Binarized Revocable Biometrics in Face Verification

Pang Ying Han and Andrew Teoh Beng Jin

***Abstract*—A two-factor revocable authentication approach, combining user-specific pseudo-random bit sequence with biometrics is presented. Through the mixing process, a distinct binary code, coined as BioBit, is formed. This provides a protection layer against biometrics fabrication because the system can cancel the compromised and reacquire a new one via bits replacement. BioBit delivers lower error rate as compared to sole biometrics when the genuine token is used by the authorized user. There are two identity theft scenarios may be raised: 1. stolen-biometrics: an imposter possesses intercepted biometrics data to be considered genuine; 2. stolen-token: an imposter has access genuine token and used it to claim himself as the genuine user. BioBit scheme shows the impressive performance (EER=0.001% and EER=0.002% tested on ORL and FERET datasets) in case 1. For the stolen-token case, this approach attains EER=1.28% and EER=1.36% on ORL and FERET datasets.**

***Index Terms*—Face recognition, Cancelable biometrics, Two-factor authentication, XOR operation.**

I. Introduction

Password- and token-based authentications are the most common forms of personal recognition today. However, these approaches are insufficient in contemporary security requirements. Passwords can be forgotten and shared. They are even easy to guess based on social engineering methods or broken by dictionary attacks. Besides, most users have the tendency to use the same password for different applications [1]. If the password is compromised, it may open many doors. On the other hand, the shortcomings for token are that it can be lost, stolen or mislaid (for example, left at home). In addition, passwords and tokens do not offer defense against repudiation [2]. User can repudiate his/ her access by claiming that the password or token is shared with a friend. Thus, there is no way to determine who the actual user is.

Biometric characteristics, such as face, fingerprints, palmprints, iris and etc., are the intrinsic aspects of a human and uniquely associated with the person. Biometrics characteristics cannot be lost or forgotten. They are extremely difficult to duplicate, share and distribute. Furthermore, biometrics-based system solves the problem of repudiation as it demands the person that to be identified to be physically present at the point of authentication [3]. However, biometrics cannot be easily revoked. If a biometrics is compromised, it is compromised forever and rendered unusable. This concern is exacerbated by the fact that a person has limited numbers of biometrics characteristics. Users may run out of biometrics for authentication if their biometrics characteristics keep being compromised and rendered unusable. Different applications might use the same biometric, an adversary who acquires a person's biometrics in one application could also be used for others.

Some researchers like Bolle et al. [4], Davida et al. [5] and Kevenaar et al. [6] have introduced the terms *cancelable biometrics* and *private biometrics* to rectify this issue. These terms are used to denote biometrics data that can be cancelled and replaced, as well as is unique to every application. Bolle et al. proposed that a high order polynomial function can be used as a transform function for fingerprint minutiae features. The cancelability issue of biometrics was also addressed by Andrew et al. [7]. They introduced the freshness into the authenticator via a randomized token. The revocation process is essentially the inner-product of a tokenized pseudo-random pattern and the biometrics information iteratively. Most recently, Savvides et al. [8] proposed a cancelable biometrics scheme, known as cancelable minimum average correlation energy (MACE) filters, which encrypted the training images used to synthesize the correlation filter for biometrics authentication. They showed that different templates can be obtained from the same biometrics by varying the random convolution kernels thus enabling the cancelability of the templates. They demonstrated that convolving the training images with any random convolution kernel prior to building the biometric filter does not change the resulting correlation output peak-to-side lobe ratios, thus preserving the authentication performance.

This paper proposes a revocable face authentication approach which combines user-specific tokenized pseudo-random bit sequence (TRBs) with biometrics data to generate a distinct binary code per person, coined as BioBit. BioBit inspires a two-factor face authentication mechanism by introducing the traditional biometrics recognition system a secondary authentication factor – a private token that constitutes the user TRBs. A two-factor authentication system has the advantage of avoiding any attack with single factor – attempting with a stolen token or a pre-captured face that works with the traditional system. This is the typical attack to the traditional token-based system where a hacker intrudes the system with a stolen or dictionary-generated password. BioBit fortifies the security level of the system where a legitimate access requires a valid token and a genuine face's feature. Besides, the utilization of TRBs also provides the revocation capability since TRBs is replaceable via token replacement. Without the presence of TRBs, sole biometrics suffers from irrevocable and privacy invasion issues, whilst sole token usage is susceptible to repudiation and stolen problems as discussed previously. We consider

Multimedia University, 75450 Melaka, Malaysia, yhpang@mmu.edu.my
Yonsei University, Seoul, South Korea, bjteoh@yonsei.ac.kr

the recognition performance of BioBit in three scenarios: when each legitimate user uses his own legitimate token for verification (*legitimate token case*); when a fraudulent verification is attempted using only intercepted biometrics data associated with the legitimate user, but without the associated token (*stolen-biometrics case*) and when a legitimate token is stolen and used by an imposter to claim as the legitimate user (known as *stolen-token case*). We show that the proposed scheme outperforms the sole biometrics and survives in two unfavorable attacks – stolen token and stolen biometrics scenarios.

