Effective Combination of Iris-based Cancelable Biometrics and Biometric Cryptosystems

Osama Ouda[1]
Department of Computer Science,
College of Computer and
Information Sciences, Jouf University,
Sakaka, Al-Jawf, Saudi Arabia

Norimichi Tsumura[2]
Department of Information and
Computer Sciences, Chiba University,
1-33, Yayoi-cho, Inage-ku,
Chiba, 263, JAPAN

Toshiya Nakaguchi[3]
Center for Frontier Medical Engineering,
Chiba University,
1-33, Yayoi-cho, Inage-ku,
Chiba, 263, JAPAN

Abstract—**The fuzzy commitment scheme (FCS) is one of the most effective biometric cryptosystems (BCs) that provide secure management of cryptographic keys using biometric templates. In this scheme, error correcting codes (ECCs) are firstly employed to encode a cryptographic key into a codeword which is then secured via linking (committing) it with a biometric template of the same length. Unfortunately, the key length is constrained by the size of the adopted biometric template as well as the employed ECC(s). In this paper, we propose a secure iris template protection scheme that combines cancelable biometrics with the FCS in order to secure long cryptographic keys without sacrificing the recognition accuracy. First, we utilize cancelable biometrics to derive revocable templates of large sizes from the most reliable bits in iris codes. Then, the FCS is applied to the obtained cancelable iris templates to secure cryptographic keys of the desired length. The revocability of cryptographic keys as well as true iris templates is guaranteed due to the hybridization of both techniques. Experimental results show that the proposed hybrid system can achieve high recognition accuracy regardless of the key size.**

Keywords*—*Biometric template protection; cancelable biometrics; biometric cryptosystems; BioEncoding; fuzzy commitment

I. Introduction

In the last few years, the marriage between biometrics and cryptography has been proven to be an effective approach to address several issues inherent to both technologies [1]. In spite of the usability advantages exhibited by biometrics, several security and privacy concerns have been raised about employing biometric characteristics in identity verification/identification. First, biometric traits are limited and therefore revoking and replacing a compromised biometric trait is not as easy as canceling and replacing a compromised password or token in traditional authentication/identification systems. Second, biometric traits are permanently associated with individuals and hence some of their private information could be revealed to adversaries if these traits are disclosed [2].

On the other hand, achieving secure management of cryptographic keys is one of the most difficult problems in cryptography. Whereas short keys are not secure, it is difficult to remember and manage several long cryptographic keys. Although user-specific tokens or cards can be employed to store long cryptographic keys, keeping such user-specific tokens secure is not assured.

Fortunately, it turned out that integrating some cryptographic concepts into biometrics systems provides practical solutions to the above-mentioned issues. This integration can take one of two main forms; namely, cancelable biometrics (CB) [3] and biometric cryptosystems (BCs) [4]. CB schemes, such as BioHashing [5], BioEncoding [6] and distorting transforms [7], aim to derive several protected templates from the original (unprotected) biometric signal employing one-way transforms. The derived templates should be revocable, renewable, noninvertible and preserve the discriminability of original templates. Basically, the CB construct is inspired from the concept of one-way hash functions in classical cryptography. However, unlike cryptographic hash functions, CB can derive similar protected templates from original biometric signals that belong to the same user. In fact, due to the intra-user variations, one should not expect that biometric samples acquired from the same user to be identical. That is why most of the existing CB techniques cannot satisfy all the requirements of the CB construct simultaneously without integrating other user-specific authentication factors in the transformation process. Moreover, although CB systems can exhibit a practical solution to the problems of template protection and privacy invasion, they are still vulnerable to some attacks that are inherent to conventional biometric systems. For example, current CB systems are neither resilient to Trojan horse attacks nor to overriding Yes/No response attack.

On the other hand, BCs, such as fuzzy extractors [8], the fuzzy vault scheme (FVS) [9] and the fuzzy commitment scheme (FCS) [10], bind/extract user-specific keys to/from biometric templates such that the key is released only if a genuine biometric sample is presented at the time of verification. In fact, this construct can be employed to protect biometric templates as well as to provide a practical approach to manage cryptographic keys. Unfortunately, although different cryptographic keys can be extracted from or linked to the same biometric template in a way that allows users to log on many systems using only their irises or fingerprints, BCs are not designed to be revocable with respect to biometric templates [11]. In other words, in BCs, it is possible to replace a compromised key and bind another key to the same biometric; however, if the biometric itself is compromised, it would not be possible to revoke it.

In order to benefit from the advantages of both approaches, several attempts of combining both CB and BCs have been proposed recently [12]-[22]. Most of these attempts employ the FCS as the BC of choice due to its simplicity and efficiency. The FCS utilizes single or concatenated ECCs to correct errors existing in different biometric samples acquired from the same user. Hence, the recognition accuracy of such hybrid systems relies primarily on the correction capability of the employed ECC(s) as well as the error rates of the adopted biometric characteristic. Moreover, the length of the key to be linked with a biometric template is constrained by both the size of that template and the used ECC(s). Therefore, to develop an effective hybrid template protection system using the FCS, all of the above issues need to be addressed.

In this paper, we propose a novel hybrid template protection system that combines both CB and BCs effectively to protect iris templates as well as cryptographic keys at the same time. A cancelable iris template (BioCode), of any desired length, is firstly generated from the most reliable bits in a true iris template using a new variant of our previously proposed cancelable transformation scheme [6]. Then, a cryptographic key is linked to the derived cancelable template employing the FCS. Iris is one of the most accurate and reliable biometric characteristics that has been successfully implemented in many real world applications with very low false rejection and acceptance rates [23]. Moreover, thanks to the proposed BioEncoding-based cancelable transformation method, the suggested hybrid system exhibits the following advantages over other existing hybrid template protection techniques: (1) no user-specific data need to be used with the proposed hybrid system (i.e., the proposed system is tokenless), (2) no restrictions are imposed on the size of the key to be secured using the proposed system, (3) both keys and iris templates could be revoked and replaced easily in case of compromise, and (4) a perfect recognition accuracy (0% ERR) can be achieved regardless of the key size.

The remainder of this paper is organized as follows. A review of the related works is presented in Section II. The FCS is reviewed in Section III. In Section IV, base BioEncoding is revisited and the proposed variant of BioEncoding is described. The proposed hybrid template protection scheme is presented in Section V. Experimental results and security analysis are presented in Sections VI and VII, respectively. Finally, conclusions are drawn in Section VIII.

II. Related Work

Several hybrid template protection schemes have been proposed in the last few years. Most of these systems utilize either the FVS or the FCS in the key binding step. On the other hand, the main difference between the cancelable transforms employed in these techniques lies in their commitment to the invertibility property.

Liu et al. [12] suggested to combine random projection based cancelable biometrics and the fuzzy vault technique to improve the security of palm-prints recognition. Sandhya and Prasad [13] utilized the fuzzy commitment scheme to protect cancelable templates constructed using Delaunay triangles from fingerprint minutiae. Kanade et al. [14]-[17] proposed several variants of a hybrid template protection scheme that depends on the FCS for binding cancelable iris and/or face templates to cryptographic keys. For obtaining cancelable templates, they suggested to shuffle bits in the original biometric templates using user-specific shuffling keys. The problem with these techniques lies in their utilization of invertible cancelable transforms rather than non-invertible ones. Thus, once these transforms are disclosed to adversaries, the original templates would be obtained easily.

