Cancelable Iris Biometrics and Using Error Correcting Codes to Reduce Variability in Biometric Data

Sanjay Kanade*, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi
Department of Electronics and Physics,
Institut TELECOM: TELECOM & Management SudParis, Evry, France.
{Sanjay.Kanade,Dijana.Petrovska,Bernadette.Dorizzi}@it-sudparis.eu

Abstract

With the increasing use of biometrics, more and more concerns are being raised about the privacy of the personal biometric data. Conventional biometric systems store biometric templates in a database. This may lead to the possibility of tracking personal information stored in one database by getting access to another database through cross-database matching. Moreover, biometric data are permanently associated with the user. Hence if stolen, they are lost permanently and become unusable in that system and possibly in all other systems based on that biometrics. In order to overcome this non-revocability of biometrics, we propose a two factor scheme to generate cancelable iris templates using iris-biometric and password. We employ a user specific shuffling key to shuffle the iris codes. Additionally, we introduce a novel way to use Error Correcting Codes (ECC) to reduce the variabilities in biometric data. The shuffling scheme increases the impostor Hamming distance leaving genuine Hamming distance intact while the ECC reduce the Hamming distance for genuine comparisons by a larger amount than for the impostor comparisons. This results in better separation between genuine and impostor users which improves the verification performance. The shuffling key is protected by a password which makes the system truly revocable. The biometric data is stored in a protected form which protects the privacy. The proposed scheme reduces the Equal Error Rate (EER) of the system by more than 90% (e.g., from 1.70% to 0.057% on the NIST-ICE database).

1. Introduction

Identity verification is a need of today's networked society. Biometrics, being intrinsically associated with the user's identity, is employed for person verification/identification. It provides high degree of assurance about the individual's identity. In general, verification involves one-to-one comparison of the test data with the reference data of the same user which means that the user discloses his identity at the time of verification. In identification, the test data is compared with the stored data of all users and the system detects who the user is. Authentication is also a one-to-one comparison similar to verification and the two terms are used synonymously in this paper. Generally, the biometric systems extract specific features from the biometric data which are called biometric templates and store these templates in a central database for future comparisons. This paper describes an iris-biometric based system in which, the biometric feature is a 1,188-bit binary string extracted from an iris image, called iris code.

With more and more applications using biometrics, there is an increased possibility of tracking personal information from one application to another by cross-matching between biometric databases and this is considered as a compromise for one's privacy. Moreover, the biometric data are permanently associated with the user. If such data is compromised in a system, it is not possible to use it again in that system and possibly in all other systems based on the same biometrics. The biometric template cannot be replaced with another data and thus it is non-revocable.

Another popular technique employed for user authentication is to use a password. Password based systems employ cryptographic techniques for user authentication. They assume that the password is known only to the genuine user and if a person provides the correct password, he is successfully authenticated. But as is obvious, the password is not strongly associated with the user and can be stolen and/or shared easily. A genuine user can also repudiate: he can claim of not accessing the system even after accessing it. In order to increase the security of such systems, it is generally combined with a second factor: a token such as smart card. The authentication is carried out based on the possession of the token and the correct password. Though this two factor scheme increases security, the authenticators (password and

*The author was supported by the French "Agence Nationale de la Recherche (ANR)" project BIOTYFUL, (ANR-06-TCOM-018).

token) are not bound to the user identity so the problems associated with password based systems still exist.

An ideal authentication system should provide a strong link between the authenticator and the user along with the important properties such as revocability, template diversity, non-repudiation, and privacy protection. In order to achieve all these characteristics, we combine biometrics with a password. The scheme presented here is tested with iris biometrics, but in general, it can be adopted to any biometric, provided the biometric features are in the form of an ordered set and the nature of biometric variability is known. A shuffling key is obtained using a password which is used to shuffle the iris code. This shuffling scheme was proposed by Kanade et al. [10] in their work on biometrics based cryptographic key regeneration. The advantage of this shuffling scheme is that it increases the Hamming distances for impostor (inter-personal) comparisons but the Hamming distances for genuine (intra-personal) comparisons remain intact.

In this paper, we treat the intra-personal variations in iris codes as errors and use Error Correcting Codes (ECC) to reduce them. In contrary to all previous works using ECC [6, 2, 10], we actually correct the errors in test iris code to obtain a modified test iris code which is closer to the reference iris code in terms of Hamming distance. ECC operate in such a way that, they correct more errors in genuine cases than in impostor cases which helps in better user separation and improves the verification performance of the system.

The rest of this paper is organized as follows: some of the works about revocable biometrics and those using ECC with biometrics are discussed in Section 2. Section 3 gives a detailed explanation of the proposed scheme. This scheme is evaluated on publicly available iris databases using experimental protocols which are discussed in Section 4. Results and system analysis are given in Section 5 and finally, Section 6 sets out conclusions and perspectives.

