IEICE TRANS. INF. & SYST., VOL.E93–D, NO.7 JULY 2010

PAPER

# BioEncoding: A Reliable Tokenless Cancelable Biometrics Scheme for Protecting IrisCodes

**Osama OUDA**[†a)], ***Nonmember***, **Norimichi TSUMURA**[†], ***and*** **Toshiya NAKAGUCHI**[†], ***Members***

**SUMMARY** Despite their usability advantages over traditional authentication systems, biometrics-based authentication systems suffer from inherent privacy violation and non-revocability issues. In order to address these issues, the concept of cancelable biometrics was introduced as a means of generating multiple, revocable, and noninvertible identities from true biometric templates. Apart from BioHashing, which is a two-factor cancelable biometrics technique based on mixing a set of tokenized user-specific random numbers with biometric features, cancelable biometrics techniques usually cannot preserve the recognition accuracy achieved using the unprotected biometric systems. However, as the employed token can be lost, shared, or stolen, BioHashing suffers from the same issues associated with token-based authentication systems. In this paper, a reliable tokenless cancelable biometrics scheme, referred to as BioEncoding, for protecting IrisCodes is presented. Unlike BioHashing, BioEncoding can be used as a one-factor authentication scheme that relies only on sole IrisCodes. A unique noninvertible compact bit-string, referred to as BioCode, is randomly derived from a true IrisCode. Rather than the true IrisCode, the derived BioCode can be used efficiently to verify the user identity without degrading the recognition accuracy obtained using original IrisCodes. Additionally, BioEncoding satisfies all the requirements of the cancelable biometrics construct. The performance of BioEncoding is compared with the performance of BioHashing in the stolen-token scenario and the experimental results show the superiority of the proposed method over BioHashing-based techniques.

***key words:*** *biometric authentication, biometric encryption, cancelable biometrics, BioEncoding, BioHashing, IrisCode*

## 1. Introduction

In today's world, identity authentication is gaining increasing importance in a growing number of applications in order to combat identity theft and fraud. Unfortunately, traditional personal authentication systems which rely on something the user knows, such as passwords and personal identification numbers (PINs), and/or something the user has, such as smart cards and USB tokens, suffer from several inherent limitations. Both tokens and passwords can be easily forgotten, lost, shared or stolen. On the other hand, biometrics-based authentication systems, which use physiological and/or behavioral characteristics such as iris, fingerprint and gait for identifying persons, do not suffer from such limitations. Biometric characteristics are permanently associated with the user and hence cannot be forgotten, lost or lent to others. Due to these usability advantages, biometric characteristics are increasingly replacing passwords and tokens in many security applications [1]. However, biometrics-based authentication systems also have their own inherent limitations. Unlike passwords and tokens which can be easily canceled and replaced if they are lost or stolen, biometrics cannot be revoked if compromised (stolen by an imposter). Furthermore, since biometrics can convey sensitive private information about individuals such as disease and genetic information, their use in personal authentication raises many privacy concerns [2].

In recent years, several approaches have been suggested to address both revocability and privacy-preserving issues of biometrics. One of the most promising approaches that have been proposed to address the abovementioned issues is the "cancelable biometrics" approach [3], [4]. The concept of cancelable biometrics is based on the notion of noninvertible transforms that can be applied to true biometric templates in order to derive multiple protected templates which can be canceled and replaced in case of compromise. However, the main challenge for the cancelable biometrics techniques proposed in the literature is the trade-off between privacy preservation and recognition accuracy preservation. It has been shown that this trade-off cannot be resolved without integrating other factors such as tokens or secure passwords in the authentication process.

*BioHashing* [5] is a two-factor form of cancelable biometrics based on iterated inner products between a set of *tokenized* user-specific random numbers (TRNs) and the true biometric features. BioHashing has been successfully applied to different biometric modalities [5]–[8] and its reported near-to-perfect results have received much attention [9], [10]. However, the main drawback of BioHashing is that its claimed performance degrades significantly when an imposter gains access to a legitimate token and tries to authenticate herself as a legitimate user. In other words, cancelable biometrics techniques based on BioHashing have the same limitations associated with traditional token-based authentication systems. Therefore, in order to benefit from the usability advantages of biometrics and protect users' privacy at the same time, new template protection methods that rely solely on biometric templates are desired.

Addressing this problem, this paper presents a novel *tokenless* cancelable template protection scheme, called *BioEncoding*, for protecting *IrisCodes*. Using the proposed method, a large number of protected templates can be derived from a true IrisCode by grouping bits in the IrisCode into fixed-sized groups and then mapping them randomly to single bit values to constitute a compact protected bit string, called *BioCode*. Rather than the true IrisCode, the

Manuscript received October 8, 2009.
Manuscript revised March 4, 2010.

[†]The authors are with the Graduate School of Advanced Integration Science, Chiba University, Chiba-shi, 263–8522 Japan.
a) E-mail: oouda@graduate.chiba-u.jp

DOI: 10.1587/transinf.E93.D.1878

generated BioCode can be used efficiently to verify the user identity without introducing significant performance degradation. Although BioEncoding also makes use of a random sequence in the transformation process, no token is needed to store that sequence because the transformation method relies on the user's IrisCode much more than the random sequence. That is, unlike BioHashing, the same random sequence can be used with all users without deteriorating the recognition performance. In addition to preserving the recognition accuracy, BioEncoding satisfies all the requirements of cancelable biometrics construct.

The rest of this paper is organized as follows. In Sect. 2, we briefly review the current state of biometric template protection techniques focusing on cancelable biometrics and BioHashing. In Sect. 3, the design details of BioEncoding are described. In Sect. 4, the diversity and noninvertibility properties of BioEncoding are discussed. Sections 5 presents a set of experiments for evaluating the recognition performance of BioEncoding and compares it with results obtained using BioHashing in the stolen-token scenario. Finally, conclusions are drawn in Sect. 6.

## 2. Biometric Template Protection

In the last few years, the issue of protecting biometric templates has gained a great deal of attention from both the biometric and cryptographic research communities, and many different approaches have been proposed to address the associated revocability and privacy concerns. Generally, template protection schemes that have been proposed in the literature so far can be broadly classified into two main categories [11]: biometric cryptosystems and cancelable biometrics schemes. Biometric cryptosystems may be further divided into two subcategories: key generation schemes and key binding schemes. Key generation schemes [12], [13] seek to derive cryptographic keys from user's biometric features directly. The main drawback of such schemes lies in the difficulty of generating exact error-free identifiers from noisy biometric features with high key entropy [11]. On the other hand, the goal of key binding schemes is to bind cryptographic keys with biometric features in such a way that makes it impossible to recover the key unless the true template is presented during authentication [14], [15]. However, the performance of key binding schemes may be affected due to the introduction of error correction schemes, which are necessary for key retrieval.

It is worth noting that although both key generation schemes and key binding schemes provide protection to biometric templates, their original objective is to secure cryptographic keys using biometric features. That is, although biometric cryptosystems preserve users' privacy, they are not designed to provide revocability [11]. Since the main focus of our work is to deal with the privacy and revocability issues of biometrics-based authentication systems, our proposed method does not fall under this category.

On the other hand, the main objective of cancelable biometrics is to generate several revocable identifiers from a given biometric template. The concept of cancelable biometrics, firstly proposed by Ratha et al. [3], is based on the application of intentional, repeatable distortions of a biometric signal using a chosen one-way transformation function. In general, a practical cancelable biometric scheme should fulfil the following requirements: [11], [16]:

**Noninvertibility** It should be impossible, or at least computationally very hard, to obtain the original unprotected template from the protected one.