Wang and Plataniotis [18] proposed a two-stage hybrid scheme for face biometrics. At the first sage, cancelable face templates are generated by quantizing the distance vectors between the extracted Principal Component Analysis (PCA) feature vector and pairs of user-dependent random vectors. Then, the FVS is utilized to bind a randomly generated key with the obtained cancelable face template. The same approach is followed by Nandakumar et al. [19] for protecting fingerprint templates. At first, a user-specific password is utilized to derive a random transformation function that is applied to the fingerprint template. Then, the transformed template is secured using the FV framework. Teoh et al. [20] proposed a two-step technique to derive personalized cryptographic keys from the face biometric. In the first step, a cancelable template (FaceHash) is generated from original face template using BioHashing (i.e., random projection followed by simple thresholding). Then, in the second step, a cryptographic key is derived from the generated FaceHash via Shamir's secret-sharing approach. A similar technique is presented by Song et al. [21] but for fingerprint templates. A cancelable fingerprint template, referred to as FingerHash, is firstly generated using BioHashing and then linked via the FCS to a cryptographic key that is encoded using a Reed-Solomon code. More recently, Feng et al [22] proposed a three-stage hybrid algorithm for face templates. Cancelable templates are generated at the first stage using random projection. At the second stage, a discriminability-preserving (DP) transform is applied to cancelable templates in order to enhance the discriminability of the original feature templates as well as converting the resulting cancelable template into a binary string. Finally, the binarized cancelable template is linked to a randomly generated binary string (the key) using the FCS. Although all the above-mentioned techniques employ non-invertible cancelable transforms, they utilize user-specific random numbers in the cancelable transformation process. That is, they suffer from the same problems associated with traditional-based authentication systems. In other words, if these random keys are compromised, the False Acceptance Rate (FAR) would increase significantly [24]. Moreover, although the employed non-invertible transforms are much harder to be reversed compared to invertible transforms, it would be much simpler for a skilled attacker to invert the transform and disclose the original features if he/she could gain access to these user-specific factors.

Unlike other CB methods, BioEncoding satisfies all the requirements of the CB construct yet does not require any user-specific passwords/keys to be employed in the cancelable transformation process. On the other hand, unlike the FCS, the FVS suffers from a number of security vulnerabilities [25] as well as some implementation difficulties. Therefore, we believe that combining both BioEncoding and the FCS into a hybrid template protection system would address most of the issues associated to other existing hybrid techniques. However, as mentioned earlier, some issues inherent to the FCS need to be dealt with. Addressing these issues is the main goal of the new BioEncoding-based cancelable transformation method proposed in this work.

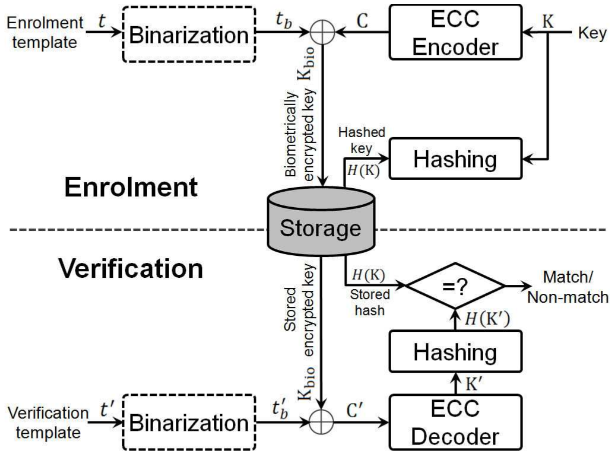


Fig. 1. Illustration of the Fuzzy Commitment Scheme.

III. Fuzzy Commitment Scheme

The FCS is one of the most attractive BCs that have been proposed so far. Due to its simplicity and efficiency, it has been applied to several biometric traits by many scientific groups [26]-[33]. As depicted in Fig. 1, the FCS requires a biometric template, t, to be represented as a binary string. Hence, a non-binary templates need to be binarized via an optional binarization module into a binary template, t_b, before the FCS can be applied. For iris biometric, since iris features are commonly represented as an ordered binary string, known as an iris code, the binarization step is not required (and hence t_b would be identical to t in Fig. 1). The FCS works as follows. On enrollment, a random binary key K is generated and encoded using an appropriate ECC(s) into a codeword C of length $n = \|t_b\|$. Both the binary template and the encoded key are then XORed to produce a biometrically encrypted key K_{bio}, also called a biometric key, as follows:

$$K_{bio} = C \oplus t_b \tag{1}$$

Furthermore, the hash value of the random key $\mathrm{H}(K)$ is computed and stored with the biometric key in a central storage or a user-specific token. At the time of verification, a binary template $t_b^{'}$ is extracted from a live biometric sample captured from the person being verified and XORed with the stored biometric key to obtain a possibly corrupted codeword C':

$$C' = t'_b \oplus K_{bio} \tag{2}$$

The obtained codeword is decoded using the ECC(s) employed on enrollment to get the verification key K'. Finally, the hash value of the recovered key, $\mathrm{H}(K')$ is computed using the same hashing function employed on enrollment and compared to the stored hash value, $\mathrm{H}(K)$. Only if the two hash values are identical, the key is released; otherwise, the authentication process fails.

IV. Generating Cancelable Templates

To bind long cryptographic keys with biometric templates, the size of such templates should be as large as possible. If (cancelable) biometric templates of any size could be generated, keys of unconstrained lengths would be employed in the key-binding phase. Unfortunately, cancelable biometrics techniques, such as BioHashing and BioEncoding, derive compact revocable templates from original biometrics templates. In BioHashing, if the generated cancelable template is not shorter than the original one, the transform could be inverted easily [34]. Also, the protected BioCodes generated using BioEncoding, as proposed in [6], are shorter than their corresponding original templates.

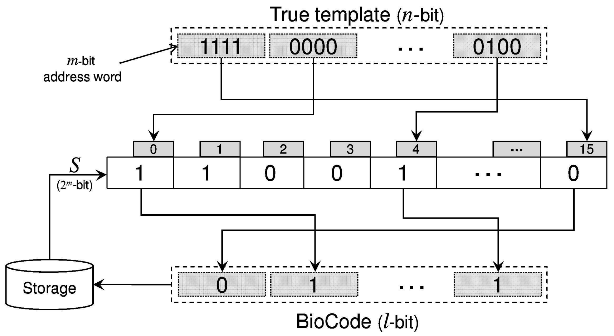


Fig. 2. Cancelable transformation of base BioEncoding.

In this paper, we propose a cancelable transformation method, based on BioEncoding, which can derive protected templates of any desired lengths from binary biometric data. In this section, we first give a brief overview of the base BioEncoding cancelable transformation scheme and then we describe the proposed variant.

A. Base BioEncoding

In BioEncoding, as illustrated in Fig. 2, bits in the true template are grouped into n/m m-bit words, where n is the bit-length of the original (unprotected) template. At the same time, a binary string S of length 2^m is randomly generated. Each word in the true iris code is mapped to a bit value in S located at the position addressed by the value of that word. For example, the first word '1111' in the true template, shown in Fig. 3, is mapped to the bit value located at position 15 $(= 1111_b)$ in S, i.e. '0'. The l $(= n/m)$ addressed bit values constitute the cancelable template (BioCode).