2. Related Works

There are many works found in literature which propose schemes that can obtain revocable biometric templates. A detailed survey of this field can be found in [4]. We broadly divide these systems into two main categories: (1) cancelable biometrics and (2) biometric based cryptographic key (re)generation. The works in the first category obtain templates which are revocable and are compared with some distance metrics. Whereas the other category includes systems in which a stable bit-string, called crypto-biometric key, is obtained using biometrics. These systems focus on the stability of the crypto-biometric keys and the keys should be perfectly matched for successful authentication. Some of the works in both these categories are discussed below.

From the first category, we can cite Ratha et al. [16] who introduced the term cancelable biometrics for their system which transforms biometric signal/features using some irreversible transformations to obtain a revocable biometric template. In their recent work [15], they proposed three different transformations, namely, Cartesian, polar, and functional transformation, in order to obtain cancelable templates using fingerprints. But, they report that with all the transformations the performance of the biometric system degrades.

Savvides et al. [17] proposed cancelable biometric filters for face recognition where they use a random kernel (which can be obtained from a PIN) to encrypt the facial images. They proved the invariance of the recognition performance to the encryption scheme and thus there is no improvement in the performance of the underlying biometric system.

The Improved BioHashing scheme [13] proposed by Lumini and Nanni which is a modified version of the BioHashing scheme of Jin et al. [7], employs random number generation and Gram-Schmidt ortho-normalization to obtain revocable biometric templates. They report improvement in the biometric system performance which is an advantage over the Ratha et al. system [15] and Savvides et al. system [17].

There is one point worth noting about some of these systems, which is that they require a random number for the transformations. This random number requires to be stored on a token but they do not provide any details about how that token is secured. Hence, the random number can be easily accessed by someone who obtains the token. In this case, when the random number is stolen, the performance of the system in [13] degrades even below the underlying biometric system.

Boult et al. [1] proposed fingerprint based biotokens which provide revocable templates. They employ robust matching techniques in encoded domain. They report an average decrease of 30% in the Equal Error Rate (EER) of the system. But, they do not provide any details about system performance when the transformation parameters are stolen.

The other category, cryptographic key (re)generation, includes some theoretical studies [9, 8, 18], as well as works with experimental results [19, 6, 2, 10]. Most of these systems treat the biometric variabilities as errors and use Error Correcting Codes (ECC) to cope with them. A random key $k, k \in [0,1]^m$, is encoded by the ECC to obtain a codeword c as, $c = ECC(k), c \in [0,1]^n$. The biometric feature vector b is XOR*ed* with c to obtain an enrollment template τ as, $\tau = c \oplus b$. At the time of verification, another biometric feature vector b' is obtained and XOR*ed* with τ to get c' which is actually c containing errors between b and b'.

$$\begin{aligned} c' &= \tau \oplus b', \\ &= c \oplus b \oplus b', \\ &= c \oplus e. \end{aligned} \tag{1}$$

If the errors e between the biometric samples b and b' are less than the error correcting capability of the ECC, the decoding process results in the exact key k and if not, the output of decoding function is a random element belonging to $[0,1]^m$. This process is shown in Fig.1. The key k (or its modified form) is generally used as a cryptographic key. The templates in these systems can be revoked by changing the random key k.

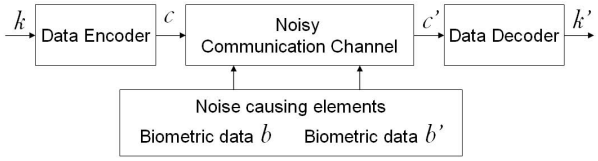


Figure 1. Block diagram showing the use of Error Correcting Codes for biometrics based cryptographic key (re)generation. This scheme is applicable to [6, 2, 10]. In the figure, k = random key, k' = regenerated key, c = encoded codeword, and c' = corrupted codeword.

A closer analysis of this scheme reveals an important aspect of these systems. The ECC act only as a classifier in these systems replacing the Hamming distance based classifier. The two iris codes act as noise causing elements in a communication channel. The ECC work on the variability in the two iris codes only to recover the locked random key. Thus, the ECC in these systems do not really correct errors in the iris codes but instead they eliminate the errors in codeword c caused by the biometric data. Moreover, this configuration can either eliminate all the errors caused due to biometrics being compared or it fails to eliminate any of them.

The novelty of this paper is that, we use the ECC in such a way that it can correct errors in the test biometric sample b' when compared to the reference sample b, and produce a new sample b'_1 such that, $HD(b,b') \geq HD(b,b'_1)$. By using the ECC properly, the proposed scheme can improve the system performance by a significant amount.