II. THE OVERVIEW OF BIOBIT SCHEME

The overview of the proposed approach is depicted in Fig. 1. Firstly, a compact facial vector $\{\boldsymbol{x}\}$ is derived from a raw biometrics image $\{\boldsymbol{I}\}$ by using pseudo Zernike moment analysis. Through the moment analysis, the high-dimensional face data is transformed into a more compact feature vector representation. Note that pseudo Zernike moment analysis is just an example as a feature extractor used in this paper and it is possible to apply any face feature extractor, such as Eigenfaces, Fisherfaces, Discrete Cosine Transform and etc., in the proposed approach. Secondly, a bitstring generation mechanism-Chang's Multi-bit Scheme (MbS) [9] is utilized to derive a stable bitstring $\{\boldsymbol{y}\}$ from the moment features. According to the degree of distingushability, each moment feature in the feature vector may contribute more than one bit to the bitstream generation. Thereby, the bitstring space is broadened in order to obstruct imposters from exhaustive search for the correct bitstring from the bitstring space. Lastly, a user-specific token-generated random bit sequence, TRBs, $\{\boldsymbol{r}\}$ is combined with $\{\boldsymbol{y}\}$ via XOR logic operation to eventually derive a binary code, coined as BioBit $\{\boldsymbol{B}\}$. The utilization of $\boldsymbol{r}$ in BioBit computation promotes cancelable biometrics mechanism since the random bit is replaceable via token replacement.

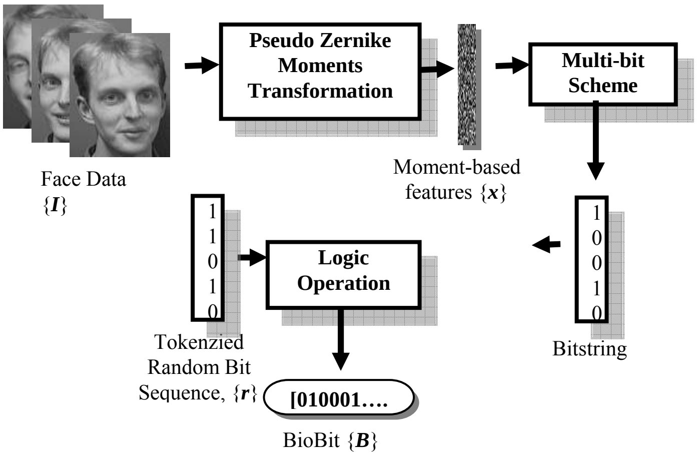

Face Data
{I}
Pseudo Zernike Moments Transformation
Moment-based features {x}
Multi-bit Scheme
1
0
0
1
0
Bitstring
1
1
0
1
0
Logic Operation
Tokenzied Random Bit Sequence, {r}
[010001….
BioBit {B}

Fig 1. The block diagram of the proposed system

III. FEATURE GENERATION: PSEUDO ZERNIKE MOMENTS

The two-dimensional pseudo Zernike moments of order p with repetition q of an image intensity function $f(r,\theta)$ are defined as [10][11]

$$\boldsymbol{PZ_{pq}} = \lambda(p,N)\int_{0}^{2\pi}\int_{0}^{1} V_{pq}(r,\theta) f(r,\theta) r dr d\theta \tag{1}$$

where pseudo Zernike polynomials $V_{pq}(r,\theta)$ are defined as,

$$V_{pq}(r,\theta) = R_{pq}(r)e^{-\hat{j}q\theta} \ ; \quad \hat{j} = \sqrt{-1} \tag{2}$$

and $r = \sqrt{x^2 + y^2}$, $\theta = \tan^{-1}\left(\frac{y}{x}\right)$, $-1 < x, y < 1$, $0 \leq |q| \leq p,\ p \geq 0$

The real-valued radial polynomials is defined as,

$$R_{pq}(r) = \sum_{s=0}^{p-|q|}(-1)^s \frac{(2p+1-s)!}{s!(p+|q|+1-s)!(p-|q|-s)!} r^{p-s} \tag{3}$$

For the implementation, a square image (N x N) is transformed and normalized over a unit circle, i.e. $x^2 + y^2 \leq 1$, in which the circle is bounding the transformed image. In this transformation [12],

$$\lambda(p,N) = \frac{4(p+1)}{(N-1)^2\pi} \ ; \ x_i = \frac{\sqrt{2}}{N-1}i + \frac{-1}{\sqrt{2}} \ ;$$

$$y_j = \frac{\sqrt{2}}{N-1}j + \frac{-1}{\sqrt{2}} \tag{4}$$

In this study, the considered features are the magnitude of $\boldsymbol{PZ_{pq}}$, $\left\{\boldsymbol{x} = \boldsymbol{PZ_{pq,mag}}\right\}$, due to their rotation invariant property; the phase information is omitted as the influence of phase information is rather insignificant especially when high order moments are included [11].