The most important advantage of BioEncoding over other CB methods is that it does not require the random sequence S to be neither unique nor secret. In other words, S need not be user specific as in BioHashing [5] and distorting transforms [7], for example. This is because even if the same sequence is employed with all users, different BioCodes will be generated due to the randomness that exists between iris codes generated from different eyes. In [34], BioEncoding was compared experimentally to BioHashing and the obtained results showed that BioEncoding, unlike BioHashing, does not deteriorate the recognition accuracy of the original biometric system even under the stolen-token scenario. Furthermore, the transform is non-invertible due to its many-to-one nature. Therefore, attackers would not be able to invert the transform, even if both a protected BioCode and S are known [6], [34]. In addition, BioEncoding offers significantly high renewability capacity. In fact, there are 2^{2^m} different binary strings that can be addressed using address words of length m. Therefore, a compromised

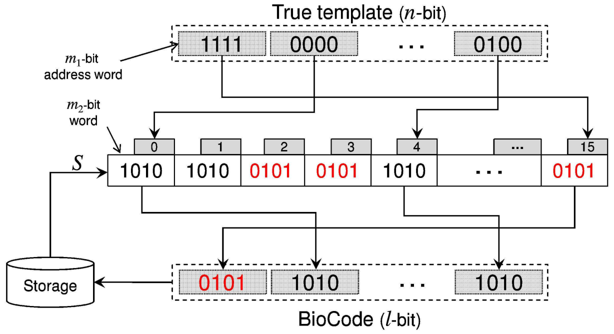


Fig. 3. Improved BioEncoding: bits replaced by words in S.

BioCode can be revoked and replaced easily by a new BioCode generated using a different string. Moreover, different random strings can be employed in different applications in order to guarantee diversity that hinders attackers from attacking different databases using cross-matching. Finally, with respect to the recognition accuracy, the performance of the original iris recognition system could be preserved if the random string S satisfies the following two conditions [6]: 1) has equal number of zeros and ones, i.e. balanced, and 2) totally different words in the true template, such as the words '0000' and '1111' shown in Fig. 2, address different bit values in S. In order to ensure that a randomly generated string satisfies these conditions, it should be tested before adopting it using the following formula:

$$f(S) = \sum_{i=0}^{2^m-2} \sum_{j=i+1}^{2^m-1} [1 \oplus (S_i \oplus S_j)].d_h(w_i, w_j) \quad (3)$$

where m is the length of any address word in the true template, S_i and S_j are the ith and jth bits in S, w_i and w_j are their corresponding address words, respectively; $\oplus$ is the XOR Boolean operator, and d_H stands for the hamming distance between the address words w_i and w_j.

Strings that have many similar bits addressed by different words in the true templates would have large f values and therefore should be avoided since they may decrease the discriminability between the generated BioCodes. However, it should be noted that although satisfying these conditions is necessary to preserve the recognition accuracy of the original unprotected system, it would restrict the renewability capacity since only $2m$ binary strings of length 2^m can strictly satisfy the above-mentioned conditions.

B. Modified BioEncoding

The basic idea behind BioEncoding lies in mapping multiple address words in the true template to one of two values; namely '0' or '1'. This many-to-one transformation is necessary to achieve the non-invertibility property required by CB. However, this does not necessarily mean that the random string S must consist of a sequence of 0's and 1's. Rather, any two distinct values could be used instead. For example, adopting a suitable matching technique in the transform domain, S might consist of two different words such as 'black' and 'white', two different symbols such as '+' and '-', or even two different binary words such as '1010' and '0101'.

In this paper, we propose to map each address word of size m_1 in the true template to one out of two binary words of size m_2, where $m_2 \geq 1$. Fig. 3 shows an example where $m_1 = m_2 = 4$. This approach exhibits two major advantages over base BioEncoding: 1) increased renewability capacity since different variations of word lengths and forms can be employed, and 2) more flexibility in choosing the length of the resultant BioCode since the length of the protected BioCode could be decided at will. The length l of the resultant BioCode could be decided according to the following formula:

$$l = (n \times m_2)/m_1 \quad (4)$$

Specifically, the second advantage is very important to our proposed hybrid template protection scheme since long BioCodes will allow for linking long enough keys as described in the next section.

V. Proposed Hybrid System

This section presents the proposed hybrid template protection system for securing cryptographic keys using arbitrary length protected iris codes. Fig. 4 illustrates the steps involved in both the enrollment and verification modules of the proposed system. The two modules are described in detail in the next subsections.

A. Enrollment

The enrollment module of the proposed system consists of two concurrent procedures. As shown in Fig. 4, the first procedure is responsible for preparing a revocable, non-invertible and protected iris template of arbitrary length n_p from a number of true iris codes of length n_o. On the other hand, the goal of the second procedure is to employ ECC(s) to encode an l-bit cryptographic key into an n_p-bit encoded string. Both processes are described below.

1) Protected templates generation: A practical biometric cryptosystem should be able to achieve high recognition accuracy, secure long-enough cryptographic keys and be robust against disclosure attacks of biometric templates. Here, we show how our proposed approach of integrating BioEncoding with the FCS can fulfill the above three requirements.

a) Extracting the most consistent bits for perfect recognition accuracy: Obviously, the recognition accuracy of the original biometric system has a significant impact on the performance of any biometric cryptosystem. Therefore, if the genuine and imposter distributions of the unprotected iris recognition system are not separated, the performance of the proposed hybrid system would be far from perfect. Generally, the only way for a biometric cryptosystem to achieve perfect accuracy is to have a clear separation between the genuine and imposter distributions of the original biometric system. Although iris is considered one of the most accurate biometric traits, such separation cannot be obtained using available public iris datasets. In fact, the near-to-perfect performance reported in some literature, such as the results reported by Hao et al. in [26] for example, is mainly due to the high quality iris images of the $private$ dataset employed in their experiments (only 3 out of 630 authentic samples, employed in [26], have a relatively high bit-error rates compared to the correction capability of the employed coding mechanism).