3. Reducing Intra-User Variability in Iris Codes and Cancelable Template Generation

As pointed out in the earlier section, the crypto-biometric systems found in literature such as [6, 2, 10] use ECC to eliminate the errors caused by the biometric data in the encoded random key. None of the systems really does error correction on the biometric data in order to reduce the variability among them. Here we introduce a novel way to use ECC by which we can successfully reduce the biometric variabilities. In particular, we take the iris as an example. Iris codes obtained from two images of an iris are generally not the same. They have two different types of variabilities which we refer to as errors: (a) *background errors*, which are random in nature, occurring due to camera noise, image capture effects, iris distortions etc., and (b) *burst errors* which generally occur due to eyelids, eye lashes, specular reflections, etc. Hao et al. [6] proposed a concatenated scheme using Reed-Solomon codes to correct burst errors and Hadamard codes to correct background errors. Kanade et al. [10] adopted the Hao et al. [6] scheme by increasing the error correction capability and using a shuffling scheme to improve the performance. But, the genuine as well as impostor errors are treated in a same way by these systems.

When we studied the causes of errors in iris codes, we found out that there is a fundamental difference between errors occurring in genuine and impostor comparisons. As stated earlier, the burst errors occur due to eye-lids, eye lashes, and specular reflections. The probability of occurrence of these errors is the same in genuine as well as impostor cases. The other type of errors, the random errors, occur due to camera noise, image capture effects, etc. But in impostor cases, the random errors are also due to the randomness of iris structures, i.e., these errors are due to the inter-personal variabilities. Hence, in general, there are more random errors in impostor comparisons than in genuine comparisons.

Using this hypothesis, we propose a scheme which will correct only the random errors up to a certain limit such that the Hamming distance between genuine comparisons will decrease by a greater amount than in impostor cases. Hadamard codes are used to correct the errors which are briefly introduced in the following subsection. Moreover, we use the iris code shuffling scheme proposed by Kanade et al. [10] which shuffles the iris codes by a shuffling key which improves the verification performance and security.

3.1. Hadamard Codes

Hadamard codes are obtained from Hadamard matrix generated by Sylvester method. Hadamard matrix is a square orthogonal matrix with elements '1' or '-1'. Hadamard code $HC(k)$ is constructed from the Hadamard matrix $H(k)$ as:

$$HC(k) = \begin{bmatrix} H(k) \\ -H(k) \end{bmatrix}. \quad (2)$$

The codewords are obtained by replacing -1 with 0 in $HC(k)$. The Hadamard code of size $n = 2^k$ has $2n$ codewords each of which is n bit long. The code has a minimum distance of 2^{k-1} and hence can correct up to $2^{k-2}-1$ errors (i.e., $\approx 25\%$).

During encoding, an input value i is encoded into a codeword w. Here i is $(k+1)$ bit and w is $n = 2^k$ bit. The matrix $HC(k)$ has $2n$ rows which are codewords. The input value i is considered as a row index and the corresponding row is taken as an output codeword. Thus an input block of $(k+1)$ bits is converted into an output block of 2^k bits.

At the time of decoding, every 0 in the received codeword w is replaced by -1 to obtain w'. Then the product,

$$w'HC^T(k) = (a_0, a_1, ..., a_r, ..., a_{2n-1}), \qquad (3)$$

is calculated. The position r, where a_r is maximum, is the decoded value. If at most $2^{k-2} - 1$ errors have occurred, the decoded value is equal to the input value, i.e., $r = i$.

3.2. Correcting Errors in Iris Code

In this section, we propose an algorithm to correct errors in iris codes. The scheme involves two distinct phases namely user enrollment phase and user verification phase. In the enrollment phase (Fig. 2), a revocable template is generated from the user provided data i.e., an iris image and a password. In the verification phase (Fig. 4), another iris image and the password are provided by the user which are used to verify the user.

We need to correct errors in the *test iris code* $\mathbf{Y} = \{y_1, y_2, ..., y_p\}$ with respect to the *reference iris code* $\mathbf{X} = \{x_1, x_2, ..., x_p\}$, where x_i and y_i are $n = 2^{m-1}$ bit binary strings. A $p \times m$ bit random bit-string $\mathbf{K}$, called a *random key*, is divided into p blocks of m bits each such that, $\mathbf{K} = \{k_1, k_2, ..., k_p\}$. Each block of $\mathbf{K}$ is encoded with a Hadamard code of size $(m - 1)$. The output of the Hadamard encoding is a set of encoded codewords which is denoted as *pseudo code* $\mathbf{S} = \{s_1, s_2, ..., s_p\}$. Each of the s_i is $n = 2^{m-1}$ bits. This *pseudo code* is XOR*ed* with the reference iris code $\mathbf{X}$ to form a *locked iris code* $\mathbf{Z} = \{z_1, z_2, ..., z_p\}$.