**Revocability** If a protected template is compromised, it should be possible to reissue a new protected template from the same original unprotected template and revoke the compromised one.

**Diversity** In order to prevent cross matching across databases, the cancelable biometrics scheme should be able to generate a large number of distinct protected templates from the same biometric signal so that different identities can be used for the same user in different applications.

**Accuracy** The cancelable biometrics scheme should not introduce significant degradation in the recognition performance of the unprotected biometric system.

Satisfying all of these requirements simultaneously is the main challenge for any cancelable biometrics scheme. Several methods that are based on cancelable biometrics concept have been proposed since its introduction. For fingerprint authentication, Ang et al. [17] proposed a key-dependent method for generating cancelable templates for fingerprints through applying a geometric transform to the extracted fingerprint features. However, this method cannot preserve the recognition performance in the transformed domain. Ratha et al. [18] proposed three different noninvertible transformations (Cartesian, radial and functional) for protecting fingerprint templates but the poor many-to-one property of these transforms makes it vulnerable to inversion attacks [19]. For face authentication, Savidges et al. [20] proposed a cancelable approach in which the face images are convolved with user-specific random kernels. However, an adversary can simply use deconvolution to recover the original template if she gained access to the random kernels. For personal authentication based on online signature, Maiorana et al. [21] proposed a non-invertible transform that protects on-line signatures signals by applying signal processing techniques to the acquired templates. However, this method introduces some degradation in the recognition accuracy and its renewability capacity is limited. For iris authentication, Jinyu et al. [22] proposed four non-invertible transforms for generating cancelable iris templates. However, the matching results of the proposed transforms drop significantly if the captured iris images are not of high quality. Moreover, similar to the methods in [17], [18], [20] and [21], the proposed transforms employ users' specific keys for deriving the protected templates from the original ones. The need for users' specific keys in such methods makes them vulnerable to the same issues associated with the tra-

ditional authentication systems since these keys can be lost, forgotten or stolen.

### 2.1 BioHashing

It has been claimed that cancelable biometrics techniques which are based on the idea of BioHashing have significant functional advantages over other cancelable biometrics schemes and even over unprotected biometrics systems with respect to the recognition accuracy. Strictly speaking, perfect or near to perfect recognition results have been reported for BioHashing-based techniques applied to many biometric characteristics [5]–[8]. Generally, BioHashing derives a compact binary vector $\mathbf{b} \in \{0,1\}^m$, called BioHash, from the true biometric feature vector $\mathbf{x} \in \Re^n$, where $m \leqslant n$, through an iterative computation of the inner products between a set of tokenized user-specific random numbers and the biometric vector $\mathbf{x}$. The base BioHashing algorithm may be summarized as follows:

1. Use a tokenized seed to generate a set of pseudorandom vectors, $\{r_i \in \Re^n | i = 1, \ldots, m\}$.
2. Transform the vectors generated in Step 1 into an orthonormal set of vectors $\{r_{\perp i} \in \Re^n | i = 1, \ldots, m\}$ using the Gram-Schmidt ortho-normalization process.
3. Compute the inner product between the biometric feature vector $\mathbf{x}$ and $r_{\perp i}$ $(\langle \mathbf{x} | r_{\perp i} \rangle), i = 1, \ldots, m$ and compute $b_i (i = 1, \ldots, m)$ as follows:

$$b_i = \begin{cases} 0 & \text{if } \langle \mathbf{x} | r_{\perp i} \rangle \leqslant \tau \\ 1 & \text{if } \langle \mathbf{x} | r_{\perp i} \rangle > \tau \end{cases}$$

   where $\tau$ is a predefined threshold.

It is important to note that BioHashing assumes that the tokenized seed is different among different users and different applications and that is why BioHashing integrates another independent factor, the token used to store the random seed, with the biometric features in the authentication process. Kong et al. [9] showed that employing the same seed among different users degrades the verification accuracy significantly. This implies that BioHashing can achieve its outstanding verification performance only under the impractical assumption that the employed tokens would never be lost, shared, or stolen. In other words, the claimed performance of BioHashing is not due to the discrimination power of the biometric features; rather, it is due to the variations found among different sets of random vectors generated using different seeds assigned to different users. Under this scenario, there would no need for combining biometric features with the generated random numbers since these numbers can serve as perfect passwords [9].

The issues introduced as a result of integrating other independent authentication factors, such as user-specific passwords and/or tokens or cards, with biometric features in current cancelable biometrics systems are the main motivation behind our proposed cancelable biometrics scheme. In BioEncoding, it is not required from the user to memorize long passwords or to keep tokens or cards in order to authenticate her identity. Rather, the user identity can be verified safely through matching her unique biometric features with the cancelable features stored centrally in the system.

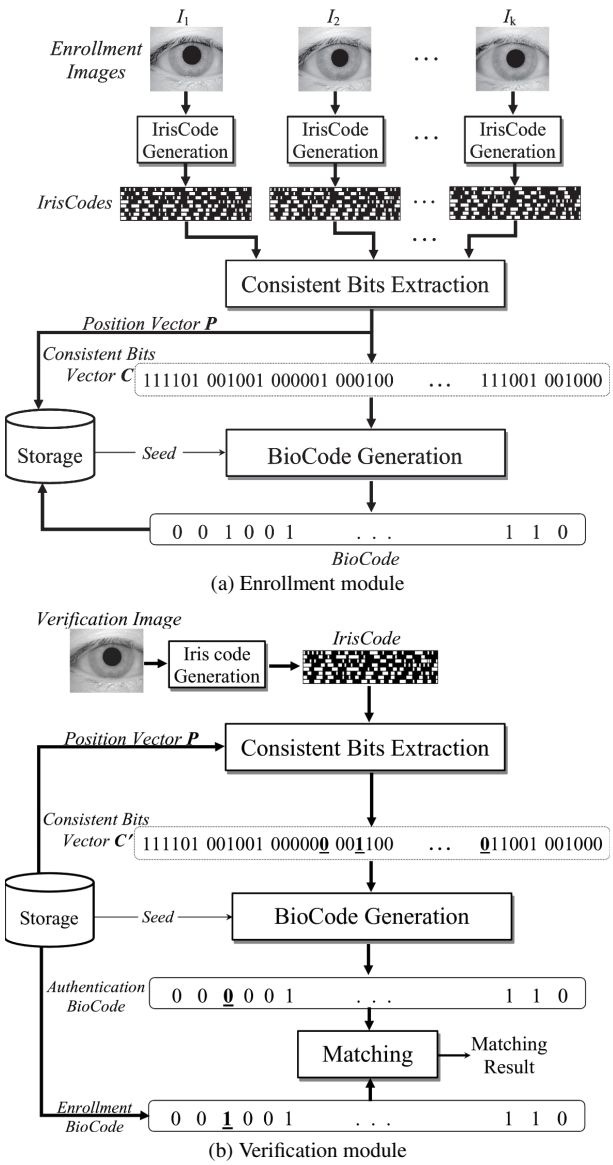
## 3. Proposed BioEncoding Scheme

Among other biometric modalities, iris is considered one of the most accurate and reliable characteristics that has been successfully implemented in many real-world applications with very low false acceptance and rejection rates [23]. The most accepted and widely used representation of iris biometric is the binary IrisCode presented by Daugman in [24]. IrisCodes are binary strings that represent the texture information found in iris patterns. Because our proposed template protection method is essentially a one-way transformation method for protecting binary representations of biometric characteristics, it can be applied directly to IrisCodes without the need for extracting binary strings from real-valued biometric templates.