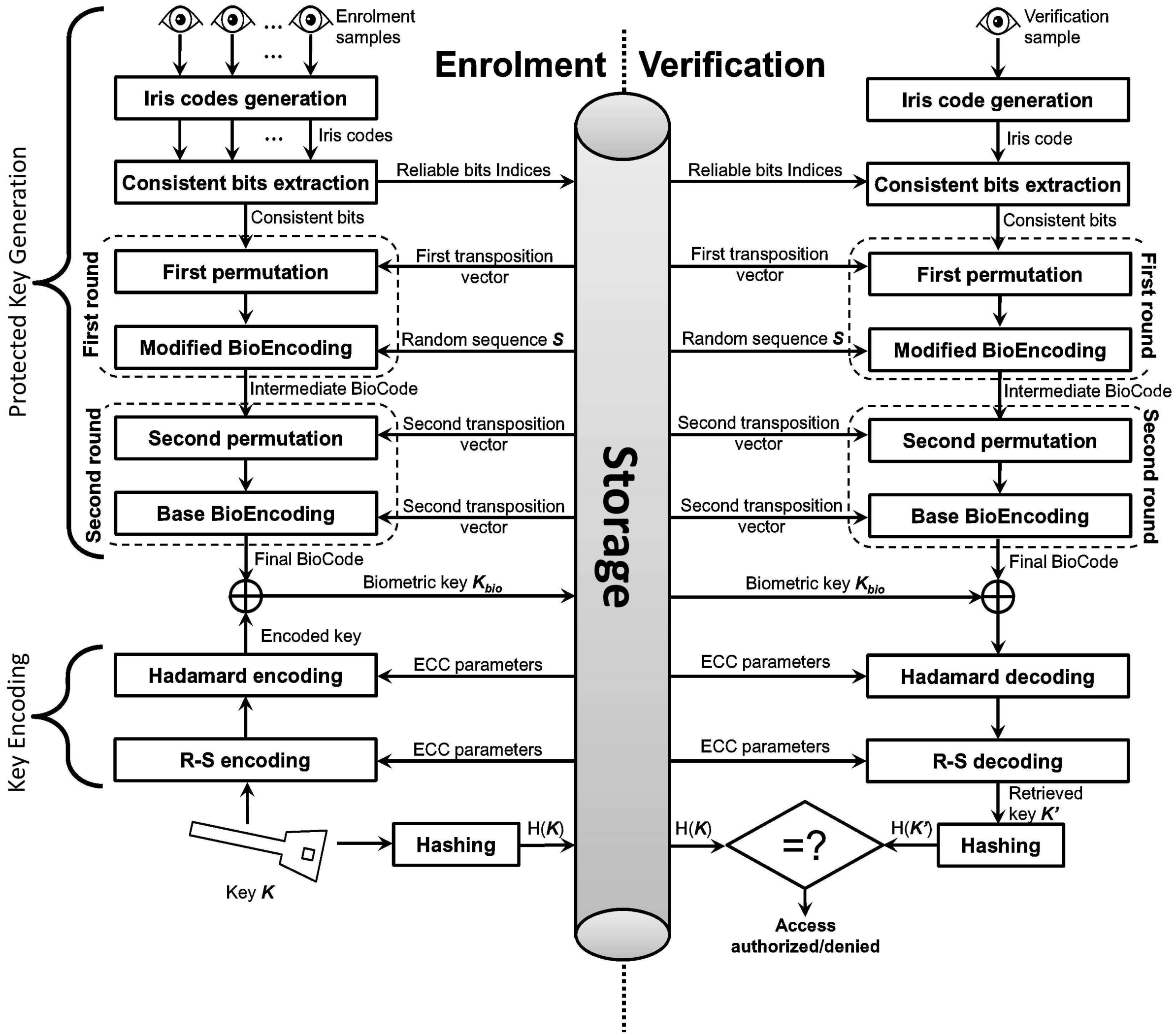


Fig. 4. Proposed hybrid template protection system.

Employing public iris datasets, a clear separation between the genuine and imposter distributions can be obtained by considering only the most consistent bits in iris codes. Consistent bits are bits that do not tend to change their values across iris codes generated from different images captured from the same eye. Hollingsworth et al. [35] showed empirically that considering only consistent bits in iris codes can improve the recognition accuracy significantly.

Therefore, in this work, a number p of iris images are acquired from a user's eye at enrollment and iris codes are generated from these images using the well known Daugman's algorithms for iris recognition [36]. Then, for all enrolled eyes in the dataset, different percentages of consistent bits (e.g. the most 50%, 40%, etc. consistent bits) are identified and extracted for each class. The genuine and imposter distributions are obtained for each percentage and the separability between the two distributions is measured using the decidability metric d' [36]:

$$d' = \frac{|\mu_i - \mu_g|}{\sqrt{(\sigma_i^2 + \sigma_g^2)/2}} \tag{5}$$

where μ_i and μ_g are the means and σ_i^2 and σ_g^2 are the variances of the imposter and genuine distributions, respectively.

Based on the obtained d' values, we adopt the largest percentage of bits that achieves a clear separation between the genuine and imposter distributions. That is, rather than applying BioEncoding on the entire iris codes, protected BioCodes are derived from the subset of bits that gives a clear separation between the two distributions and hence a perfect recognition accuracy in the original domain can be obtained.

Consistent bits are extracted by first aligning the p enrollment iris codes and then searching for bits that do not change their values across these codes. This can be done be summing up the corresponding bits and then sorting them according to the sum results. Bits that sum to 0 or p are referred to as perfectly consistent bits. For a specific number of bits $n_c < n_o$, if the number of perfectly consistent bits is larger than n_c, we randomly select only n_c bits from them; otherwise, we proceed with bits that sum to 1 or $p-1$, 2 or $p-2$, and so on, until we get a number of bits that is equal to or greater than the specified number n_c. The separation between the resulting genuine and imposter distributions is then investigated. If the two distributions are separated, we adopt this number or percentage of consistent bits; otherwise a smaller number of consistent bits is tested until the required separation is reached. The positions of these consistent bits in the iris code of user i are collected in a position vector P_i and stored in the centralized storage to be used at the time of verification.

b) Generating arbitrary length BioCodes: In order to generate non-invertible BioCodes of any desired length, two consecutive permutation and BioEncoding rounds are applied to the extracted consistent bits. The role of the first round is to obtain protected templates, which we refer to as intermediate BioCodes, of lengths that are much larger than the length of the original templates in order to make it possible for the final BioCodes (obtained after the second round) to be linked with long cryptographic keys.

For security reasons, which are described below, these intermediate BioCodes cannot be linked directly to cryptographic keys using the FCS construction. Therefore, to make the proposed hybrid system robust against attacks that exploit employing the ECCs in biometric cryptosystems to retrieve the bound/generated key [37], a second round is required to randomize any information that might be inferred from the intermediate BioCode by deriving a random and non-invertible template, called the final BioCode, from the intermediate BioCode.

In the first round, the extracted consistent bits are firstly permuted using a random permutation key and then BioEncoded using the modified BioEncoding scheme proposed in Section III. This permutation step is necessary for two reasons. The first is to diminish the local correlation that may exist among the extracted consistent bits. Experiments conducted by Daugman [36] on a large iris dataset showed that only 249 degrees of freedom exist in 2048-bit iris codes. Therefore, permuting the extracted consistent bits is necessary for randomizing those bits and increasing their entropy. The second reason is to secure the protected BioCodes against correlation attacks [38]. One important drawback of BioEncoding is that although it is noninvertible for a single protected template, it might be possible to recover the original iris template by correlating several protected templates created from the same iris. Therefore, it is important to change the value of address words in original templates before applying BioEncoding in every new application to hinder attackers from exploiting this type of attacks. Permuting bits in the true iris codes employing different permutations in different applications is a simple yet efficient way to achieve this objective [38].

After obtaining the permuted bits, the modified BioEncoding is applied to extract the intermediate BioEncoded template. As described in Section III, using this version of BioEncoding, protected BioCodes of an arbitrary length n_i can be generated from the n_c consistent bits. Assuming that the true consistent bits are grouped into address words of size m_1, the size r of any word in the random string S can be calculated as follows:

$$r = n_i \times m_1 / n_c \tag{6}$$

As will be described later, in this work, the Hadamard ECC is employed to encode the randomly generated cryptographic key. Therefore, unfortunately, the obtained intermediate template cannot be linked directly with the encoded key. This is because it is possible to fully recover a Hadamard codeword if only few bits are known. Specifically, knowing only 7 bits, he could completely break the biometric cryptosystem of Kanade et al. [14] in which the (32, 6, 8) Hadamard ECC is employed. To boost the correction power of the Hadamard ECC, authors in [14] insert two zeros after every three bits in the original iris template. Hence, there are at least 12 known bits in every 32-bit Hadamard codeword. This is more than enough to recover all the codewords and hence the entire key.