The iris code shuffling scheme proposed by Kanade et al. [10], is also used in the proposed system. The *reference iris code* $\mathbf{X}$ is shuffled with a randomly generated user specific shuffling key to obtain a *shuffled iris code*, $\mathbf{X}_{shuf}$.

The *shuffled iris code* $\mathbf{X}_{shuf}$, *locked iris code* $\mathbf{Z}$, shuffling key, and the Hash values of all $k_i's$ together form a user template. The template is protected by a password using standard security mechanisms. This is called the user enrollment phase as shown in Fig. 2.

The most distinct feature of the proposed scheme compared to the existing schemes [6, 2, 10] is in the decoding part. The decoding and error correction are carried out block-wise by processing one block at a time. The flowchart of this process is shown in Fig. 3. The i^{th} block of $\mathbf{Y}$, y_i, is XOR*ed* with the i^{th} block of the locked code $\mathbf{Z}$, z_i, producing s'_i, which is the corresponding block of pseudo code $\mathbf{S}$, s_i, contaminated with the errors between the two iris code blocks. s'_i is decoded by the Hadamard code to obtain k'_i which is compared with the original random key block k_i. If it is found that $k'_i = k_i$, k'_i is re-encoded using Hadamard code to obtain s''_i. Since $k'_i = k_i$, $s''_i = s_i$. This s''_i is XOR*ed* with z_i to obtain y'_i. Since, $s''_i = s_i$, $s''_i \oplus z_i = s''_i \oplus s_i \oplus x_i = x_i$.

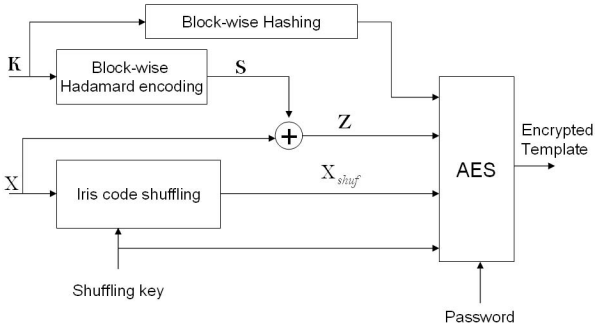


Figure 2. Block diagram showing the enrollment process for the proposed scheme. Here, $\mathbf{K}$ is a random key, $\mathbf{X}$ = reference iris code, $\mathbf{S}$ = pseudo code, $\mathbf{Z}$ = locked iris code, $\mathbf{X}_{shuf}$ = shuffled iris code, and AES = Advanced Encryption Standard.

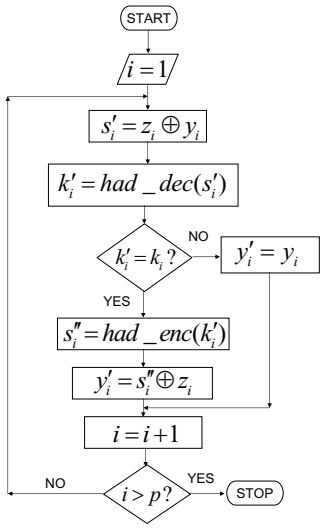


Figure 3. Algorithm for applying ECC to reduce variability in iris codes where z_i = locked iris code block; y_i = test iris code block; y'_i = modified test iris code block; k_i = random key block; had_enc = Hadamard encoding; had_dec = Hadamard decoding.

If the $k'_i \neq k_i$, then k'_i is not re-encoded and without further processing, $y'_i = y_i$. This process is carried out for all $i, (i = 1, 2, ..., p)$ resulting in $\mathbf{Y}' = \{y'_1, y'_2, ..., y'_p\}$ which is the modified test iris code. As can be seen, this code is the test iris code with some blocks replaced by the regenerated reference iris code blocks. Consequently, when this modified test iris code is compared with the reference iris code, the Hamming distance is decreased. In case of genuine users, the random errors are generally less than 25% (which is the Hadamard code error correction capability) whereas for impostors, the errors are generally more than 25%. Hence, this error correction scheme helps reduce the genuine Hamming distance by a significant amount which ultimately helps in a better user separation.