In this section, we describe our proposed template protection scheme for IrisCodes. First, a general overview of the proposed method is provided, then the involved steps are explained in more detail.

### 3.1 BioEncoding Overview

The steps involved in the enrollment and verification modules of BioEncoding are shown in Fig. 1. The base procedure of BioEncoding for IrisCodes is comprised of three major stages: (a) IrisCode generation, (b) consistent bits extraction, and (c) BioCode generation. The three stages have to be conducted at both enrollment and verification modules. As illustrated in Fig. 1 (a), during enrollment phase, $k$ sample iris images are captured from the eye being enrolled. Since the quality of the acquired images may have a significant impact on the accuracy of the overall system, it should be ensured that the acquired $k$ images are of sufficient quality to support iris recognition. In order to determine whether an acquired image is suitable for use or not, iris image quality metrics [25] can be employed. The IrisCodes are generated for the captured images and collected in $k$ binary vectors. Then, the most consistent bits are extracted from the $k$ IrisCode vectors. The most consistent bits are the bits that have lower probability of flipping across several IrisCodes generated from several samples of the same iris compared to other bits [26]. The extracted consistent bits are collected in a consistent bit vector $C$ and their positions, in the true IrisCodes, are collected in a position vector $P$ and stored in a centralized storage. Finally, bits in $C$ are randomly encoded to another set of bits using a secret seed that can be kept in the centralized storage. The generated random bits constitute the protected BioCode which is stored in the centralized storage, instead of the original IrisCode that can be discarded safely at this point, in order to verify the user identity during the authentication phase. It is worth noting that although the generated BioCodes are stored in a centralized




(a) Enrollment module

(b) Verification module

**Fig. 1** Enrollment and verification modules of BioEncoding.

database, users' privacy is assured since it is not possible to obtain the original codes from the protected ones as will be discussed in a later section.

At verification, a live iris image is captured from the person being verified and the true IrisCode for the captured image is generated as shown in Fig. 1 (b). Using $P$, the most consistent bits are extracted from the generated IrisCode and collected in a consistent bit vector $C'$. Using the random seed stored in the centralized storage and $C'$, the BioCode for the user being authenticated is generated and matched with her stored BioCode. If the matching result exceeds a predefined threshold, the user is authenticated; otherwise authentication fails.

### 3.2 IrisCode Generation

Generally, a typical iris recognition system consists of three major modules: 1) eye image acquisition, 2) preprocessing, and 3) feature extraction and encoding. Although there are many approaches to iris recognition [27], the pioneering approach presented by Daugman [24] is still the most important. In Daugman's approach, texture information is encoded as a binary string called IrisCode, and matching is done using the normalized Hamming distance which is the fraction of bits that disagree in the matched IrisCodes. The normalized Hamming distance (HD) between two IrisCodes A and B is defined as:

$$HD(A,B) = \frac{1}{n}\sum_{i=1}^{n} A_i \oplus B_i \tag{1}$$

where $A_i$ and $B_i$ are the $i$th bits of the IrisCodes $A$ and $B$, repectively; $n$ is the length of IrisCodes and $\oplus$ is the XOR Boolean operator.

A captured iris image usually contains unwanted regions, such as eyelids, eyelashes and pupil, which need to be excluded before proceeding to subsequent steps. The goal of preprocessing is to segment the captured iris image in order to isolate the iris region and then normalize that region into a fixed dimension for further processing. In this work, the circular Hough transform is employed to find the pupillary and limbic boundaries [28]. The linear Hough transform is used to isolate eyelids by fitting a line to the upper and lower eyelids and then drawing horizontal lines which intersect with the top and bottom eyelid lines at the edge that is closest to the pupil. Finally, the eyelashes are isolated using a simple thresholding technique [29]. After extracting the iris region from the eye image, it is necessary to normalize the extracted region into fixed dimensions in order to account for dimensional inconsistencies between different captured iris images that result from varying imaging conditions such as illumination and imaging distance. The rubber sheet model presented by Daugman [30] is employed to normalize the extracted region into a fixed two dimensional array where the number of rows represents the angular resolution and the number of columns represents the radial resolution. In this paper, as illustrated in Fig. 2, the angular resolution is

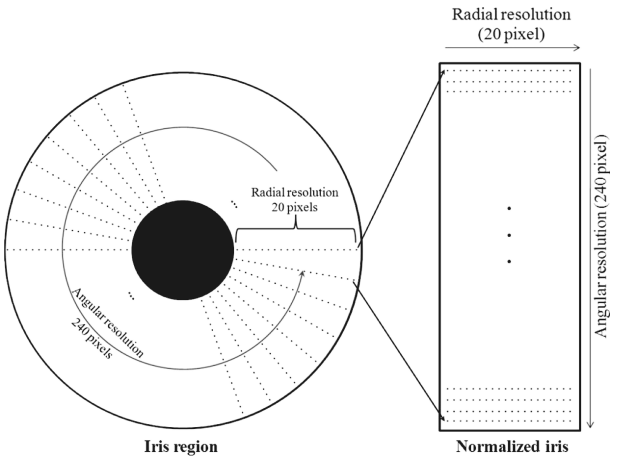



**Fig. 2** Illustration of iris region normalization.

set to 240 and the radial resolution is set to 20.

Feature extraction involves extracting the most discriminating iris features by analyzing the normalized iris texture. Iris texture information is extracted by treating each row in the normalized iris pattern as 1D signal and convolving it with 1D Log-Gabor wavelets [29]. For feature encoding, the output of filtering is phase quantized to four levels. Each filter produces two bits of data for each complex coefficient. The first bit is set to '1' if the real part of the coefficient is positive; otherwise it is set to '0', and the second bit is set to '1' if the imaginary part of the coefficient is positive; otherwise it is set to '0' [24]. Therefore, the total size of the generated IrisCode is 9600 (240x20x2) bits.

### 3.3 Consistent Bits Extraction

Once the true IrisCodes are generated from the enrollment images, the most consistent bits are extracted by identifying fragile bits and masking them out. Fragile bits are bits that have a substantial probability of being '0' in some images of an iris and '1' in other images of the same iris. Hollingsworth et al. [26] showed that masking out fragile bits and using only consistent bits could improve recognition performance significantly.

Before extracting the most consistent bits, the generated $k$ IrisCodes are aligned in order to account for rotational inconsistencies that may be caused due to head tilt during the acquisition of iris images. In order to achieve iris recognition with rotation invariance, Daugman [30] suggested comparing the gallery IrisCode to several different shifts of the probe IrisCode and then taking the shift that gives the smallest Hamming distance as the correct orientation of the probe image (rather than rotating the image itself using the Euler matrix). One angular step circular shift of an IrisCode is equivalent to rotating the original iris image by $360/r$ degrees, where $r$ is the angular resolution. As shown in Fig. 2, since the angular resolution is set to 240 in our work, shifting the IrisCode by a single angular step is equivalent to rotating the iris image by 1.5 degrees. Figure 3 shows an example of aligning two IrisCodes of size (2×6) using one shift in both directions, where the angular resolution is represented horizontally and the radial resolution is represented vertically, using the same technique. Here, we use the same procedure to align the $k$ IrisCodes generated during enrolment. First, the first IrisCode is set as a reference template. Then, the following steps are repeated for each of the remaining $k-1$ IrisCodes:

1. Compute the Hamming distance between the reference IrisCode and the IrisCode to be aligned with it.
2. Shift the IirsCode to be aligned with respect to the reference IrisCode by one angular step at a time eight times in both directions (this corresponds to shifting the original iris image 12 degrees in both directions step 1.5 degrees) and compute the Hamming distance between this IrisCode and the reference IrisCode after each shift.