Although no bits are explicitly known in the intermediate template obtained from the first round, a similar attack could be applied to our system if the encoded key is committed directly using this intermediate BioCode. Consider, for example, that words in the true iris template are mapped to one of the following two values in S: S_1 = “10110010” or S_2 = “01001101” where the length of words in S is 8. For the first Hadamard codeword, the attacker may assume that it starts with S_1 and hence the first 8 bits will be known. Although this is not sufficient for recovering the entire codeword, the search space will be reduced dramatically. If, on the other hand, the first Hadamard codeword starts with S_2, the result will be complementary to the S_1 case. The same procedure can be applied to the remaining parts of the Hadamard codeword at hand to reduce the search space further.

In order to hinder this type of attacks, an extra round of random permutation and (base) BioEncoding is applied. In this round, the resulting intermediate BioCode is first permuted using a second random permutation key in order to increase the system robustness against invertibiliy attacks. To realize the importance of this permutation step, let use consider the above example again assuming further that the resulting intermediate BioCode is divided into 4-bit address words in the subsequent BioEncoding step. If no permutation is applied, there would be only two possible address words for every bit in S. That is, every odd-numbered bit in the resulting BioCode could be addressed by either “1011” or “0100”. Likewise, every even-numbered bit could be addressed by either “0010” or “1101”. It is worth noting that although the permutation key need not be user specific, and hence the same key can be employed for all users enrolled at the same application, different permutation keys should be employed in different applications to prevent correlation attacks [38].

Using base BioEncoding, bits in the permuted template are grouped into m_2-bit words and each word is mapped to a single bit in a binary string S of length 2^{m_2}. The length n_p of the resulting (final) BioCode will be:

$$\begin{aligned} n_p &= n_i / m_2 \quad (7) \\ &= r \times n_c / (m_1 \times m_2) \quad (8) \end{aligned}$$

According to the length of the consistent bit vector, extracted from the enrollment iris codes, as well as the required length of the protected BioCode, different values of the BioEncoding's (base and modified) parameters can be chosen. For example,

TABLE I. EXAMPLES OF DIFFERENT PARAMETER SETTINGS OF BIOENCODING WHEN $n_c = 1024$ AND $n_p = 2048$

	n_c	m_1	r	n_i	m_2	n_p
1	1024	4	32	8192	4	2048
2	1024	8	64	8192	4	2048
3	1024	4	64	16384	8	2048

if it is required to generate a 2048-bit protected BioCode from an 1024-bit consistent bit vector, there would be more than one choice for values of m_1, r, and m_2. Table I shows some examples of these values.

2) Cryptographic Key Encoding and Linking: In this module, an l-bit cryptographic key K is randomly generated and encoded into an n_p-bit codeword C using the two-layer error correcting (EC) scheme described in [26]. This concatenated EC scheme combines Hadamard and Reed-Solomon ECCs to deal with background and burst errors in iris codes, respectively. In the first layer, K is encoded using a Reed-Solomon ECC as follows. Bits in K are divided into k_{RS} blocks of k-bit each. This set of blocks is then represented as a message of k_{RS} symbols over F_{2^k} and encoded into a codeword of n_{RS} symbols using a (n_{RS}, k_{RS}, t_{RS}) Reed-Solomon code that has a correction capacity $t_{RS} = (n_{RS} - k_{RS})/2$. In the second layer, each of the resulting n_{RS} symbols is represented as a k-bit word and encoded into a 2^{k-1} bit codeword using a (2^{k-1}, k, 2^{k-2}) Hadamard code. Such Hadamard code is generated from a Hadamard matrix of order $k - 1$ and can correct $2^{k-3} - 1$ erroneous bits in each codeword. The length n_p of the final encoded key can be calculated as follows:

$$n_p = n_{RS} \times 2^{k-1} \tag{9}$$

The correction capability of this concatenated EC scheme depends primarily on the values of both n_{RS} and k_{RS}. The lowest correction capacity is obtained when $n_{RS} = k_{RS}$ (i.e., only Hadamard encoding is employed). That is, the described two-layer EC scheme can correct at least up to approximately 25% of the encoded codeword, since the correction capability of sole Hadamard coeds is up to 25% [39]. On the other hand, the correction capability of the two-layer scheme is increased by increasing the difference between n_{RS} and k_{RS}.

In this paper, the (128, 8, 64) Hadamard code is adopted and different Reed-Solomon codes are employed based on the required key length as well as the intra- and inter-user distributions of the generated BioCodes.

Finally, the resulting codeword is XORed with the generated protected BioCode to get the biometric key K_{bio}. At the same time, the hash value of the key $H(K)$ is computed using any secure hash function. Finally, the original iris template as well as the protected one is discarded safely and only the biometric key along with its hash value are stored in the centralized storage for further processing during verification.

B. Verification

At the time of verification, as illustrated in Fig. 4, a single iris image is captured from the eye being verified and its iris code is generated using the same procedure applied on enrollment. Using the stored position vector, the most n_c consistent bits are extracted from the generated code. It should be noted that due to the misalignment that may be found between the enrolled images and the image captured at verification, the generated iris code is shifted eight times in the left and right directions, as suggested by Daugman [36], and the process of extracting the most consistent bits is repeated after each shift. The first permutation vector, stored on enrollment, is then applied to the extracted consistent bits and the modified BioEncoding cancelable transformation is employed to derive the intermediate protected BioCode from the permuted original bits. Then, the second permutation followed by the base BioEncoding transformation process are applied to the intermediate BioCode to obtain the final protected iris code. To retrieve the secured key, the final protected code is XORed with the stored biometric key, K_{bio}, and the resulting bit string, is decoded using the concatenated scheme used on enrollment. Finally, the hash of the retrieved key, K', is computed and compared to the stored hash. Only if the two hash values $H(K)$ and $H(K')$, coincide, the key is released; otherwise, the authentication process fails.

VI. EXPERIMENTAL RESULTS

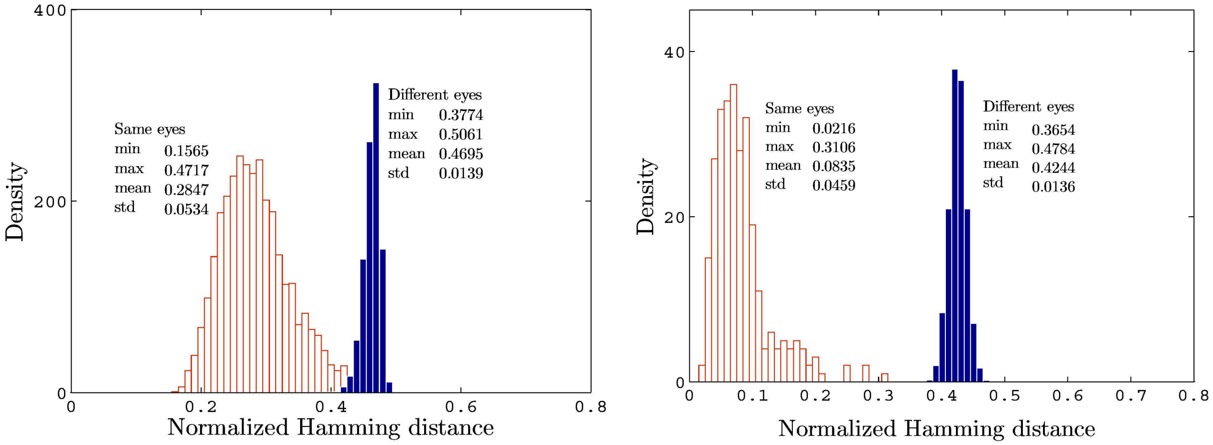
The publicly available CASIA-IrisV3-Interval iris images dataset [40] was used to evaluate the proposed system. This database contains 2639 8-bit gray scale images, with a resolution of 320 × 280 pixels, captured from 396 different classes (eyes). However, many classes in this dataset have just a small number of images. Therefore, a subset contains 700 images of 70 different classes, classes that have 10 images in the database, was used.