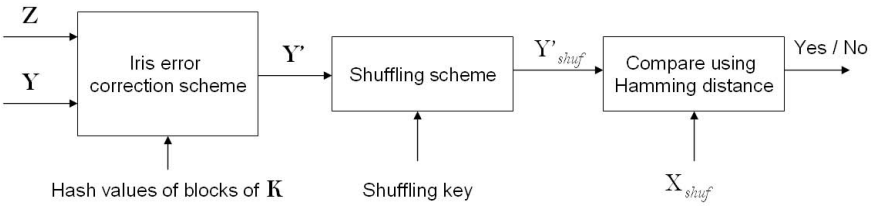


Figure 4. Block diagram showing the user verification process for the proposed scheme. The decryption of the template data is not shown. Here, **K** is a random key, **Y** = test iris code, **Z** = locked iris code, **Y'** = modified (error corrected) test iris code, $\mathbf{X}_{shuf}$ = shuffled reference iris code, and $\mathbf{Y'}_{shuf}$ = shuffled modified test iris code.

One of the main goals of this system is to protect user privacy. The biometric data can be recovered from the locked iris code if an impostor knows the key **K**. In order to provide security to the biometric data, the key **K** is not stored in the system. Instead, a one-way hash function is used to hash each block of the key and the hash values are stored as part of the template.

Moreover, the iris code comparison is not carried out in a usual way. Conventional iris code matching algorithm uses iris noise masks which represent the locations of possible errors in the iris codes. This helps to suppress the possible error bits from the iris code comparison. This results in symmetric error correction assuming that both, the reference and test iris codes, may contain errors. The Hamming distance considering masks (HD_{mask}) between two iris codes is calculated using the following formula:

$$HD_{mask} = \frac{\|(Code_1 \oplus Code_2) \cap Mask_1 \cap Mask_2\|}{\|Mask_1 \cap Mask_2\|}, \tag{4}$$

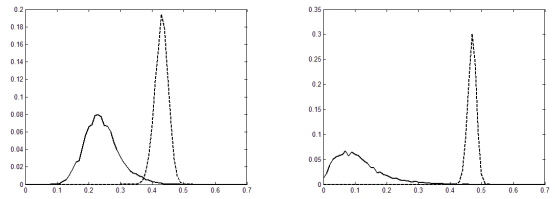
where, $Code_1$, $Code_2$ represent the reference and test iris codes respectively and $Mask_1$, $Mask_2$ are their respective noise masks. It is clear from equation (4) that both the iris masks are needed to be logically AND*ed* and thus the reference mask must be stored in the system. This mask can leak vital information about the iris code making the system weaker. Also using such masks will enable an impostor to select only a certain number of bits from the iris code and he can get more easily accepted by the system. Hence from a security point of view, we prefer not to use the masks. In fact, using the noise masks helps to improve the performance of the system. Thus by opting not to use the masks, we are making the verification process more difficult. Similar approach was followed by Hao et al. [6].

We have used the iris code shuffling scheme proposed by Kanade et al. [10] which further improves the user separation capability of the system. In this shuffling scheme, an iris code is shuffled with a user specific random shuffling key. The iris code is divided into blocks and these blocks are aligned with the shuffling key bits. If a bit in the shuffling key is 1, the corresponding block is taken into part 1 and if the bit is zero, the corresponding block is taken into part 2. The concatenation of the two parts gives a shuffled iris code. The Hamming distance between shuffled reference iris code $\mathbf{X}_{shuf}$ and shuffled modified test iris code $\mathbf{Y'}_{shuf}$ is considered in order to make the decision. But, there is a possibility of miss-alignment of the iris codes which is generally a result of rotation. In order to cope with that, the normalized iris image is translated horizontally in both the directions 20 times (10 times in each direction). All of these translated images are processed to generate 21 shuffled modified test iris codes and the Hamming distances between each of them and the shuffled reference iris code are calculated. The minimum of these Hamming distances is considered for making the final decision. The user verification phase is shown in Fig. 4.

As shown in [10], the advantage of this shuffling scheme is that it increases the Hamming distance for impostor comparisons but the Hamming distance for genuine comparisons remains intact. Thus, the ECC scheme decreases the Hamming distance for genuine comparisons and the shuffling scheme increases the Hamming distance for impostor comparisons which results in improvement in the verification performance. This effect can be seen from the Hamming distance distribution curves in Fig. 5. Note that the systems described in Hao et al. [6] and Bringer et al. [2] do not attempt to change the Hamming distance distributions and hence they cannot improve the performance of the underlying biometric system. The Kanade et al. [10] system improves the distribution but the improvement is achieved only due to increase in impostor Hamming distances. The scheme proposed in this paper works on both, the genuine as well as the impostor Hamming distances and hence the improvement is significantly higher than that in the Kanade et al. [10] scheme.