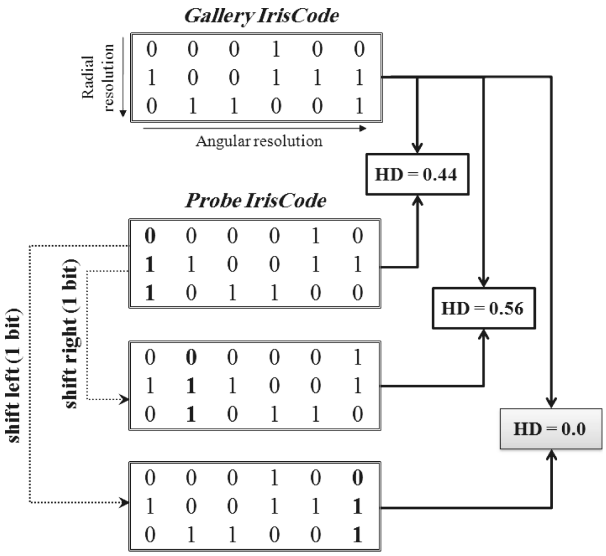



**Fig. 3** An example of aligning two IrisCodes.

3. Take the shift that corresponds to the smallest Hamming distance from the distances computed in Steps 1 and 2 as the correct orientation of the template under consideration.

After alignment, the corresponding bits in the $k$ iris vectors are summed up and only bits that give a sum of '0' or '$k$' are considered consistent and collected in a consistent bit vector $C$, and their positions are collected in a position vector $P$.

### 3.4 BioCode Generation

BioCode generation is the most important stage of BioEncoding. In this stage, the consistent bits of the true IrisCode are protected by mapping them to another set of random bits that comprise the compact BioCode. This random mapping process is illustrated in Fig. 4 and can be summarized in the following steps:

1. Group bits in the consistent bit vector $C$ into $n/m$ addressing words of length $m$ each, where $n$ is the number of bits in $C$ and the operator '/' denotes integer division with truncation of the result toward zero. For example, if the number of bits in a given consistent bit vector is 63 and the value of $m$ was set to 6, the number of addressing words would be 10.

2. Use a random seed to generate a pseudorandom sequence $S$ of length $2^m$. For example, as shown in Fig. 4, if $m = 6$, the length of $S$ would be 64. The seed is determined randomly by the verification system and stored in a secure centralized enrollment server. It is important to note that this seed need not be user-specific and therefore all the enrolled users can use the same seed (and hence the pseudorandom sequence). However, different seeds need to be generated for different applications to prevent cross-matching across databases.

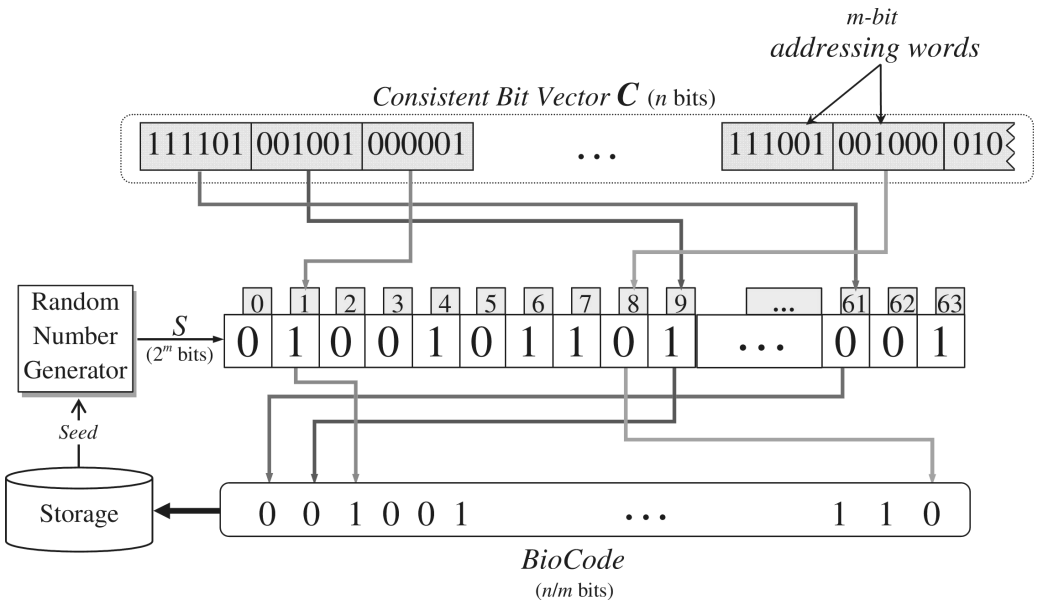



**Fig. 4** Illustration of BioCode generation process where $m = 6$.

3. Map each addressing word in $C$ to the bit value in $S$ whose position is addressed by the value in that word. For example, in Fig. 4, the address binary word $111101_b$ (which is equal to the decimal value 61) is mapped to 0 since the value at position 61 in $S$ is equal to 0, while the word $001001_b$ (decimal 9) is mapped to 1 since the value at position 9 in $S$ is equal to 1, and so on.

4. Constitute the protected BioCode from the set of $n/m$ addressed bits in $S$.

5. Store the BioCode in the centralized enrolment sever and discard the true (unprotected) IrisCode safely.

It is clear from the above procedure that the BioCode generation process is noninvertible since it is based on a many-to-one transformation. On average, half of the words in the true IrisCode are mapped to '1' while the remaining words are mapped to '0'. Thus, even if both the generated BioCode and the random sequence are known, it is impossible to determine which word in the true IrisCode addresses a given '0' (or '1') in the generated BioCode.

From the recognition perspective, since IrisCodes from different eyes are statistically independent [24], the corresponding BioCodes are ensured to be statistically independent as well regardless of whether IrisCodes from different eyes are employed to address different random sequences or even are used to address the same random sequence. In other words, the transformation process does not rely on the random sequence in the sense that different sequence should be assigned to each user. This advantage makes it possible to use sole IrisCodes in the authentication process without sacrificing user's privacy since there is no need to store user-specific data in a user-specific token. On the other hand, due to the similarity between IrisCodes that are generated from different samples of the same eye, the Hamming distances between BioCodes derived from these codes would be less than the Hamming distances between BioCodes derived from IrisCodes that are generated from samples of different eyes and therefore the separability between genuine and imposter distributions could be maintained. This implies that, unlike BioHashing, the random sequence need not to be user-specific. That is because if the same random sequence is used by many users, a unique BioCode would be generated for each user since corresponding words in different IrisCodes would address different bits in a shared random sequence.

Moreover, the cancelability of BioEncoding is straightforward since changing the random sequence would generate different BioCodes. That is, if the data stored in the centralized database is compromised, the compromised BioCodes can be revoked and replaced by other codes simply by changing the random sequence and re-enrolling the users.