IV. STABLE BITSTRING GENERATION

To generate stable face representation in binary bitstring, Multi-bit Scheme (MbS), proposed by Chang et al. [9], is employed. The features could only be considered for the formulation of bitstring if and only if its distance between the authentic mean and the global mean is larger than k_a times of the authentic standard deviation [9][13]. Features that fall within a certain standard deviation from the genuine mean will be assigned multiple bits for the bitstring generation. Enlarged bitstring space hinders imposters from exhaustive search for a correct bitstring from the space.

Let $\boldsymbol{x}$ be a vector representing the features extracted from a face image via pseudo Zernike moments. Based on the information of $\boldsymbol{x}, \boldsymbol{y}$ of each face is generated by the following steps [9]:

(1) Compute the left and right boundaries, LB and RB, respectively, for each feature:

$$LB = \min\left(m_g - k_g\sigma_g, m_a - k_a\sigma_a\right)$$

$$RB = \max\left(m_g + k_g\sigma_g, m_a + k_a\sigma_a\right)$$

where m_a and σ_a are the mean and standard deviation of the authentic feature distribution, and m_g and σ_g are those of global feature distribution. A suitable value is set to k_g in order to cover almost 100% of the global distribution and k_a is used to control the range of authentic feature distribution $\left[m_a - k_a\sigma_a, m_a + k_a\sigma_a\right]$ to be specified as the authentic region. In our case, k_g is set to 3, while k_a is set to 0.3.

(2) Determine the number of segment with the same size as the authentic region in the ranges of (a) from LB to the left boundary of the authentic region, LB_a, and (b) from the right boundary of the authentic region, RB_a , to RB.

The number of segment from LB to LB_a is $LS = \dfrac{m_a - k_a\sigma_a - LB}{2k_a\sigma_a}$ segment(s)

The number of segment from RB_a to RB is $RS = \dfrac{RB - m_a - k_a\sigma_a}{2k_a\sigma_a}$ segment(s)

Thus, there are LS+RS+1 segments in the range of [LB RB]. At least $\log_2(LS+RS+1)$ bits are sufficient to specify each segment with a unique binary index. Please refer [9] for more detail.

V. BioBit generation and its possible attacks

In this stage, two authentication credentials are combined to generate a user-specific BioBit, ***B***. The credentials are bitstring, ***y***, and TRBs, ***r***. ***r*** is computed based on a seed stored in a token microprocessor through a random bit generator. Same seed is used for both enrollment and verification processes to a same user, but is different among different users and different applications. Employment of random data, ***r***, as one of the authentication credentials for BioBit generation enables the cancelability of the biometrics template and provides multiple application-specific verification templates from a same biometrics.

Based on the ***y*** and ***r*** information, ***B*** is generated by the following steps:

(1) Use token to generate a set of pseudo-random binary bits with the length of m, $\boldsymbol{r} = \{r_j \mid j = 1,\ldots,m\}$

(2) Compute ***B*** by coupling ***y*** and ***r*** via bit-wise XOR-logic operation, $\boldsymbol{B} = <\boldsymbol{r} \oplus \boldsymbol{y}>$.

The decision is then done based on the hamming distance matching between the reference BioBit and input BioBit by referring to a preset threshold value.

The abovementioned matching process is considered as *legitimate token scenario*. There are two possible attacks may occur in this scheme, namely *stolen-biometrics* and *stolen-token* attacks. For the stolen-biometrics case, we consider the scenario that an imposter manages to steal the genuine biometrics data with high quality and claims himself as the genuine user by taking the stolen biometrics data mixed with the imposter's token. The stolen-token scenario occurs when the genuine token is stolen and presented by an impostor as the genuine token.

However, we anticipate that the BioBit scheme still survives on these attacks based on the analysis below:
Let $\boldsymbol{r}_A$ the random number set that generated by the genuine user with his token and sealed with $\boldsymbol{y}_{EA}$ (the enrolled biometric-derived bitstring representation A) and $\boldsymbol{y}_{TA}$ (the test biometric-derived bitstring representation A) using XOR operation ($\oplus$). The output of the mixing is denoted as ***B***. Based on the unique property of XOR operation, i.e if $X \oplus Y = Z$ then $X \oplus Z = Y$ or $Y \oplus Z = X$, we have:

Case 1- legitimate-token:
Each genuine user uses his own biometric ($\boldsymbol{y}_{Ti}$) and token ($\boldsymbol{r}_i$) , where i represents different person, to access the system. The matching of the provided identity and the claimed identity will generate genuine and imposter distributions. Genuine distribution is the scattering where each BioBit of the class is matched against all other BioBits in the same class; whereas, imposter distribution is the scattering where the BioBit of each class is matched against the BioBit of all the other classes.