4. Databases and Experimental Protocols

We developed our system on the publicly available Casia-BioSecure (CBS) database [12] (OKI device subset). In order to prove the portability of our system, we tested the system on the Iris Challenge Evaluation (ICE) database [14] with the parameters obtained from the CBS database tests. The benchmarking protocols given with the



(a) (b)

Figure 5. Effect of Error Correction and shuffling on genuine (thick line) and impostor (dotted line) Hamming distance distributions for the ICE-Exp-1; (a) the baseline biometric system, (b) the proposed system using ECC and shuffling.

respective databases were followed.

The CBS database has two parts: BioSecureV1 and CasiaV2. According to the CBS database protocol described in [12], 6,000 tests for genuine as well as impostor comparisons were carried out for each of the two parts. This protocol results in tests between images captured in different sessions, illumination conditions, and between images of eyes with and without spectacles.

The ICE database consists of 2,953 images from 244 different eyes. These images are divided into two parts: right eye images and left eye images. Separate experiments were carried out for each of the two parts: Exp-1 – with right eyes, and Exp-2 – with left eyes. All possible comparisons between iris images were carried out for the two experiments: for Exp-1, 12,214 genuine, and 1,002,386 impostor comparisons, and for Exp-2, 14,653 genuine, and 1,151,975 impostor comparisons.

The Open Source Iris recognition system, OSIRIS [12, 11] is used to extract iris codes from the iris images from the databases. Iris code is a binary string of phase information extracted from the Gabor filter decomposed iris images. The OSIRIS has two main parameters: filters and analysis points. We select 6 filters and 198 analysis points which yield 1,188 bit iris codes. In order to match the iris code structure, we set the number of bits in the shuffling key to be 198 and number of bits in each iris code block to be 6 for the shuffling algorithm. Thus the shuffling key length is set at 198 bits which will be protected with a password of eight characters.

5. Results and System Analysis

When reporting the performance of any biometric system, it is important to consider the database, experimental protocol, etc. The performance can vary according to the nature and size of the database and hence we tested our system on two different databases. Moreover, when performance of a cancelable biometric system is to be reported, it is worthwhile mentioning the change in performance of the system with reference to the baseline biometric system.

One of the most common ways to report the system performance is to report the Equal Error Rate (EER). We found out that after applying the ECC and shuffling scheme, the EER decreases by more than 90%, e.g., on ICE-Exp-1, the EER decreases from 1.70% for the baseline biometric system to 0.057% for the proposed system. The detailed results are reported in Table 1. The Detection Error Trade-off (DET) curves are plotted in Fig. 6. These curves clearly show the improvement achieved by the proposed system over the baseline system.

Table 1. Verification results in terms of EER in %, BIO – baseline biometric system; B1 – CBS-BioSecureV1, C2 – CBS-CasiaV2

Test	B1	C2	ICE-Exp-1	ICE-Exp-2
BIO	2.73	3.05	1.70	1.78
Proposed	0.15	0.057	0.057	0.125

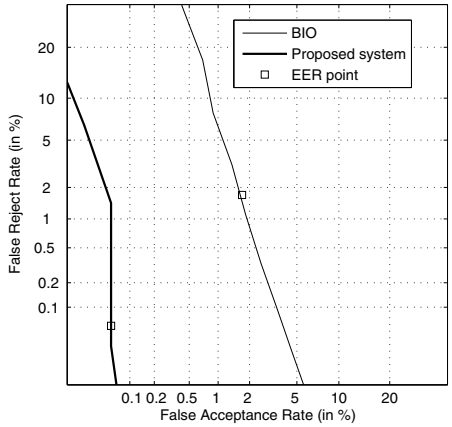


Figure 6. Comparison of the performance of the proposed system and the baseline biometric system using Detection Error Trade-off (DET) curves on the ICE-Exp-1.

The thresholds at EER for all the four experiments are nearly constant which indicates that the system is portable on various databases.

Another popular way to report the biometric system performance is to report the values of False Rejection Rate (FRR) at fixed values of False Acceptance Rates (FAR). The results of the NIST-ICE evaluations were reported in this manner. In order to facilitate the comparison, the results on ICE database are reported in Table 2. The performance of the proposed system is better than the best reported result (0.1-0.2% FRR at 0.1% FAR for the SAGEM algorithm) in ICE [14].

The proposed system provides protection to the templates and here we estimate the security theoretically. The system needs two inputs from the user: an iris image and

Table 2. Verification results in terms of FRR at specified values of FAR for the ICE database, (all values are in %); (a) baseline biometric system, (b) proposed system.