## 4. Revocability and Noninvertibility of BioEncoding

As mentioned in Sec. 2, a practical cancelable biometrics technique has to meet a number of requirements. The advantage of BioEncoding over other template protection schemes is that it meets all the requirements of cancelable biometrics construct without introducing significant degradation to the recognition performance. In this section, we discuss the revocability, renewability capacity and noninvertibility properties of BioEncoding whereas the recognition accuracy is discussed in the next section.

### 4.1 Revocability and Renewability Capacity

It is straightforward to show that BioEncoding is revoca-

ble since a compromised BioCode can be canceled and replaced by another one simply by replacing the random seed value. Using a different seed value, a different (pseudo-) random sequence would be generated and the derived BioCode would be changed accordingly. The number of different random sequences that can be generated for each user relies on the length of the addressing words in the divided IrisCode. Since a word of length $m$ can address $2^m$ bits, there are $2^{(2^m)}$ different sequences that can be generated for IrisCodes divided into words of length $m$. For example, if $m = 4$, a compromised BioCode can be replaced by another one from $2^{16}$ candidates. However, it is important to note that there may be some correlation between some generated sequences. Therefore, in case a given random sequence is compromised and another random sequence has to be generated to replace the compromised one, the statistical independence between the two sequences should be tested. If they did not pass the test, another sequence should be generated until we get a sequence that passes the test.

### 4.2 Noninvertibility Analysis

Here we show formally the difficulty of obtaining the unprotected consistent bit vector (and hence the true IrisCode) from a protected BioCode even if both the BioCode and the random seed (and hence the random sequence) are compromised. From the security point of view, recovering the consistent bit vector of an IrisCode is as dangerous as recovering the IrisCode itself since it contains the most significant bits in that IrisCode. Therefore, in this section, the terms consistent bit vector and IrisCode are used interchangeably.

True-, as well as pseudo-, random binary sequences, which are often employed in cryptographic systems, should exhibit a number of statistical attributes [31]. One of these attributes is that the number of 1's, in such sequences, should be approximately equal to the number of 0's. Since the binary sequence $S$, which is used in the transformation process, is randomly generated, we can say, roughly, that the number of 1's in $S$ equals the number of 0's equals $2^{m-1}$, where $m$ is the length of any addressing word in an IrisCode. Therefore, as illustrated in Fig. 5, each bit in a BioCode may be originated from $2^{m-1}$ different addressing words. As a result, an attacker needs to perform $2^{l(m-1)}$ different trials in order to try all the possible combinations, where $l$ is the length of the BioCode. Since $l = n/m$, where $n$ is number of bits in an IrisCode, then the number of trials would be $2^{n(m-1)/m}$. For large $m$, the required number of trials would be $\approx 2^n$. This implies that recovering the true IrisCode from the protected BioCode is nearly as difficult as randomly guessing all bits in the original unprotected IrisCode, which is computationally infeasible.

On the other hand, since BioHashing can be considered as a quantized under-determined system of linear equations [7], the BioHashed biometric features could be inverted partially, if the token is compromised, via pseudo-inverse operation. However, while BioHashing prevents attacks via record multiplicity, in which an adversary uses multiple different pairs of protected templates and transformation random numbers to recover the true template, BioEncoding may be vulnerable to this type of attack. That is why it is very important in BioEncoding not to let users to specify the random seed by themselves. If users were allowed to specify the random seed, which is used in the transformation process freely, an adversary could obtain many pairs of BioCodes and random sequences and hence would be able to narrow down the candidates of the IrisCodes.

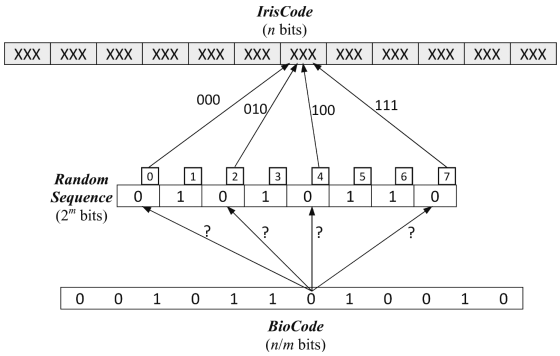



**Fig. 5** An example of *BioCode* inversion where $m = 3$.

## 5. Experimental Results

In order to evaluate the performance of BioEncoding and compare it with that of BioHashing [7], several experiments have been conducted using the publicly available iris image database collected by the Chinese Academy of Science – Institute of Automation, CASIA v1.0 [32]. The main reason for choosing this dataset is to make a fair comparison between our proposed method and the BioHashing-based method in [7] since their reported results are based on the same dataset. This dataset contains 756 iris images from 108 unique eyes, with 7 different samples from each eye. All the images are 8-bit grayscale images with a resolution of 320×280 pixels. Seven IrisCodes were generated for each eye. From these codes, six different codes were used to extract the most consistent bits for that eye. By changing the codes that are used to extract the consistent bits, seven different consistent bit vectors were obtained for each eye. We found that, on average, 20.46% of the bits in the generated IrisCodes were perfectly consistent; that is, the average number of bits that were always equal to the same binary value across all IrisCodes for an iris is 1964 bits. The iris with the smallest fraction had 332 bits (3.46%) that are perfectly consistent and the iris with the highest fraction has 3737 perfectly consistent bits (38.93%). That means, on average, the generated BioCodes would be of length $1964/m$. This length is long enough for $m \leq 6$, both from the security point of view and with respect to the maximum BioHash dimension tested in [7].

It is important to note that the reason for obtaining some too short consistent bit vectors, which could make the task of inverting the generated BioCodes simpler, is due to the employment of all the images in the adopted dataset, for

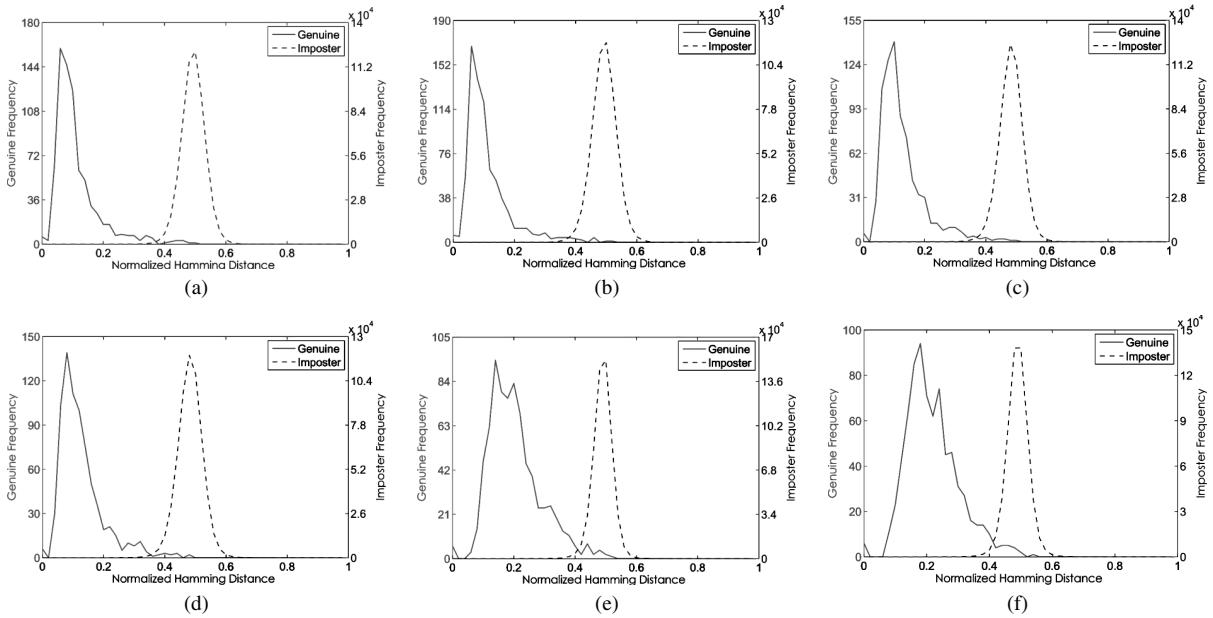



**Fig. 6** The genuine and imposter distributions for (a) true IrisCodes, BioCodes generated using word lengths of (b) 2, (c) 3, (d) 4, (e) 5 and (f) 6 bits.

identifying the consistent bits during enrolment, regardless of their quality. Fortunately, this is not the case in practical systems in which only images that have sufficient quality to support recognition are used on enrollment.