Genuine distribution:

claimed identity $\boldsymbol{r}_A \oplus \boldsymbol{y}_{EA} = \boldsymbol{B}_{AA}$

provided authenticators $\boldsymbol{r}_A \oplus \boldsymbol{y}_{TA} = \boldsymbol{B}_{AA}'$

Matching based on Hamming distance metric

$$\boldsymbol{B}_{AA} \oplus \boldsymbol{B}_{AA}' = \boldsymbol{B}_{AA} \oplus \left(\boldsymbol{r}_A \oplus \boldsymbol{y}_{TA} \right)$$

Assume that $\boldsymbol{y}_{EA}$ is very close to $\boldsymbol{y}_{TA}$, $\boldsymbol{y}_{EA} \approx \boldsymbol{y}_{TA}$, so

$$\begin{aligned}\boldsymbol{B}_{AA} \oplus \boldsymbol{r}_A \oplus \boldsymbol{y}_{TA} &\approx \boldsymbol{B}_{AA} \oplus \boldsymbol{r}_A \oplus \boldsymbol{y}_{EA} \\ &\approx \boldsymbol{B}_{AA} \oplus \boldsymbol{B}_{AA} \approx 0\end{aligned}$$

The genuine population should be peak at zero distance if and only if $\boldsymbol{y}_{EA}$ is exactly equal to $\boldsymbol{y}_{TA}$. However, practically, it is impossible to get such result due to the biometric's inherent high degree of uncertainty, but near to zero distance is obtained.

Imposter distribution:

claimed identity $\boldsymbol{r}_A \oplus \boldsymbol{y}_{EA} = \boldsymbol{B}_{AA}$

provided authenticators $\boldsymbol{r}_A \oplus \boldsymbol{y}_{To} = \boldsymbol{B}_{oo}$

Since $\boldsymbol{r}_A \neq \boldsymbol{r}_A$ and $\boldsymbol{y}_{EA} \neq \boldsymbol{y}_{To}$,

$\boldsymbol{B}_{AA} \oplus \boldsymbol{B}_{oo}$ = large distance

Since $\boldsymbol{r}_A \neq \boldsymbol{r}_o$ and $\boldsymbol{y}_{EA} \neq \boldsymbol{y}_{To}$, the matching of BioBits of different classes in legitimate token case shifts the imposter distribution towards larger distance.

Case 2- stolen-biometrics:
An imposter o has stolen the biometric of a genuine user A and claimed himself as genuine A by using his token ($\boldsymbol{r}_o$) and the user A's biometric ($\boldsymbol{y}_{TA}$).

claimed identity $\boldsymbol{r}_A \oplus \boldsymbol{y}_{EA} = \boldsymbol{B}_{AA}$

provided authenticators $\boldsymbol{r}_o \oplus \boldsymbol{y}_{TA} = \boldsymbol{B}_{oA}$

Matching based on Hamming distance metric

$$\begin{aligned}\boldsymbol{B}_{AA} \oplus \boldsymbol{B}_{oA} &= \boldsymbol{B}_{AA} \oplus \left(\boldsymbol{r}_o \oplus \boldsymbol{y}_{TA} \right) \\ &= (\boldsymbol{B}_{AA} \oplus \boldsymbol{y}_{TA}) \oplus \boldsymbol{r}_o\end{aligned}$$

Assume that $\boldsymbol{y}_{EA}$ is close to $\boldsymbol{y}_{TA}$, so

$$\boldsymbol{B}_{AA} \oplus \boldsymbol{y}_{TA} \approx \boldsymbol{B}_{AA} \oplus \boldsymbol{y}_{EA}$$

then,

$$\begin{aligned}\boldsymbol{B}_{AA} \oplus \boldsymbol{B}_{oA} &= (\boldsymbol{B}_{AA} \oplus \boldsymbol{y}_{TA}) \oplus \boldsymbol{r}_o \\ &\approx (\boldsymbol{B}_{AA} \oplus \boldsymbol{y}_{EA}) \oplus \boldsymbol{r}_o \\ &\approx \boldsymbol{r}_A \oplus \boldsymbol{r}_o\end{aligned}$$

The result of this case will be similar to the result of the tokenized random number matching, if and only if $\boldsymbol{y}_{\mathrm{EA}}$ near to $\boldsymbol{y}_{T\mathrm{A}}$.

Case 3- stolen-token:

An imposter o has stolen the token of a genuine user A and claimed himself as genuine A by using his biometric data ($\boldsymbol{y}_{T\mathrm{B}}$) and the user A's token ($\boldsymbol{r}_o$).

claimed identity $\quad \boldsymbol{r}_A \oplus \boldsymbol{y}_{EA} = \boldsymbol{B}_{AA}$

provided authenticators $\quad \boldsymbol{r}_A \oplus y_{To} = \boldsymbol{B}_{Ao}$

Matching based on Hamming distance metric

$$\begin{aligned}\boldsymbol{B}_{AA} \oplus \boldsymbol{B}_{Ao} &= \boldsymbol{B}_{AA} \oplus \left(\boldsymbol{r}_A \oplus y_{To} \right) \\ &= (\boldsymbol{B}_{AA} \oplus \boldsymbol{r}_A) \oplus y_{To} \\ &= y_{EA} \oplus y_{To}\end{aligned}$$

The result of this case will be similar to the result of the biometric-derived bitstring matching (MbS recognition performance).