The open source MATLAB implementation for iris recognition provided in [41] was employed to generate 9600-bit iris codes, together with their corresponding 9600-bit noise masks, for all images in the selected subset. The normalized Hamming distances between all possible iris code pairs, considering all bits in the generated iris codes and taking the noise masks into account, were calculated using the following formula [36]:

$$d_H = \frac{\|(CodeA \oplus CodeB) \cap MaskA \cap MaskB\|}{\|MaskA \cap MaskB\|} \tag{10}$$

where $CodeA$ and $CodeB$ are the two iris templates being matched, $MaskA$ and $MaskB$ are the noise masks corresponding to $CodeA$ and $CodeB$ respectively and $\cap$ represents the bitwise AND operation. Fig. 5(a) shows the genuine and imposter normalized Hamming distances distributions for the adopted iris images. A clear overlap between the two distributions can be seen in the figure. The separation between the two distributions measured using the decidability metric (d'), defined in Eq. (5), is 4.061.

The separation between the genuine and imposter distributions was checked for different numbers of consistent bits. Fig. 6 shows the decidability values that result from comparing only consistent bits that represent different percentages (from 5% to 25% step 5%) of the entire length of an iris code (that is, 9600). We used bits that are perfectly consistent in fewer images, i.e. 5 images, then 4 images, etc. for eyes whose perfectly consistent bits are smaller than the tested percentage. It is clear from Fig. 6 that the decidability value decreases apparently when the number of the tested consistent bits exceeds 10%(960 bits) of



(a) (b)

Fig. 5. Genuine and imposter distributions for (a) true iris codes, (b) the most 1024 consistent bit vectors extracted from the tested iris codes.

the entire length of the true iris code. Therefore, we decided to extract the most consistent 1024 (nearest power of two) bits from all iris codes used in our experiments. For each class in the dataset, the extracted 1024 consistent bits were collected in a consistent bit vector that is assigned to that class. Fig. 5(b) shows the genuine and imposter distributions resulted from comparing those consistent bit vectors. As shown in this figure, the statistics of both distributions imply that they are separable enough for achieving perfect recognition accuracy.

The size of the cryptographic key to be committed securely using a protected template of length n_p relies on the parameters of the employed ECCs. Hence, it is not known in advance how long the protected template should be to reliably secure a 128-bit key, for example. We began by deriving 2048-bit protected templates from the extracted 1024-bit consistent vectors and evaluating the performance of the proposed system for all possible key lengths that can be linked with these protected templates using the employed EC scheme. Recall that, as indicated from the examples in Table I, BioEncoding can be configured in many different ways to obtain 2048-bit protected templates from 1024-bit true templates.

As shown in Table II, employing the $(64, 7, 32)$ Hadamard code, it is possible to employ the derived protected templates to commit keys of lengths up to 224-bit (N.B. $224 \div 7 \times 64 = 2048$). In this case, the RS code is not used and hence the matching accuracy will be affected since no block errors are corrected. Results in Table II shows that with the introduction of RS codes, better error rates can be obtained at the expense of the key length. Perfect accuracy was achieved for $|K| = 42$.

To achieve this perfect performance for longer keys, longer protected templates should be generated. We repeated the above experiment, employing the same HC, using 4096-bit protected keys. Protected templates of such size can be obtained simply by increasing the length r of words in the random string employed in the first round of the protected template generation module of the proposed hybrid system. Based on the results obtained from the previous experiment, shown in Table II, we began by evaluating the system performance for key length $|K| = 56$ and we continued checking the accuracy for longer keys until a non-perfect performance was observed. Table III shows that a cryptographic key of size up to 154-bit can be committed and released perfectly (with 0% error rates) using the proposed hybrid system when 4096-bit protected templates were employed.

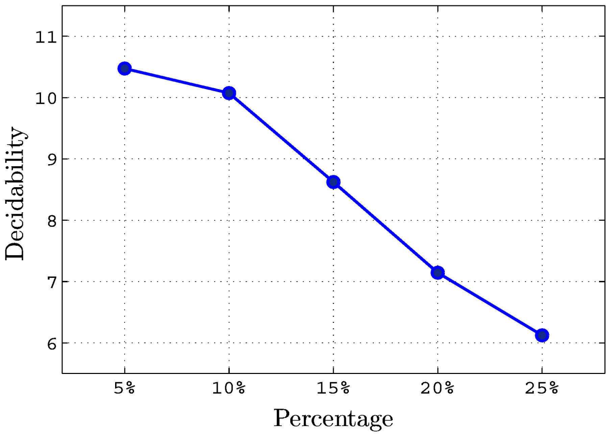


Fig. 6. Decidability values that result from comparing consistent bit vectors of different lengths (lengths represent different percentages of the entire length of an iris code).

The above experiment was repeated again using protected templates of length $n_p = 6144$. Employing the same EC scheme, we checked the system performance for different key sizes (starting from 168-bit keys). The obtained results, shown in Table IV, indicates that employing protected templates of the mentioned size can secure cryptographic keys of long-enough sizes (up to 280-bit).

In general, since the proposed hybrid system can generate protected iris templates of an arbitrary length, the obtained results indicate that the proposed hybrid system exhibits perfect recognition accuracy regardless of the size of the key that is secured using the generated revocable iris template. A comparison between the proposed system and a number of recent iris cryptosystems is shown in Table V.

TABLE II. System performance employing $HC(64, 7, 32)$ when $n_p = 2048$

$\|K\|$	t_{RS}	FRR(%)	FAR(%)
224	0	4.64	0
210	1	2.86	0
196	2	2.86	0
182	3	2.14	0
168	4	2.14	0
154	5	1.78	0
140	6	1.78	0
126	7	1.07	0
112	8	1.07	0
98	9	1.07	0
84	10	0.71	0
70	11	0.71	0
56	12	0.36	0
42	13	0	0

TABLE III. System performance employing $\mathrm{HC}(64, 7, 32)$ when $n_p = 4096$

$\|K\|$	t_{RS}	FRR(%)	FAR(%)
56	28	0	0
70	27	0	0
84	26	0	0
98	25	0	0
112	24	0	0
126	23	0	0
140	22	0	0
154	21	0	0
168	20	0.36	0

TABLE IV. System performance employing $\mathrm{HC}(64, 7, 32)$ when $n_p = 6144$

$\|K\|$	t_{RS}	FRR(%)	FAR(%)
168	36	0	0
182	35	0	0
196	34	0	0
210	33	0	0
224	32	0	0
238	31	0	0
252	30	0	0
266	29	0	0
280	28	0	0
294	27	0.36	0

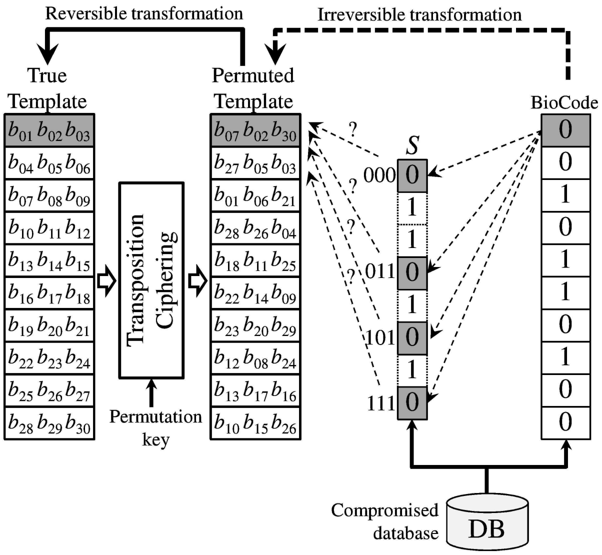


Fig. 7. Obtaining the true iris template from the protected one is infeasible due to the many-to-one nature of the transform.