For inter-class comparisons, every consistent bit vector for each eye is considered as the enrollment IrisCode and matched against the corresponding bit vectors that are extracted from the same positions in other IrisCodes generated from all the eye-samples captured from other eyes, yielding 566244 different imposter comparisons. That is, if a consistent bit vector, extracted from some given eye, contains $n$ consistent bits where the positions of these consistent bits are stored in the position vector for that eye, matching is carried out between these bits and the $n$ bits found in the same positions in other IrisCodes. This implies that the number of the compared bits is different from eye to eye and relies on the number of consistent bits found in each eye. For intra-class comparisons, each iris image is matched against the consistent bit vector extracted from other images of the same iris, leading to 7 genuine comparisons for each eye and a total of 756 genuine comparisons.

At first, consistent vectors of the original (unprotected) IrisCodes were matched against each other before applying any transformation. Then, two sets of experiments were conducted. In the first set, we applied BioEncoding with different addressing word lengths, $m$ = 2, 3, 4, 5, and 6 bits; and matching between the BioCodes derived after each experiment was performed. Since our interest is in performing authentication using sole IrisCodes and not relying on user-specific random numbers, the same random sequence $S$ was used with all users. Figure 6 shows the imposter and genuine distributions for true IrisCodes and for BioCodes generated using different word lengths.

We used the decidability metric $d'$ in order to investigate the impact of applying BioEncoding on the recognition accuracy. The decidability metric measures the separability between genuine and imposter distributions in terms of the difference between their means normalized by some function of their standard deviations and is given by [30]:

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

where $\mu_i$ and $\mu_g$ are the means and $\sigma_i^2$ and $\sigma_g^2$ are the variances of the imposter and genuine distributions, respectively. High decidability values imply better separation between genuine and imposter distributions, which allows for more accurate recognition.

**Table 1** Means, variances and decidability values of imposter and genuine distributions for BioCodes of different word lengths.

| Word length ($m$) | $\mu_i$ | $\sigma_i^2$ | $\mu_g$ | $\sigma_g^2$ | $d'$ |
|---|---|---|---|---|---|
| True IrisCodes | 0.4931 | 0.0015 | 0.1123 | 0.0057 | 6.35 |
| 2 | 0.4930 | 0.0016 | 0.1122 | 0.0059 | 6.24 |
| 3 | 0.4820 | 0.0014 | 0.1259 | 0.0055 | 6.04 |
| 4 | 0.4819 | 0.0016 | 0.1258 | 0.0054 | 6.04 |
| 5 | 0.4855 | 0.0010 | 0.1869 | 0.0069 | 4.74 |
| 6 | 0.4809 | 0.0011 | 0.2062 | 0.0071 | 4.45 |

Table 1 shows the means, variances and decidability values of genuine and imposter distributions for true IrisCodes and for the protected BioCodes generated using the mentioned $m$ values. The results in Fig. 6 and Table 1 show that the recognition performance is preserved for $m$ = 2, 3 and 4, while slightly degraded for $m$ = 5 and 6. This is because as $m$ increases, the probability of a match between corresponding addressing words decreases and hence the match probability between the corresponding addressed bits in the compared BioCodes decreases conse-

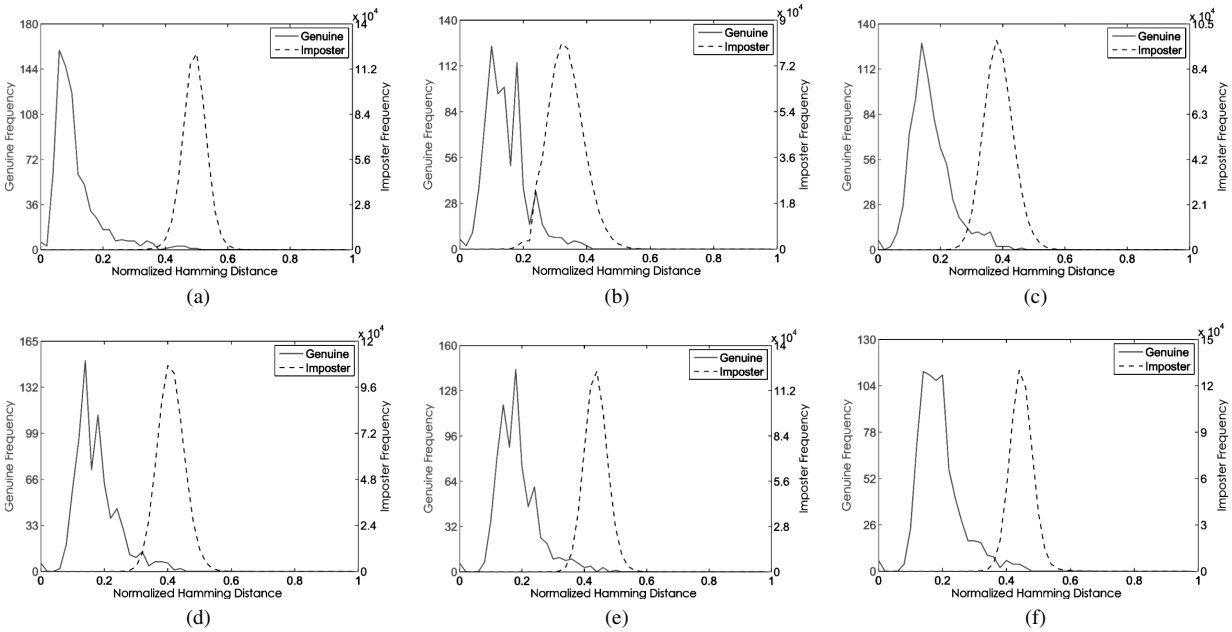



**Fig. 7** The genuine and imposter distributions for (a) true IrisCodes and BioHashes of dimensions, (b) 100, (c) 150, (d) 200, (e) 300 and (f) 350 bits.

quently. This implies that while the recognition accuracy can be enhanced as a result of decreasing the length of address words ($m \leq 4$), larger values of $m$ ($\geq 4$), on the other hand, increases the renewability capacity and strengthens the noninvertibilty property of BioCodes. Among the tested $m$ values, address words of length 4 can provide an appropriate compromise between the requirements of renewability and noninvertibility on the one side and the requirement of accuracy preservation on the other side. That is because using address words of length 4, a large number of random sequences could be employed at the expense of only a slight (insignificant) degradation in recognition accuracy. Generally speaking, small $m$ values should be employed in applications that require high recognition accuracy while large $m$ values are preferred in applications where high security is required.

