Tokenless Cancelable Biometrics Scheme for Protecting IrisCodes

Osama Ouda, Norimichi Tsumura and Toshiya Nakaguchi
Graduate School of Advanced Integration Science, Chiba University, Chiba 263-8522, Japan
oouda@graduate.chiba-u.jp, tsumura@faculty.chiba-u.jp, nakaguchi@faculty.chiba-u.jp

Abstract

In order to satisfy the requirements of the cancelable biometrics construct, cancelable biometrics techniques rely on other authentication factors such as password keys and/or user-specific tokens in the transformation process. However, such multi-factor authentication techniques suffer from the same issues associated with traditional knowledge-based and token-based authentication systems. This paper presents a new one-factor cancelable biometrics scheme for protecting IrisCodes. The proposed method is based solely on IrisCodes; however, it satisfies the requirements of revocability, diversity and noninvertibility without deteriorating the recognition performance. Moreover, the transformation process is easy to implement and can be integrated simply with current iris matching systems. The impact of the proposed transformation process on the recognition accuracy is discussed and its noninvertibility is analyzed. The effectiveness of the proposed method is confirmed experimentally using CASIA-IrisV3-Interval dataset.

1. Introduction

Unlike passwords and tokens, biometric templates cannot be canceled and replaced easily if they are compromised (stolen by an imposter). In addition, many privacy concerns have been raised regarding using biometrics in human verification/identification. As a result, the concept of cancelable biometrics [9] has been introduced as a means for generating multiple protected biometric templates through the application of an intentional and repeatable transformation to biometrics signals. A cancelable transform should be noninvertible, revocable, able to generate a large number of distinct protected templates, and should not introduce significant degradation to the recognition accuracy[8].

In order to satisfy the above-mentioned requirements, cancelable biometrics systems that have been proposed so far, such as methods described in [2],[4],[6],[8], make use of other user-specific independent authentication factors. Since the employed keys and/or tokens still can be shared, lost, or stolen, cancelable biometrics schemes that rely on such factors suffer from the same issues associated with traditional knowledge-based and token-based authentication systems.

In this paper, we present a novel cancelable biometrics scheme for protecting IrisCodes. In contrast with other existing methods, the feature transformation of the proposed method is non-invertible even if it is known to an attacker. Moreover, the recognition accuracy is preserved even if the same key is employed with all users. That is, the transformation key need not be kept unique nor secret.

2. Proposed method

The proposed method consists of three stages: (1) IrisCode generation, (2) consistent bits extraction and (3) cancelable code (BioCode) generation. As shown in Fig. 1, at enrolment, k sample iris images are captured...
from the eye being enrolled and IrisCodes for the captured images are generated and collected in $k$ binary vectors. Then, the most consistent bits are extracted from the $k$ vectors and collected in a consistent bit vector $C$; and their positions, in the true IrisCodes, are collected in a position vector $P$. The most consistent bits are bits that have lower probability of flipping across several IrisCodes generated from several samples of the same iris compared to other bits [5]. In this paper, a bit is considered consistent if it does not change its value across the $k$ binary vectors. Then, bits in $C$ are randomly mapped to another set of bits to constitute the protected BioCode as described in Section 2.1. Finally, the position vector and the protected BioCode are stored in a centralized storage rather than the original IrisCode which can be discarded safely.

At verification, as shown in Fig. 2, a single iris image is captured from the eye being verified and the true IrisCode is generated for it. Using position indices stored in $P$, the most consistent bits are extracted from the generated IrisCode and collected in a consistent bit vector $C'$. Finally, the BioCode for the eye being authenticated is derived from $C'$ and matched against the stored BioCode.

2.1. BioCode generation

The steps involved in the random encoding process, referred to as BioEncoding, that encodes consistent bits extracted from the true IrisCodes into the protected BioCode are shown in Fig. 3. These steps can be summarized as follows:

1. Group bits in the consistent bit vector $C$ into address words of fixed length $m$.
2. Use a random seed, stored in the centralized storage, to generate a (pseudo-) random sequence $S$ of length $2^m$.
3. Map each word in $C$ to a single bit value in $S$ whose location is addressed by the value in that word.
4. Constitute the protected BioCode from the set of $n/m$ addressed bits, where $n$ is the length of $C$.
5. Store the BioCode in the centralized storage and discard the original (unprotected) IrisCode.

The random sequence $S$ need not be user-specific; rather, it can be made public since different IrisCodes would produce totally different BioCodes even if the same sequence is used. Moreover, the revocability of the proposed transformation process is straightforward since changing the random sequence $S$ would generate different BioCodes for all users. 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. However, it is worth noting that although there are $2^{2m}$ different sequences that can be addressed using $m$-bit address words, some of these sequences are not suitable for our scheme. For example, using the sequence $S_1$, shown in Table 1, would produce BioCodes that consist of only zeros, and using the sequence $S_2$ would degrade the discriminability of the system since totally different word pairs such as $(w_0: 000, w_7: 111)$ and $(w_1: 001, w_6: 110)$ address similar bit values, ‘1’ and ‘0’ respectively; that is, similar BioCodes would be generated from different IrisCodes. On the other hand, using sequences such as $S_3$ and $S_4$ would ensure accuracy preservation as well as noninvertibility of BioCodes.