The above analysis can be validated by the experiments that will be done in section 6.0.

VI. EXPERIMENTAL RESULTS AND DISCUSSIONS

The proposed method is evaluated on images taken from two different types of face databases, which are Olivetti (also known as ORL) face database and Face Recognition Technology (FERET) face database (FERET). In ORL database, there are ten (10) different images of each of 40 distinct subjects (classes). For some subjects, the images were taken at different times, between April 1992 and April 1994, with the variations of lighting, facial expressions (open/ closed eyes, smiling/ not smiling) and facial details (glasses/ no glasses) [14]. FERET is a large and well-designed face database. FERET contains face images associated with 1196 individuals. Since our normalization approach – Hambridge feature location [15][16] is constraint to frontal images, a substantial number of these (i.e. left- and right-profile views) are unsuitable for our experiments. Thus, we have selected a subset of 230 users, each having six essentially normalized frontal images with variations in pose (i.e. within ± 25 degrees rotation in depth) scale and illumination. Many of these images were taken over an extended period, and are highly varied in terms of eyewear (absence and presence thereof) and illumination [17].

In this paper, the following are the abbreviations used for brevity in our subsequent discussion:

- *PZM* denotes pseudo Zernike moments analysis.
- *mPZM* denotes Multi-bit Scheme.
- *mPZM_XOR* denotes BioBit scheme with XOR logic operation for integrating $\boldsymbol{y}$ and $\boldsymbol{r}$.

In our experiment, each face sample in a same class (denoted as intra-class sample) is combined with the same distinct TRBs to generate BioBit template per sample, known as intra-class BioBit. Different sets of TRBs are combined with different face classes in order to produce inter-class BioBit templates. To generate genuine distribution, intra-class BioBit*s* are matched against each other. In other words, each BioBit of the class is matched against all other BioBit*s* in the same class. This process is repeated for all the other classes. For imposter distribution, the first BioBit of each class is matched against the first BioBit of all the other classes and the same matching process is repeated for the subsequent BioBit*s*.

To simulate the stolen-biometrics scenario, a face data is mixed to all sets of random bit sequences from all classes and the matching is done according to the imposter match described above and we name the imposter distribution that obtained as pseudo-genuine 1. For stolen-token scenario, we consider the worst case whereby only one set of random bit sequence is mixed to all face images of all classes and the matching is done according to the imposter match - pseudo-genuine 2. We repeat the same process ten times and the results are averaged to reduce the statistical frustration caused by random numbers. For matching classification, Hamming distance is used for binary outputs generated by *mPZM* and *mPZM_XOR*, while a simple Euclidean distance metric is adopted for real value output generated by *PZM*. In this paper, we compare our proposed method, BioBit scheme, with a new cancelable biometrics approach based on BioHashing (BioHash), proposed [7][18] in the legitimate token, stolen-biometrics and stolen-token cases.

A. Performance evaluation

The first experiment is conducted for the purpose of determining a pertinent and optimal moment-based feature vector which is able to provide an optimal representation for the biometrics identity. At this stage, the performance of *PZM* in term of equal error rate (EER) is evaluated. EER is the average value of False Accept Rate (FAR, probability of a not-enrolled individual being identified) and False Reject Rate (FRR, probability of an enrolled individual not being identified). Table I shows the error rates of *PZM* with different feature length setting for both ORL and FERET databases. The table demonstrates that larger number of feature length (or higher moment order) provides better verification rate because higher order moments provide more and finer details about the face image. However, this is only true to a certain point as the verification rate will level off or even worse if the feature length is extended further. *PZM*'s EERs show increasing after feature length =160 in ORL database and feature length =100 in FERET database. This implies that excessive high order moment features contain redundant information and unwanted signals (e.g. noise) which might influence the representation capability. From the table, we can see that the overall error rates of *PZM* in both databases are relatively high. This is because the significant illumination and facial expression variations in the face images create a serious degradation on biometrics, influencing the representation capability of *PZM*.

From Table II, we can observe that *mPZM* attains much better recognition performance than *PZM*. This implies that Multi-bit Scheme (*mPZM*) is able to generate a more stable and distinguishable bitstring, even through its input is a non-stable moment feature vector. From the figures, in general, the performances of *mPZM_XOR* are increasing for larger feature length. This explains that larger feature length comprises more refined feature representation with higher accuracy. We also can see that the proposed scheme, *mPZM_XOR*, is just slightly poorer than BioHash in the legitimate token case. But, an impressive performance is still