On the other hand, in the second set of experiments, we simulated BioHashing using the same BioHash dimensions, $m$ = 100, 150, 200, 300, and 350 bits; adopted in [7]. However, rather than using a unique seed among different irises, as performed in [7], the same random numbers were used for all users, as done with BioEncoding, since it should be assumed that imposters have an access to the the user-specific token in a more realistic scenario.

Figure 7 shows the imposter and genuine distributions for true IrisCodes and for each tested dimensions of BioHash and Table 2 shows the means, variances and decidability values of these distributions. In contrast with the results obtained in [7], it is clear from the results shown here that BioHashing degrades the recognition accuracy achieved using unprotected iris recognition system significantly. These results conform with the analysis presented in [9] and prove that the perfect results obtained in [7] are based on the *impractical* assumption that an imposter cannot gain access to the random number of a genuine user.

**Table 2** Means, variances and decidability values of imposter and genuine ditsributions for different dimensions of BioHash.

| Dimension ($m$) | $\mu_i$ | $\sigma_i^2$ | $\mu_g$ | $\sigma_g^2$ | $d'$ |
|---|---|---|---|---|---|
| True IrisCodes | 0.4931 | 0.0015 | 0.1123 | 0.0057 | 6.35 |
| 100 | 0.3393 | 0.0032 | 0.1488 | 0.0041 | 3.18 |
| 150 | 0.3870 | 0.0022 | 0.1700 | 0.0046 | 3.74 |
| 200 | 0.4287 | 0.0015 | 0.1823 | 0.0046 | 4.41 |
| 300 | 0.4391 | 0.0013 | 0.1864 | 0.0045 | 4.68 |
| 350 | 0.4471 | 0.0012 | 0.1903 | 0.0046 | 4.76 |

However, it can be noticed from the results shown in Fig. 7 and Table 2 that when the BioHash dimension increases, the recognition accuracy increases consequently. Therefore, we tested the accuracy using higher BioHash dimensions, $m$ = 400, 450, 500, 550, and 600 bits. We found from the obtained decidability values of all the tested dimensions, shown in Fig. 8, that the recognition accuracy starts to fall down for $m > 350$.

From the obtained results, we can conclude that BioHashing introduces a noticeable degradation in recognition accuracy where the best accuracy was obtained when $m$ (BioHash dimension) = 350. Whereas, for BioEncoding, the recognition accuracy is preserved for $m$ (word length) $\leq$ 4. For ease of comparison, Figure 9 shows the ROC curves for different BioHash dimensions and Figure 10 shows the ROC curves for BioEncoding, for the tested $m$ values. The corresponding equal error rates (EER) are listed in Table 3 for BioHashes and in Table 4 for BioCodes. The obtained results show clearly the efficiency of our proposed cancelable biometrics scheme compared to BioHashing.

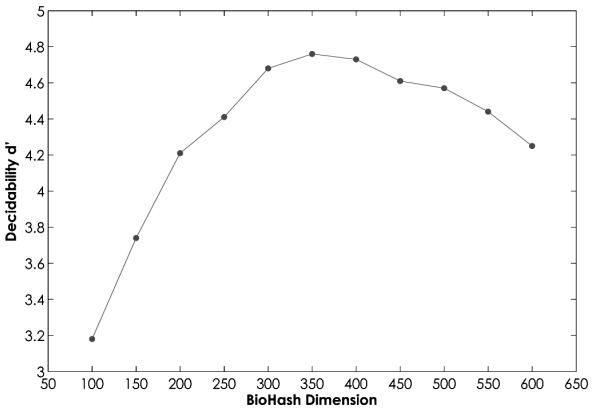



**Fig. 8** The decidability values corresponding to the tested BioHash Dimensions.

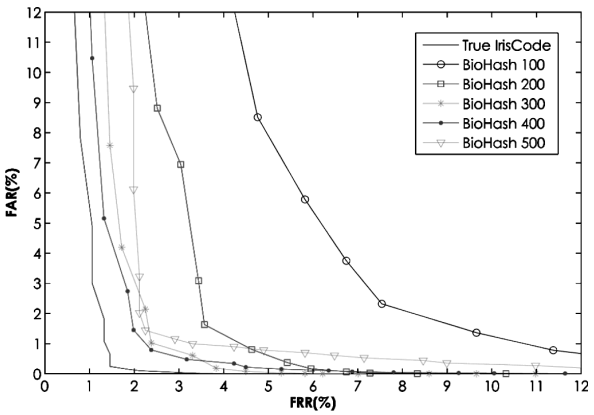



**Fig. 9** ROC curves for BioHash dimensions adopted in [7].

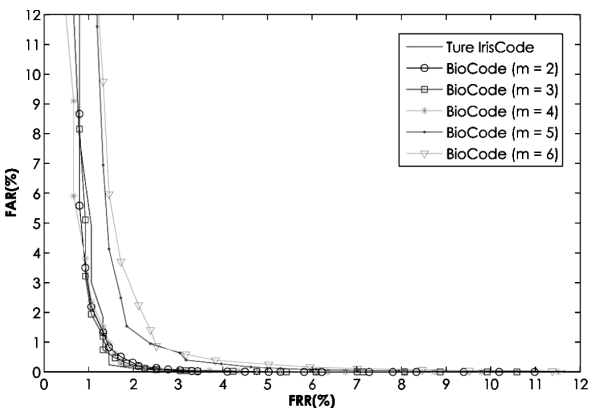



**Fig. 10** ROC curves for BioCodes generated using the tested word lengths.

**Table 3** EER values for BioHashes generated using different dimensions.

| BioHash Dimension ($m$) | ERR |
|---|---|
| True IrisCodes | 1.3 |
| 100 | 5.8 |
| 150 | 4.8 |
| 200 | 3.4 |
| 300 | 2.2 |
| 350 | 2.2 |

**Table 4** EER values for BioCodes generated using different $m$ values.

| Word length ($m$) | ERR |
|---|---|
| True IrisCodes | 1.3 |
| 2 | 1.3 |
| 3 | 1.3 |
| 4 | 1.4 |
| 5 | 1.8 |
| 6 | 2.1 |

## 6. Conclusion

In this paper, we have presented an effective cancelable biometrics scheme that is based on random sequence addressing for protecting IrisCodes. The proposed scheme satisfies the requirements of the cancelable biometrics construct without sacrificing the recognition accuracy and hence allows for enhanced security and privacy in iris recognition systems. The diversity and revocability properties of the proposed scheme have been discussed and its non-invertibility has been analyzed. Experiments using CASIA dataset have been conducted in order to compare the impact of applying the proposed scheme on the recognition accuracy with that of BioHashing techniques in the stolen-case scenario. Experimental results showed the superiority of our scheme compared to the true performance of BioHashing and it has been shown that the recognition accuracy of the proposed method has not been affected by employing the same random sequence with all users and therefore it can be adopted without using any tokens for storing user-specific information as required in the case of BioHashing techniques.

A promising area for future research is the application of the proposed scheme to the binary representations of other biometrics characteristics, such as fingerprint and face.

