Verbitskiy, P. Tuyls, D. Denteneer and J.P. Linnartz, "Reliable biometric authentication with privacy protection," in *Proc. of 24th Benelux Symposium on Information Theory*, pp. 125-132, 2003. Article (CrossRef Link)
[22] W.K. Yip, A. Goh, D.C.L. Ngo and A.B.J. Teoh, "Generation of replaceable cryptographic keys from dynamic handwritten signatures," in *Proc. of 1st International Conference on Biometrics*, Lecture Notes in Computer Science, vol. 3832, pp. 509-515, 2006. Article (CrossRef Link)

Meng-Hui Lim obtained his BEng in 2006 from Multimedia University in Malaysia and MEng degree in 2009 from Dongseo University in South Korea. He is currently a Ph.D student in EE Department, College Engineering of Yonsei University in South Korea. His research interests include machine learning, pattern recognition, cryptography, and discretization.

Andrew Beng Jin Teoh obtained his BEng (Electronic) in 1999 and Ph.D degree in 2003 from National University of Malaysia. He is currently an assistant professor in EE Department, College Engineering of Yonsei University, South Korea. His research interest is in biometrics security and pattern recognition. He had published around 160 international journal and conference papers in his area.