TABLE I
THE ERROR RATES OF *PZM* WITH DIFFERENT FEATURE LENGTH SETTING FOR BOTH ORL AND FERET DATABASES

Database	Feature length	EER (%)	Database	Feature length	EER (%)
ORL	20	20.55	FERET	20	28.04
	40	17.74		40	26.72
	60	17.76		60	25.89
	80	17.23		80	25.46
	100	17.02		100	25.30
	120	16.60		120	26.00
	140	16.33		140	26.39
	160	16.25		160	26.80
	180	16.38		180	27.13
	200	16.40		200	27.29
	220	16.36		220	27.63
	240	16.40		240	27.57
	260	16.42		260	27.78
	280	16.51		280	27.95
	300	16.54		300	27.96

obtained by *mPZM_XOR*, yielding near to zero EER (EER=0.012% when tested on ORL and EER=0.016% on FERET, with feature length 100). Besides that, through the analysis, BioBit scheme is able to provide an impressive performance with EER=0.001% and EER=0.002% tested on ORL and FERET datasets in the stolen-biometrics case. In the stolen-token case, *mPZM_XOR* achieves much superior recognition result than BioHash. The proposed method is able to attain EER=1.28% on ORL and EER=1.36% on FERET when stolen-token case occurred. Unfortunately, BioHash shows a great degradation in performance, by obtaining EER=24.5% on ORL and EER=31.5% on FERET, in the stolen-token case, as pointed by [19][20].

We also demonstrate the performance of *PZM*, *mPZM*, *mPZM_XOR* and BioHash in the form of Receiver Operating Characteristic (ROC) curves as plot of the genuine accept rate (GAR) against the false accept rate (FAR) for all possible operating points in Fig. 2. The nearer the curve to the upper left corner, the better the performance is. It can be seen that our proposed method, as well as BioHash, achieves excellent recognition performance, nearly to 100% GAR at FAR=0.01%, in the legitimate token case. However, BioHash obtains serious performance degradation when the stolen-token scenario occurred. Fortunately, our proposed system, BioBit, is found less suffered to this problem. BioBit (*mPZM_XOR*) attains false accept rate for pseudo-genuine 2 (stolen-token case) is 0.002% in ORL and is 0.001% in FERET when we set a threshold along with the genuine accept rate of 90%.

Fig. 3(a) depicted the genuine, imposter and pseudo-genuine populations of *mPZM_XOR* tested on ORL in the legitimate token, stolen-biometrics and stolen-token cases, same to BioHash in Fig. 3(b). For the legitimate token and stolen-biometrics cases, there is a good separation of the genuine and imposter distributions, as well as pseudo-genuine I distribution, in both cancelable biometrics techniques. This justifies that the two approaches are robust and able to achieve excellent recognition performance, nearly to 100% GAR, in both legitimate token and stolen-biometrics cases. However, for the stolen-token case, the pseudo-genuine II distribution is slightly overlapped with the genuine and hence it deteriorates the performance to EER= 1.28% in our proposed system. Nevertheless, it is far better than the BioHash which obtains EER=24.5%, having a strong overlapping in between the pseudo-genuine

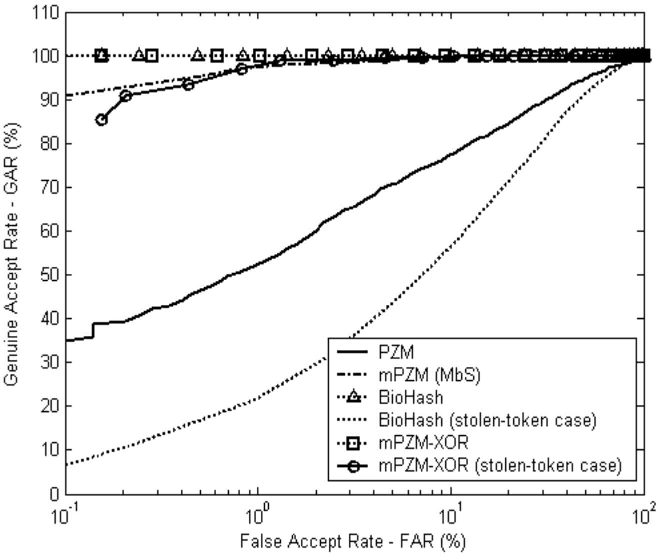
TABLE II
PERFORMANCE EVALUATION IN TERM OF EER FOR *PZM*, *MPZM* AND *MPZM_XOR* FOR DIFFERENT FEATURE LENGTHS, TESTED ON (A) FERET DATASET AND (B) ORL DATASET.