### References

[1] S. Nanavati, M. Thieme, and R. Nanavati, Biometrics: Identity Verification in a Networked World, John Wiley & Sons, 2002.
[2] A. Cavoukian and A. Stoianov, "Biometric encryption: A positive-sum technology that achieves strong authentication, security and privacy," Technical Report, Office of the Information and Privacy Commissioner of Ontario, Toronto, Ontario, Canada, March 2007.
[3] N. Ratha, J. Connell, and R. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," IBM Syst. J., vol.40, no.3, pp.614–634, 2001.
[4] R. Bolle, J. Connell, and N. Ratha, "Biometrics perils and patches," Pattern Recognit., vol.35, no.12, pp.2727–2738, Dec. 2002.
[5] A.B.J. Teoh, D.C.L. Ngo, and A. Goh, "BioHashing: Two factor authentication featuring fingerprint data and tokenised random number," Pattern Recognit., vol.37, no.11, pp.2245–2255, Nov. 2004.
[6] A. Goh and D. Ngo, "Computation of cryptographic keys from face biometrics," Lect. Notes Comput. Sci., vol.2828, pp.1–13, Springer, 2003.
[7] C. Chin, A. Teoh, and D. Ngo, "High security iris verification system based on random secret integration," Computer Vision and Image Understanding, vol.102, no.2, pp.169–177, May 2006.
[8] T. Connie, A. Teoh, M. Goh, and D. Ngo, "PalmHashing: A novel

approach for cancelable biometrics," Inf. Proces. Lett., vol.93, no.1, pp.1–5, Jan. 2005.
[9] A. Kong, K.-H. Cheung, D. Zhang, M. Kamel, and J. You, "An analysis of BioHashing and its variants," Pattern Recognit., vol.39, no.7, pp.1359–1368, July 2006.
[10] K.-H. Cheung, A. Kong, D. Zhang, M. Kamel, J. You, and H. -W. Lam, "An analysis on accuracy of cancelable biometrics based on bioHashing," Lect. Notes Comput. Sci., vol.3683, pp.1168–1172, Springer, 2005.
[11] A.K. Jain, K. Nandakumar, and A. Nagar, "Biometric template security," EURASIP J. Advances in Signal Processing, Jan. 2008.
[12] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," Eurocrypt 2004, Lect. Notes Comput. Sci., vol.3027, pp.523–540, 2004.
[13] E.C. Chang and S. Roy, "Robust extraction of secret bits from minutiae," Proc. 2nd International Conference on Biometrics, pp.750–759, Aug. 2007.
[14] A. Jules and M. Sudan, "A Fuzzy vault scheme," Proc. IEEE International Symposium on Information Theory, p.408, 2002.
[15] F. Hao, R. Anderson, and J. Daugman, "Combining crypto with biometrics effectively," IEEE Trans. Comput., vol.55, no.9, pp.1081–1088, Sept. 2006.
[16] D. Maltoni, D. Maio, A.K. Jain, and S. Prabhakar, Handbook of Fingerprint Recognition, Springer Publishing Company, 2009.
[17] R. Ang, R. Safavi-Naini, and L. McAven, "Cancelable key-based fingerprint templates," Proc. 10th Australian Conference on Information security and Privacy, pp.242–252, July 2005.
[18] N.K. Ratha, S. Chikkerur, J.H. Connell, and R. Bolle, "Generating cancelable fingerprint templates," IEEE Trans. Pattern Anal. Mach. Intell., vol.29, no.4, pp.561–572, 2007.
[19] F. Quan, S. Fei, C. Anni, and Z. Feifei, "Cracking cancelable fingerprint template of ratha," International Symposium on Computer Science and Computational Technology, pp.572–575, 2008.
[20] M. Savvides, B.V.K. Vijaya Kumar, and P.K. Khosla, "Cancelable biometric filters for face recognition," Proc. 17th International Conference on Pattern Recognition (ICPR 2004), pp.922–925, Aug. 2004.
[21] E. Maiorana, M. Martinez-Diaz, P. Campisi, J. Ortega-Garcia, and A. Neri, "Template protection for HMM-based on-line signature authentication," IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008), Workshop on Biometrics, June 2008.
[22] Z. Jinyu, N. Ratha, and J. Connell, "Cancelable iris biometric," Proc. 19th International IAPR Conference on Pattern Recognition (ICPR 2008), pp.1–4, 2008.
[23] J. Daugman, "Probing the uniqueness and randomness of iris codes: Results from 200 billion iris pair comparisons," Proc. IEEE, vol.94, no.11, pp.1927–1935, Nov. 2006.
[24] J. Daugman, "High confidence visual recognition of persons by a test of statistical independence," IEEE Trans. Pattern Anal. Mach. Intell., vol.15, no.11, pp.1148–1161, 1993.
[25] N. Kalka, J. Zuo, N. Schmid, and B. Cukic, "Image quality assessment for iris biometric," Proc. SPIE Conference on Biometric Technology for Human Identification III, vol.6202, pp.6202:D1D11, April 2006.
[26] K. Hollingsworth, K. Bowyer, and P. Flynn, "The best bits in an iris code," IEEE Trans. Pattern Anal. and Machi. Intell., vol.31, no.6, pp.964–973, 2009.
[27] K. Bowyer, K. Hollingsworth, and P. Flynn, "Image understanding for iris biometrics: A survey," Computer Vision and Image Understanding, vol.110, no.2, pp.281–307, May 2008.
[28] R. Wildes, "Iris recognition: An emerging biometric technology," Proc. IEEE, vol.85, no.9, pp.1348–1363, Sept. 1997.
[29] L. Masek, "Recognition of human iris patterns for biometric identification," Bachelor's Thesis, University of Western Australia, 2003.
[30] J. Daugman, "How iris recognition works," IEEE Trans. Circuits Syst. Video Technol., vol.14, no.1, pp.21–30, 2004.
[31] A. Menezes, P. van Oorschot, and S. Vanstone, Handbook of Applied Cryptography, CRC Press, 1996.
[32] CASIA iris image database, Available: http://www.cbsr.ia.ac.cn/Dat-abases.htm

**Osama Ouda** received the B.Sc. from the faculty of Computers and Information Sciences, Mansoura University, Egypt in 2000 and the M.Sc. from the faculty of Computers and Information Sciences, Ain Shams University, Egypt in 2007. He is currently pursuing the doctoral degree at Chiba University, Chiba, Japan. His scientific interests include information security, biometric encryption and pattern recognition.

**Norimichi Tsumura** received the B.E., M.E. and D.E. in applied physics from Osaka University in 1990, 1992 and 1995, respectively. He moved to the Department of Information and Image Sciences, Chiba University in April 1995, as assistant professor. He is currently associate professor since 2002, and also researcher at PREST, Japan Science and Technology Corporation (JST). He was visiting scientist in University of Rochester from March 1999 to January 2000. He got the Optics Prize for Young Scientists (The Optical Society of Japan) in 1995, and Applied Optics Prize for the excellent research and presentation (The Japan Society of Applied Optics) in 2000. He received the Charles E. Ives award in 2002 from the IS&T. He is interested in the color image processing, computer vision, computer graphics and biomedical optics.

**Toshiya Nakaguchi** received the B.E., M.E., and Ph.D. degrees from Sophia University, Tokyo, Japan in 1998, 2000, and 2003, respectively. He was a research fellow supported by Japan Society for the Promotion of Science from April 2001 to March 2003. From 2006 to 2007, he was a research fellow in Center of Excellence in Visceral Biomechanics and Pain, in Aalborg Denmark, supported by CIRIUS, Danish Ministry of Education from 2006 to 2007. Currently, he is an Assistant Professor of imaging science at the Graduate School of Advanced Integration Science, Chiba University, Chiba Japan. His current research interests include the computer assisted surgery and medical training, medical image analysis, real-time image processing, and image quality evaluation.