VII. Security Analysis

The advantage of the proposed hybrid system over other hybrid template protection schemes is that it can protect cryptographic keys as well as true iris templates at the same time. Unfortunately, because other hybrid template protection techniques employ invertible cancelable transformations for deriving the protected biometric templates, it would be very simple to obtain the true templates from the protected ones if the keys, linked with the cancelable templates, are disclosed for any reason. In the proposed system, on the other hand, if a linked key is compromised, obtaining the true iris template from the protected one will be computationally very hard due to the many-to-one nature of the BioEncoding cancelable transformation process.

As shown in Fig. 7, even if both the random string S and the permutation key are known. Every bit in the protected BioCode could be originated from 2^{m-1} different address words in the true iris template where m is the size of any address word (in the example shown in Fig. 7, every BioCode bit could be originated from four different address words). Therefore, recovering all bits in the BioCode requires $2^{l(m-1)}$ trials, where l is the BioCode length. Since $l = n/m$, where n is the length of the true iris template, recovering all BioCode bits would require $2^{n(m-1)/m} \approx 2^n$, if m is large. That is, recovering the true template from the protected BioCode is approximately as difficult as guessing all bits in the true template [38]. It is important to note that although the permutation process is a reversible process, knowing the permutation key is useless unless the (irreversible) BioEncoding process is reversed. That is, the two permutation keys as well as the random strings employed in our proposed system need not be user-specific; rather, they can be treated as public data without affecting the security of the system.

On the other hand, it is important to evaluate the system security with respect to the committed cryptographic key. As shown in Fig. 5(b), the maximum intra-user fractional Hamming distance is 0.3106 and therefore a perfect accuracy can be achieved as long as the correction capability of the employed EC scheme is larger than 31.06%. As a result, in order to recover the committed key successfully, it is enough for an attacker to find a 1024-bit string which is ≤ 319-bit Hamming distance from the consistent iris bit vector. Accordingly, the key strength can be measured in terms of the entropy of the key (E) as follows [26]:

$$E = log_2 \frac{2^{1024}}{\binom{1024}{318}} = 114 \tag{11}$$

That is, the attacker needs at least 2^{114} computations to recover the key successfully. It is worth noting that, however, every single computation involves two rounds of BioEncoding

TABLE V. Comparison with recent iris cryptosystems

Work	Iris templates revocable?	Dataset/subset (# samples/# classes)	Key size	FRR (%)	FAR (%)
Hao et al. [26]	no	private (700/70)	140	0.47	0
Yang et al. [31]	no	CASIA ver.1 (756/108)	92	0.8	0
Lee et al. [42]	no	BERC ver.1 (990/99)	128	0.775	0
Bringer et al. [27]	no	CASIA Ver.1 (756/108) ICE 2005 (2953/244)	42	0.0665 0.0562	0 $< 10^{-5}$
Kanade et al. [14]	no	ICE 2005 (2953/244)	198	1.04	0.055
Zhang et al. [33]	no	private (348/128)	938	0.52	0
Ziauddin et al. [28]	no	Bath (500/25)	260	0	0
Chai et al. [43]	no	CASIA ver3-Interval (868/124)	200	3.63	0
Proposed	yes	CASIA ver3-Interval subset(700/70)	unlimited	0	0

and transposition processes followed by and XORing operation with the biometric key. This implies that it would be computationally very expensive, and might be infeasible, to obtain the cryptographic key committed using our proposed hybrid template protection scheme.

VIII. Conclusion and Discussion

This paper presented a new hybrid template protection scheme for protecting iris codes as well as securing cryptographic keys. A novel cancelable transformation method, based on BioEncoding, has been proposed to derive cancelable iris templates of any desired size form the most consistent bits in original iris codes. The derived cancelable templates were employed to secure cryptographic keys using the fuzzy commitment scheme. The proposed iris cryptosystem exhibits four major advantages over other existing systems. First, all parameters and variables employed either in the cancelable transformation process or in the binding process need not be user-specific and therefore the proposed scheme is tokenless. Second, thanks to the proposed cancelable transformation method, the presented iris cryptosystem can secure cryptographic keys of an arbitrary length. Third, since cryptographic keys are secured using cancelable iris templates rather than original ones, the revocability requirement is satisfied for both cryptographic keys and iris templates. Finally, thanks to extracting the most consistent bits from the true iris codes, experimental results showed that the proposed system achieves perfect recognition accuracy (0% ERR) regardless of the key size. This is achieved at the expense of storing the positions of the consistent bits in the application database in order to successfully match the probe sample with the gallery sample during authentication. If the adversary could gain access of these position indices, he might try to cross-match them across different applications. As a future work, we intend to deal with this issue via utilizing the challenge-response protocol in order to recover these indices without explicitly storing them in the application database.

IX. Acknowledgment

We would like to thank the Deanship of Scientific Research at Jouf University for providing the financial support for publishing this work.

References

[1] D. Sadhya, S. K. Singh and B. Chakraborty, ”Review of key-binding-based biometric data protection schemes.” IET Biometrics, vol. 5, no. 4, pp. 263-275, 2016.

[2] Nita, Stefania, Marius Mihailescu and Valentin Pau, ”Security and cryptographic challenges for authentication based on biometrics data.” Cryptography vol. 2, no. 4, pp. 1-22, 2018.

[3] Nitin Kumar, ”Cancelable Biometrics: a comprehensive survey”, Artificial Intelligence Review, vol. 2019, pp. 1-44, 2019.

[4] Ravi Das, Biometric technology: authentication, biocryptography, and cloud-based architecture. CRC press, 2014.

[5] Y. Zheng, Y. Cao and C. Chang, ”Facial biohashing based user-device physical unclonable function for bring your own device security,” IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, pp. 1-6, 2018.

[6] O. Ouda, N. Tsumura and T. Nakaguchi, ”Tokenless Cancelable Biometrics Scheme for Protecting Iris Codes,” in Proc. IAPR Conference on Pattern Recognition (ICPR 2010), August, 2010.

[7] H. Kaur and K. Pritee, ”Non-invertible biometric encryption to generate cancelable biometric templates.” In Proceedings of the World Congress on Engineering and Computer Science, vol. 1, pp. 1-4. 2017.

[8] A. Schaller, T. Stanko, B. Škorić and S. Katzenbeisser, ”Eliminating Leakage in Reverse Fuzzy Extractors,” in IEEE Transactions on Information Forensics and Security, vol. 13, no. 4, pp. 954-964, April 2018.