Therefore, a generated sequence $S$ should be tested before use in order to check whether it meets the following two conditions: (1) the number of 0’s in $S$ should equal (or approximately equal) to the number of 1’s, and (2) totally (or mostly) different address words, such as ‘1011’ and ‘0100’, should address different bit values in $S$. Although fulfilling these conditions may restrict the renewability capacity of the proposed scheme, they have a significant role in preserving the discrim-
For address words of size \( m \), there are \( 2^m(2^m-1)/2 \) different pairs of address words. The Hamming distances between words in these pairs range from 1 to \( m \). Table 2 shows all the 28 word pairs for \( m = 3 \) and the Hamming distances between words in these pairs. In general, the Hamming distances between address words that address similar bits in a given sequence \( S \) should be small as much as possible. Therefore, in order to decide whether a given sequence fulfills the above-mentioned conditions or not, we calculate the sum of the Hamming distances between the address words in the pairs that correspond to every two similar bits in \( S \) as follows:

\[
f(S) = \sum_{i=0}^{2^m-2} \sum_{j=i+1}^{2^m-1} [S(i) \oplus S(j)] \cdot HD(w_i, w_j)
\]

where \( m \) is the length of any address word, \( S(i) \) and \( S(j) \) are the \( i \)th and \( j \)th bits in \( S \), \( w_i \) and \( w_j \) are their corresponding address words, respectively; \( \oplus \) is the XNOR Boolean operator which equals 1 if and only if the two bits \( S(i) \) and \( S(j) \) are similar, and HD stands for the Hamming distance (the number of disagreeing bits) between the address words \( w_i \) and \( w_j \).

The \( f \) values for the sequences shown in Table 1 indicate that sequences that have small \( f \) values, such as \( S_3 \) and \( S_4 \) are sequences that satisfy the two conditions required for ensuring noninvertibility and recognition accuracy preservation. On the other hand, sequences with relatively large \( f \) values, such as \( S_1 \) and \( S_2 \), should be avoided since they may decrease the discriminability between the generated BioCodes.

### 3. Noninvertibility analysis

As stated in condition (1), the number of 0’s in \( S \) should be (approximately) equal to the number of 1’s equal to \( 2^{m-1} \), where \( m \) is the length of any address word in an IrisCode. Therefore, as shown in Fig.4, each bit in a protected BioCode may be originated from

![Figure 4: Example of BioCode inversion where \( m = 3 \). 2\(^{m-1} \) different address words. As a result, an attacker needs to perform \( 2^{(m-1)} \) different trials in order to try all the possible combinations, where \( l \) is the length of the BioCode. Since \( l = nm \), 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 even if both the generated BioCode and the random sequence \( S \) are known.

### 4. Experimental results

In order to testify the effectiveness of our proposed scheme with respect to the recognition accuracy, we conducted several experiments using the publicly available iris database CASIA-IrisV3-Interval [1]. This database contains 2655 images captured from 396 different classes (eyes). However, most of these classes have few number of images in the dataset. Therefore, a subset contains 790 images of 79 different classes, classes that have at least 10 images in the database, was used. For each class, six images were employed to

![Table 3: Selected sequences for different word lengths](Image 50x40 to 350x60)

The results presented in the Table 3 show that our scheme guarantees noninvertibility and recognition accuracy preservation.
extract the consistent bit vector \( C \) used in enrollment, while the remaining four images were used in verification. To make our work reproducible, we used the open-source MATLAB implementation provided by Masek and Kovesi [7] to generate the true IrisCodes. Then, we applied the described transformation method in order to generate the cancelable BioCodes using different word lengths, \( m = 2, 3 \) and 4 bits; and matching between BioCodes derived after each experiment was performed. Since our interest is in performing authentication using sole IrisCodes and not relying on other authentication factors, the same (public) random sequence \( S \) was used with all classes. We calculated the \( f \) values for all the sequences that can be addressed by the tested \( m \) values, then we selected sequences that have small \( f \) values as shown in Table 3. We ran our simulation for all the selected sequences (4, 14 and 232 times for \( m = 2, 3 \) and 4 respectively) and the results are averaged to reduce the statistical fluctuations. We measured the impact of the proposed transformation process on the separability between the genuine and imposter distributions using the decidability metric [3]:

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

where \( \mu_i \) and \( \mu_g \) are the means and \( \sigma_i^2 \) and \( \sigma_g^2 \) are the variances of the impostor and genuine distributions, respectively. Table 4 shows the means, variances and decidability values of genuine and impostor distributions. Results in Table 4 show that the statistics of the genuine and impostor distributions have been very slightly affected by the transformation process. Table 5 shows the equal error rate (ERR) values and false reject rate (FRR) values when false accept rate (FAR) = 0, and Fig. 5 shows the ROC curves for both IrisCodes (\( m = 1 \)) and BioCodes generated using the tested \( m \) values. Results shown in Table 5 and Fig. 5 show no or little performance degradation compared to the unprotected iris recognition system.

Table 5: EER, and FRR (when FAR = 0) values of BioCodes

<table>
<thead>
<tr>
<th>Word length(( m ))</th>
<th>EER</th>
<th>FRR (FAR = 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.30</td>
<td>6.96</td>
</tr>
<tr>
<td>2</td>
<td>2.31</td>
<td>6.96</td>
</tr>
<tr>
<td>3</td>
<td>2.44</td>
<td>6.96</td>
</tr>
<tr>
<td>4</td>
<td>2.43</td>
<td>6.96</td>
</tr>
</tbody>
</table>

5. Conclusion

In this paper, we have proposed a novel cancelable biometrics scheme for protecting IrisCodes. Unlike current cancelable biometrics techniques, the proposed scheme use sole IrisCodes in the authentication process and does not rely on other independent authentication factors yet satisfies all the requirements of the cancelable biometrics construct. The preliminary experimental results showed that our transformation method introduces no or little performance degradation compared to a conventional iris recognition scheme.

Analysing the security of the proposed method rigorously, considering more intelligent attacks such as hill climbing attack, is the main goal of our future research.

References