(a) FERET dataset

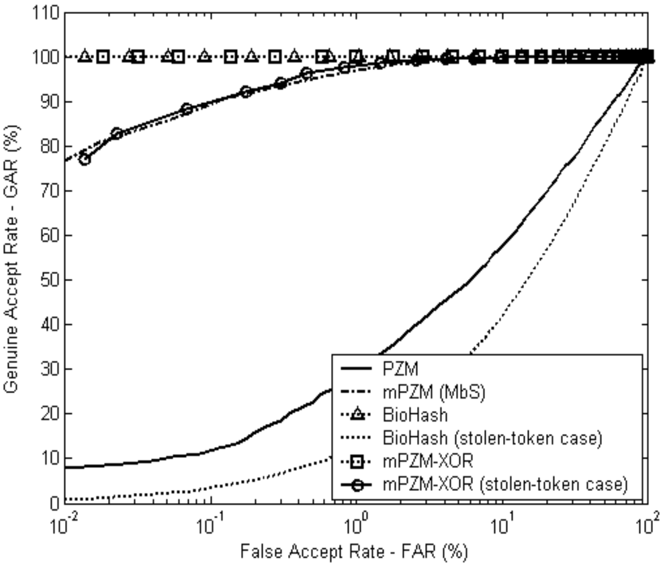
Method	Equal Error Rates (EERs) (%)				
	20	40	60	80	100
PZM	28.04	26.72	25.89	25.46	25.3
mPZM	13.68	7.44	4.11	2.34	1.76
BioHash	2.04	0.62	0.18	0.03	0.00
BioHash (stolen-biometrics case)	1.68	0.38	0.12	0.08	0.05
BioHash (stolen-token case)	35.44	36.99	35.36	36.55	31.54
mPZM_XOR/ BioBit	4.80	1.20	0.53	0.17	0.01
mPZM_XOR (stolen-biometrics case)	0.24	0.16	0	0	0
mPZM_XOR (stolen-token case)	10.66	5.71	3.10	1.88	1.36

(b) ORL database

Method	Equal Error Rates (EERs) (%)				
	20	40	60	80	100
PZM	20.55	17.74	17.76	17.23	17.02
mPZM	15.08	5.73	3.20	1.61	1.61
BioHash	0.64	0.22	0	0	0
BioHash (stolen-biometrics case)	0.67	0.05	0.01	0	0
BioHash (stolen-token case)	36.08	32.66	27.48	26.46	24.51
mPZM_XOR/ BioBit	5.44	1.34	0.78	0.35	0.012
mPZM_XOR (stolen-biometrics case)	3.14	0.38	0.06	0	0
mPZM_XOR (stolen-token case)	14.45	7.45	4.12	2.51	1.27

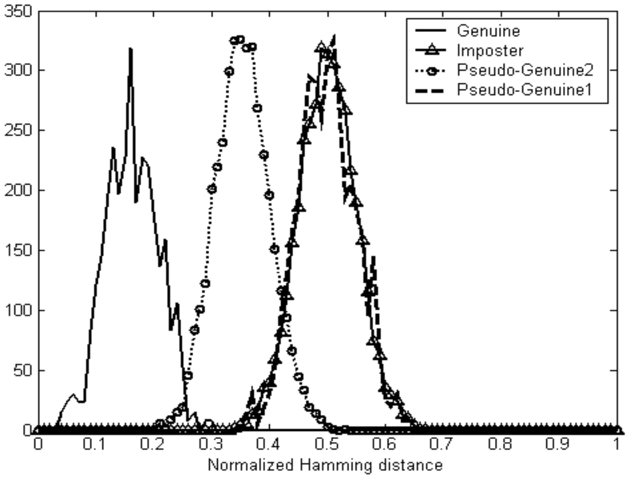


(a) ORL dataset

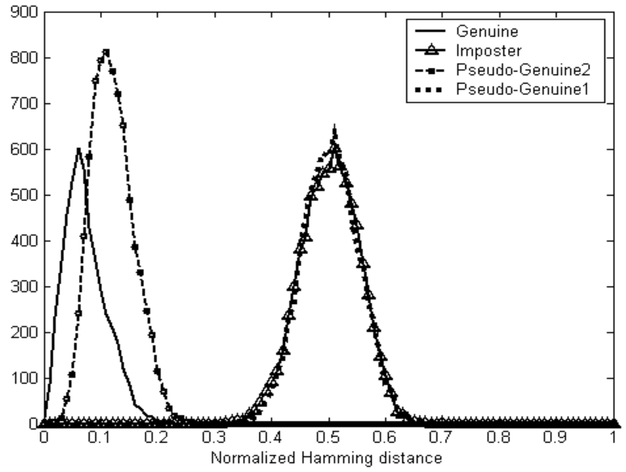


(b) FERET dataset

Fig. 2. The Receiver Operating Characteristic (ROC) curves of genuine, imposter and pseudo-genuine2 (stolen-token case) for (a) ORL dataset and (b) FERET dataset, at *m*=100.



(a)



(b)

Fig. 3. The distributions of genuine, imposter, pseudo-genuine1 (stolen-biometrics case) and pseudo-genuine2 (stolen-token case) for (a) *mPZM_XOR* and (b) BioHash, in ORL database with *m*=100.

II and genuine distributions. The results justify the analysis in section 5.0 regarding the robustness of BioBit in legitimate token, stolen-token and stolen-biometrics cases.