Experiment	FRR		
	FAR=0	**FAR=0.001**	**FAR=0.1**
ICE-Exp-1 (a)	15.82	8.07	3.20
ICE-Exp-1 (b)	0.17	0.07	0.057
ICE-Exp-2 (a)	29.47	11.40	4.29
ICE-Exp-2 (b)	0.348	0.16	0.125

a password. We propose to use an 8-character randomly generated password which can have 52-bit entropy [3]. The iris code itself contains some correlations and following the procedure given by Daugman [5], we estimate the degrees of freedom in the 1,188-bit iris codes to be 561. The Hadamard code acts on 32-bit blocks of the iris code thus, on average, each block has $f \approx 15$ degrees of freedom. The Hadamard code can correct at most 7 error bits occurring in 32-bit block resulting in 22% error correction. Considering this error tolerance (i.e., $g = 0.22 \times f \approx 3$) and following the procedure given in Hao et al. [6], an impostor will need,

$$BF \approx \frac{2^f}{\binom{f}{g}} \approx \frac{2^{15}}{\binom{15}{3}} \approx 72, \tag{5}$$

brute force calculations to successfully guess the random key block. But, since the random key block is only 6 bits long, the maximum number of calculations required per block will be 64. To guess the random key completely, $64 \times 37 \approx 2^{11.2}$ calculations will be required. Thus the overall security provided by the system is $52 + 11 = 63$ bits. This security can be significantly increased by limiting the maximum number of login attempts.

The system proposed in this paper also provides template diversity, i.e., it is able to generate completely different templates from one iris. This is achieved by changing the shuffling key. The two templates generated from one iris using two different shuffling keys do not match with each other. In order to prove our point, we carried out two experiments on the ICE-Exp-1 data: (i) each iris code is shuffled with two different shuffling keys and the resultant shuffled iris codes are compared with each other; and (ii) different iris codes from different images of the same eye (i.e., genuine iris code comparisons) are shuffled with different shuffling keys and compared with each other. The plots of the Hamming distances resulting from these two experiments are shown in Fig. 7. For comparison purposes, the impostor hamming distances (which are also shown in Fig. 5(b)) are also shown in the same figure. It is clear from the figure that the system can generate diverse templates for each user which are comparable to random iris templates.

The second experiment for template diversity also proves an important point that is the system restricts the access to

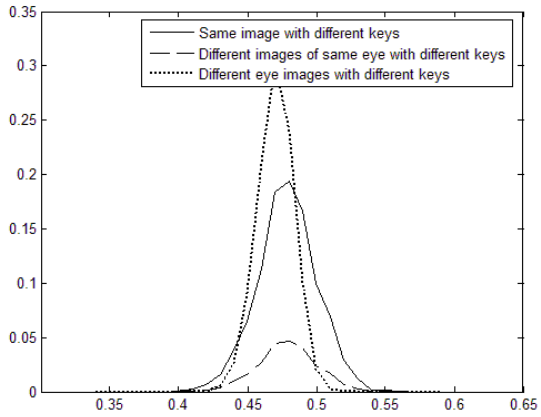


Figure 7. Hamming distance distributions (for ICE-Exp-1) proving that the system provides template diversity.

an impostor who obtains a genuine iris sample but does not have the right password. When an impostor provides an incorrect password, he cannot obtain the correct shuffling key. If he uses a wrong shuffling key with the stolen iris code, he is treated as a random impostor user and is most probably rejected.

There is another security issue in the proposed system that the shuffling key can also be stolen (by stealing the password). In this scenario, the Hamming distance between the two iris codes being compared does not increase but it remains unchanged and thus the performance of the system in this scenario does not decrease below the baseline biometric system. When we tested the system for such a case on CBS-BioSecureV1 database, we found out the EER to be 2.70% which is nearly the same as that of the baseline biometric system.

The use of the password is a required step in order to make the system truly revocable. By true revocability we mean that, the system must require an impostor to break the security of the template every time it is revoked. In the proposed system, the shuffling key transforms the iris code in an irreversible way such that the original iris code cannot be obtained from the shuffled iris code without the actual shuffling key. But, the shuffling key is a long bit-string which is generally not possible to remember and hence needs to be stored somewhere. If it is not protected by a user provided secrete, it is accessible by a common access mechanism. In case of compromise detection (e.g., an authentic iris code is obtained by an attacker), a new template will be issued by the system by changing the random key $\mathbf{K}$ and the shuffling key. If password is not being used, the data being provided by the user is only the iris code. Hence, the attacker, who already has the iris code, can easily access the system which means the new template cannot avoid such attacks. Instead,

if the system also employs a password, the attacker has to crack the encryption system every time to get the shuffling key. Thus, password is a required aspect of such systems to make them truly revocable.