[9] A. Panwar, P. Singla and M. Kaur. ”Techniques for enhancing the security of fuzzy vault: a review.” In Progress in Intelligent Computing Techniques: Theory, Practice, and Applications, pp. 205-213. Springer, Singapore, 2018.

[10] S. Chauhan, A. Sharma, ”Improved fuzzy commitment scheme.” International Journal of Information Technology, pp. 1-11, 2019.

[11] R.K. Bharathi, S. D. Mohana, ”A Review on Biometric Template Security.” In: V. Sridhar, M. Padma, K. Rao (eds) Emerging Research in Electronics, Computer Science and Technology. Lecture Notes in Electrical Engineering, vol. 545, pp. 589-596, Springer, Singapore, 2018.

[12] H. Liu, D. Sun, K. Xiong, Z. Qiu, ”A hybrid approach to protect palmprint templates.” The Scientific World Journal, 2014.

[13] M. Sandhya and M. V. Prasad, ”Cancelable fingerprint cryptosystem based on convolution coding.” In Advances in Signal Processing and Intelligent Recognition Systems, pp. 145-157, Springer, 2016.

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

[15] S. Kanade, D. Petrovska-Delacr´etaz and B. Dorizzi, ”Multi-Biometrics Based Cryptographic Key Regeneration Scheme,” In IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS), September 2009.

[16] S. Kanade, D. Camara , D. Petrovska-Delacr´etaz and B. Dorizzi, ”Application of Biometrics to Obtain High Entropy Cryptographic Keys,” in Proceedings of World Academy of Science, Engeneering and Technology, vol. 39, 2009.

[17] S. Kanade, D. Petrovska-Delacr´etaz and B. Dorizzi, ”btaining cryptographic keys using feature level fusion of iris and face biometrics for secure user authentication,” In IEEE Computer Vision and Pattern Recognition Workshops (CVPRW), June 2010, 138-145.

[18] Y. Wang and K. N. Plataniotis, ”Fuzzy vault for face based cryptographic key generation,” in Biometric Consortium Conference, Baltimore, September 2007.

[19] K. Nandakumar, A. Nagar and A. K. Jain, ”Hardening Fingerprint Fuzzy Vault Using Password,” in Proceedings of Second International Conference on Biometrics, Seoul, South Korea, pp. 927-937, 2007.

[20] A. B. J. Teoh, D. C. L. Ngo and A. Goh, ”Personalised cryptographic key generation based on FaceHashing,” Comput. Secur. 23:606-614, 2004.

[21] O. T. Song, A. B. J. Teoh and D. C. L. Ngo, ”Application-Specific Key Release Scheme from Biometrics,” International Journal of Network Security, vol. 6, no. 2, pp. 127-133, March 2008.

[22] Yi C. Feng, Pong C. Yuen and Anil K. Jain, ”A Hybrid Approach for Generating Secure and Discriminating Face Template,” IEEE Transactions on Information Forensics and Security, 5(1):103-117, 2010.

[23] J. Daugman, ”Information theory and the iriscode.” IEEE transactions on information forensics and security, vol. 11, no. 2, pp. 400-409, 2015.

[24] B. Topcu, C. Karabat, M. Azadmanesh and H. Erdogan, ”Practical security and privacy attacks against biometric hashing using sparse recovery.” EURASIP Journal on Advances in Signal Processing, vol. 2016, no. 1, 2016.

[25] M. Lafkih, P. Lacharme, C. Rosenberger, M. Mikram, S. Ghouzali, M. El Haziti, D. Aboutajdine, ”Vulnerabilities of fuzzy vault schemes using biometric data with traces.” In International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 822-827, 2015.

[26] F. Hao, R. Anderson and J. Daugman, ”Combining crypto with biometrics effectively,” IEEE Transactions on Computers, vol. 55, no. 9, pp. 1081-1088, 2006.

[27] J. Bringer, H. Chabanne, G. Cohen, B. Kindarji and G. Z´emor, ”Theoretical and practical boundaries of binary secure sketches,” IEEE Transactions on Information Forensics and Security, 3(4):673-683, 2008.

[28] S. Ziauddin, M. N. Dailey, ”Robust iris verification for key management,” Pattern Recognition Letters, vol. 31, no. 9, pp. 926-935, 2010.

[29] T. A. T. Nguyen, D. T. Nguyen and T. K. Dang, ”A multi-factor biometric based remote authentication using fuzzy commitment and non-invertible transformation.” In Information and Communication Technology-EurAsia Conference, pp. 77-88, 2015.

[30] M. Yasuda, T. Shimoyama, N. Abe, S. Yamada, T. Shinzaki, and T. Koshiba, ”Privacy-preserving fuzzy commitment for biometrics via layered error-correcting codes.” In International Symposium on Foundations and Practice of Security, pp. 117-133, 2015.

[31] S. Yang, I. Verbauwhede, ”Secure Iris Verification,” Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp. II-15-20, Honolulu, HI, April 2007.

[32] S. Adamovic, M. Milosavljevic, M. Veinovic, M. Sarac and A. Jevremovic, ”Fuzzy commitment scheme for generation of cryptographic keys based on iris biometrics.” IET Biometrics, vol. 6, no. 2, pp. 89-96, 2016.

[33] L. Zhang, Z. Sun, T. Tan and S. Hu, ”Robust biometric key extraction based on iris cryptosystem,” In Proceedings of the 3rd International Conference on Biometrics 2009 (ICB09) LNCS: 5558, pp. 1060-1070, 2009.

[34] O. Ouda, N. Tsumura and T. Nakaguchi, ”BioEncoding: a reliable tokenless cancelable biometrics scheme for protecting IrisCodes,” IEICE Trans. Inf & Syst., vol.E93-D, no.7, July 2010.

[35] K. P. Hollingsworth, K. W. Bowyer and P. J. Flynn, ”The best bits in an iris code,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 6, pp. 964-973, 2009.

[36] J. Daugman, ”Evolving methods in iris recognition.” In IEEE International Conference on Biometrics: Theory, Applications, and Systems,(BTAS07), 2016.

[37] D. Akdoğan, D. K. Altop and A. Levi, ”‘Secure key agreement using pure biometrics.” In 2015 IEEE Conference on Communications and Network Security (CNS), pp. 191-199, 2015.

[38] O. Ouda, N. Tsumura and T. Nakaguchi, ”Security enhanced BioEncoding for protecting iris codes,” in Biometric Technology for Human Identification VIII, Proc. SPIE 8029, 80291U, 2011.

[39] C. Francisco, T. A. Oliveira, A. Oliveira and L. Grilo, ”Hadamard matrices and links to information theory.” In AIP Conference Proceedings, vol. 1978, No. 1, p. 460008, AIP Publishing, 2018.

[40] CASIA iris image database, Available: http://www.cbsr.ia.ac.cn/Databases.htm.

[41] Libor Masek, Peter Kovesi. MATLAB Source Code for a Biometric Identification System Based on Iris Patterns. The School of Computer Science and Software Engineering, The University of Western Australia. 2003.

[42] Y.J. Lee, K. Bae, S.J. Lee, K.R. Park, J. Kim, ”Biometric key binding: Fuzzy vault based on iris images,” in: Springer LNCS 4642: International Conference on Biometrics, August 2007, pp. 800-808.

[43] T. Y. Chai, B. M. Goi, Y. H. Tay, and Z. Jin, ”A New Design for Alignment-Free Chaffed Cancelable Iris Key Binding Scheme.” Symmetry, vol. 11, no. 2, p. 164, 2019.