VII. Conclusion and Future Works

In practical, this is hard to get nearly zero FAR and FRR errors in the biometrics recognition systems due to the fact that the biometrics signals have high uncertainty and the classes are difficult to completely separate in the measurement space. This paper presents a two-factor recognition system by using both face data and a unique token deposited with random binary data. The proposed BioBit scheme offers 3 main advantages:

(1) The combination of biometrics face data and user-specific random bit sequence provides a perfect verification performance. The proposed scheme is able to provide a clear separation of the genuine and imposter populations in the legitimate token case. On the other hand, the proposed technique is still able to obtain an encouraging result when the stolen-token case occurred, compared with BioHash.

(2) In this proposed approach, two authentication factors are needed in order to derive a BioBit for verification. This can strengthen the security of the recognition system as two-factor authentication fortifies the security by prohibiting any single-factor attack;

(3) The proposed scheme also addresses the invasion of privacy issue, such as biometrics fabrication. The compromised BioBit could be alleviated through the user-specific credential revocation via token replacement. Furthermore, multiple templates can be generated from a same biometrics for multiple applications using different sets of bit sequences.

Future works in the implementation of other logic operations (such as AND, OR, etc), a combination of logic operations or other forms of mixing mechanism are directed towards the improvement of the practicability of the BioBit recognition system.

Acknowledgment

Authors would like to thank the Speech, Vision and Robotics Group of the Cambridge University Engineering Department and NIST for face database provision.

References

[1] E.J. Vargas, "Introduction to authentication," [Online]. Available: http://www.jyestudio.com/docs/introauth.pdf

[2] O.G. Lawrence, "Comparing passwords, tokens, and biometrics for user authentication," in *Proc. of the IEEE, 91(12)*, 2003, pp. 2019-2040.

[3] C. Sharat, C. Viraj, G. Venu, "A study on the convergence of biometrics and cryptograph," [Online]. Available: http://www.eng.buffalo.edu/~ssc5/research/papers/biometrics_and_cryptography.pdf

[4] R.M. Bolle, J.H. Connel, N.K. Ratha, "Biometric perils and patches," *Pattern Recognition, 2002, 35*, pp. 2727 2738.

[5] G. Davida, Y. Frankel, B.J. Matt, "On enabling secure applications through off-line biometric identification," in *Proc. Symposium on Privacy and Security*, 1998, pp. 148-157.

[6] T.A.M. Kevenaar, G.J. Schrijen, van der Veen M, A.H.M. Akkermans, F. Zuo, "Face recognition with renewable and privacy preserving binary templates," in *Proc. of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies AUTOID '05, IEEE Computer Society*, 2005, pp. 21-26.

[7] T. Andrew, N. David, G. Alwyn, "An integrated dual factor verification based on the face data and tokenised random number," *In Proc. of LNCS, Springer-Verlag,* 3072, David Zhang, Anil Jain (Eds.), 2004, pp. 117-123.

[8] M. Savvides, B.V.K. Vijaya Kumar, P.K. Khosla, "Cancelable biometric filters for face recognition," in *Proc. of the 17th International Conference on Pattern Recognition (ICPR'04)*, 2004, pp. 922-925.

[9] Y.J. Chang, W.D. Zhang, T.H. Chen, "Biometrics-based cryptographic key generation," in *Proc. of the 2004 IEEE International Conference on Multimedia and Expo, vol. 3*, 2004, pp. 2203-2206.

[10] C.H. Teh, R.T. Chin, "On image analysis by the methods of moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1988, 10, pp. 496-512.

[11] A. Khotanzad, "Invariant image recognition by Zernike moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, 12, pp. 489-497.

[12] C.W. Chong, P. Raveedran, R. Mukundan, "Translation invariant of Zernike moment," *Journal of the Pattern Recognition Society, 2003, 36*, pp. 1765-1773.

[13] F. Monrose, M.K. Reiter, Q. Li, S. Wetzel, "Cryptographic key generation from voice," in *Proc. of the 2001 IEEE Symposium on Security and Privacy*, 2001, pp. 202-213.

[14] ORL face database. [Online]. Available:
http://www.uk.research.att.com/facedatabase.html

[15] J. Hambridge . The elements of dynamic symmetry. Yale University Press, New Haven, USA, 1926.

[16] N. David, G. Alwyn, "Facial feature extraction via Dynamic Symmetry Modeling for user identification," *Pattern Recognition Letter*, 2003.

[17] FERET face database. [Online]. Available:
http://www.itl.nist.gov/iad/humanid/feret/feret_master.html

[18] N. David, T. Andrew, G. Alwyn, "Eigenspace-based face hashing," *In Proc. of LNCS, Springer-Verlag,* 3072*,* David Zhang, Anil Jain (Eds.), 2004, pp. 195-199.

[19] K.H. Cheung, A. Ong, D. Zhang, M. Kamel, J. You, H.W. Lam, "An analysis on accuracy of cancelable biometrics based on BioHashing," *KES 2005, LNAI 3683*, 2005, pp. 1168-1172.

[20] Loris Nanni, A. Lumini, "Human authentication featuring signatures and tokenized random numbers," *NeuroComputing, vol. 69*, March 2006, pp. 858-861.