6. Conclusions and Perspectives

We proposed a novel approach to use Error Correcting Codes (ECC) for reducing variabilities (which are treated as errors) in iris codes. After carefully studying the causes of errors in iris codes, we designed an ECC scheme that can correct more errors in genuine iris codes than in impostors. Moreover, we use an iris code shuffling scheme which shuffles the iris code with a user specific randomly generated shuffling key. The shuffling scheme increases the Hamming distance for impostor comparisons whereas for the genuine comparisons, the Hamming distance remains the same. The combination of the two techniques enables the system to distinguish genuine users from impostors with high accuracy. The shuffling key is protected by a password which makes the system truly revocable. The templates do not store personal biometric data in its usual form thereby protecting the user privacy. The use of password and shuffling key does not allow access to the system even if a stolen iris image is provided. The system was first developed on the CBS database and to prove its portability, it was tested on the ICE database with the same tuning parameters. The proposed system improves the verification performance of the underlying biometric system by reducing the EER by more than 90%, e.g., for ICE-Exp-1, the EER reduces from 1.70% to 0.057%. This system performs better than the best performing system in ICE.

The idea of using ECC to *reduce* the biometric variability, in general, can be used with any biometric having ordered binary feature set provided the ECC is tuned according to the nature of errors in that biometric. This system is also suited for high security applications since it has very low FRR (e.g., 0.17% for ICE-Exp-1) at 0% FAR.

References

[1] T. E. Boult, W. J. Scheirer, and R. Woodworth. Revocable fingerprint biotokens: Accuracy and security analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.

[2] J. Bringer, H. Chabanne, G. Cohen, B. Kindarji, and G. Zmor. Optimal iris fuzzy sketches. In *IEEE Conference on Biometrics: Theory, Applications and Systems*, 2007.

[3] W. E. Burr, D. F. Dodson, and W. T. Polk. Electronic authentication guideline: Recommendations of the National Institute of Standards and Technology, April 2006.

[4] A. Cavoukian and A. Stoianov. Biometric encryption: A positive-sum technology that achieves strong authentication, security and privacy. White paper, Information and privacy commissioner of Ontario, March 2007.

[5] J. Daugman. The importance of being random: Statistical principles of iris recognition. *Pattern Recognition*, 36(2):279–291, February 2003.

[6] F. Hao, R. Anderson, and J. Daugman. Combining crypto with biometrics effectively. *IEEE Transactions on Computers*, 55(9):1081–1088, 2006.

[7] A. T. B. Jin, D. Ngo, C. Ling, and A. Goh. Biohashing: two factor authentication featuring fingerprint data and tokenised random number. *Pattern Recognition*, 37(11):2245–2255, November 2004.

[8] A. Juels and M. Sudan. A fuzzy vault scheme. In A. Lapidoth and E. Teletar, editors, *Proc. IEEE Int. Symp. Information Theory*, page 408. IEEE Press, 2002.

[9] A. Juels and M. Wattenberg. A fuzzy commitment scheme. In *Proceedings of the Sixth ACM Conference on Computer and communication Security (CCCS)*, pages 28–36, 1999.

[10] S. Kanade, D. Camara, E. Krichen, D. Petrovska-Delacrétaz, and B. Dorizzi. Three factor scheme for biometric-based cryptographic key regeneration using iris. In *The 6th Biometrics Symposium 2008 (BSYM2008)*, September 2008.

[11] E. Krichen. *Reconnaissance des personnes par l'iris en mode dégradé*. PhD thesis, Institut National des Télécommunications, 2007.

[12] E. Krichen, B. Dorizzi, Z. Sun, S. Garcia-Salicetti, and T. Tan. Iris Recognition. In D. Petrovska-Delacrétaz, G. Chollet, and B. Dorizzi, editors, *Guide to Biometric Reference Systems and Performance Evaluation*, pages 25–50. Springer-Verlag, 2009.

[13] A. Lumini and L. Nanni. An improved biohashing for human authentication. *Pattern Recognition*, 40(3):1057–1065, March 2007.

[14] National Institute of Science and Technology (NIST). Iris Challenge Evaluation, 2005. http://iris.nist.gov/ice.

[15] N. K. Ratha, S. Chikkerur, J. H. Connell, and R. M. Bolle. Generating cancelable fingerprint templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):561–572, April 2007.

[16] N. K. Ratha, J. H. Connell, and R. M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal*, 40(3):614–634, 2001.

[17] M. Savvides, B. V. Kumar, and P. Khosla. Cancelable biometric filters for face recognition. In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR04)*, volume 3, pages 922–925, August 2004.

[18] C. Soutar, D. Roberge, A. Stoianov, R. Gilroy, and B. V. Kumar. *Biometric Encryption*, chapter 22. McGraw-Hill, 1999.

[19] U. Uludag and A. Jain. Securing fingerprint template: Fuzzy vault with helper data. In *Proc. of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, pages 163–170, June 2006.