IMPROVED FUZZY HASHING TECHNIQUE FOR BIOMETRIC TEMPLATE PROTECTION

T. S. ONG
Faculty of Information Science and Technology,
Multimedia University, Melaka, Malaysia

ANDREW B.J TEOH
School of Electrical and Electronics Engineering,
Yonsei University, Korea

W. H. KHOH
Centre for Diploma Programme,
Multimedia University, Melaka, Malaysia

ABSTRACT

Biometrics provides a new dimension of security to modern automated applications since each user will need to prove his identity when attempting an access. However, if a stored biometric template is compromised, then the conventional biometric recognition system becomes vulnerable to privacy invasion. This invasion is a permanent one because the biometric template is not replaceable. In this paper, we introduce an improved FuzzyHashing technique for biometric template protection purpose. We demonstrate our implementation in the context of fingerprint biometrics. The experimental results and the security analysis on FVC 2004 DB1 and DB2 fingerprint datasets suggest that the technique is highly feasible in practice.

Keywords: Fuzzy Hashing, Template Protection, Fingerprint Biometrics

INTRODUCTION

In the recent year, research to provide reliable biometric authentication system has been conducted extensively. Biometrics provides good alternative to traditional token (e.g ID card) or knowledge (e.g. password) for personal authentication system, as it is unlikely biometric data will get lost or forgotten. A conventional approach to biometric authentication is to capture the biometric template of all users during enrollment stage and store the templates in a reference database to be used during authentication stage by matching against another live captured biometric measurement. Nevertheless, there are some security and privacy issues posed to this approach, and one of the major concern is the invasion of privacy and non-replaceable of biometrics, e.g., the digital storage of biometric data, known as templates in digital form in a reference database is raising concern about their protection and usage of such information for all sorts of purposes, without the knowledge or consent of the user (Boult, Scheirer and Woodworth, 2007). Adding to this, the authentication established using physical or knowledge possession like a token of password or PIN is re-issuable. This can be seen when an ATM card is lost or stolen, the banks just issues a new password or pin to protect the customer whereas this is not possible in the case of biometrics since we cannot change our own biometric components even we desire.

Generally, template protection technique is an emerging research direction spurred by the need to resolve the security and privacy risks that associated with biometric template storage.

There are two possible solutions to tackle the above mentioned issues, namely cancelable biometrics (Ratha, Connell and Bolle, 2001) and biometric encryption (BE) (Cavoukian and Stoianov, 2007). For cancelable biometrics (CB), the main idea is to converts a real-valued biometrics into a transformed template with the storage of some optional tokenized random number that could be used during the matching process. The general idea is to secure a biometric system through concealing raw biometric data to preserve user privacy. On the other hand, biometric encryption (BE) encompass the design of non-invertible of biometric representation via incorporating of helper parameters with an externalized identifier. i.e., random number generator (RNG) that served as protection mechanism to protect the security of storage template through template binding or hiding, thus no original biometric template is needed to be tied to an individual. In certain extent, BE could enhance the accuracy of template matching with the capability of template rectification and error correction using the helper parameters to the key authenticator in the model. A typical example of this abstraction is fuzzy commitment scheme (Juels and Wattenberg, 1999), which will be discussed in section 2.

RELATED WORK AND MOTIVATION

The most notable one of CB was proposed by Ratna et al. in year 2001, where the biometric signal is distorted intentional in the same fashion at both enrollment and authentication mode. Their proposal is realized in (Ratha, Chikkerur, Connell and Bolle 2007) by transforming fingerprint into various cartesian, polar and surface folding of the minutia positions. Based on empirical analysis, the authors have proven that the various proposed algorithms can achieve diversity, non-invertibility and performance practicality as required for template protection.

Another approach proposed by Teoh, Ngo and Goh (2004) which coined as Biohash. This technique first generate and orthonormalize a set of user-specific random number matrices using Gram-Schmidt process and then combines the tokenized random numbers with their biometric feature vectors via inner-product operation. The formulation has been tested with various biometrics such as fingerprint (Teoh, Ngo, and Goh, 2004), iris (Chong, Teoh, and Ngo, 2005), palm print (Connie, Teoh, Goh, and Ngo, 2004) and face (Teoh and Ngo, 2005) with nearly zero error rates were reported for all types of biometrics. However, their performance degrades substantially in the stolen token scenario when the genuine token was being used by imposter for verification purpose. Following the this, the author has proposed a variant of such method, known as Multiple Random Projection (MRP) (Teoh and Chong, 2007) to address the problem of recognition performance and the experimental result showed that the original performance as in sole biometrics was retained even in the stolen token scenario.

On the other hand, BE was first outlined by Davida, Frankel and Matt (1998). Either a biometric template itself, or a hashed value derived from it, was used as an identifier in this case. It was suggested to use error correcting codes (ECC) to compensate the bit variations and a one-way hash function to conceal the biometric template. However, a direct use of ECC in biometrics is vulnerable as it leads to data leakage. The fuzzy commitment scheme (FCS) of Juels and Wattenberg (1999) further extended the idea of Davida et al (1998). This scheme works as follows: at enrolment, let the binary biometric feature be denoted as b, the identifier, Id, the mixture, M=enc(Id)+b, where enc(•) is a ECC encoding function. The + operation is implemented as bitwise XOR in this context as the authors assumed binary biometric representation. M itself reveals no information about Id as well as b if both are uniformly random; though this assumption is difficult to materialize in practice. A fuzzy commitment {M, h(Id)} is formed where h(•) is a one-way hash function. Given a query biometric sample b' during verification, Id'=M+b'is computed, decoded and hashed. h(Id') = h(Id) if and only if $b \approx b$'. A few practical constructs were

reported in (F. Hao, Anderson and Daugman (2000), Tuyls , Akkermans, Kevenaar and Greert (2005), Kevenaar, Schrijen, Akkermans and Zuo (2005)). In addition, Juels and Sudan (2002) proposed an extension of fuzzy commitment via a secret sharing scheme to allow a non-exactly ordered biometric representation, namely fuzzy vault.

In this work, a secure template protection method known as improved Fuzzy Hashing (FZH) is proposed. FZH generates a revocable but irreversible template directly from biometric data. The proposed technique is hybrid approach of cancellable biometrics as well as biometric encryption. The former requires a transformed template to be stored and the decision is made based on a pre-selected system threshold, but this is not required in FZH. Besides, FZH possess revocable capability and rectifies the non-uniform problem of biometric feature, which is lacking in the BE approach. In certain extent, both FZH and cryptographic hash share some common properties, such as one-way transformation (non-invertibility), uniformity and randomness. However, since biometric data is fuzzy and subject to intra-user variations such as rotation and alignment issue of fingerprint or illumination and pose variation in face etc. In this context, error correction method is applied to reduce and rectify the differences of biometric data for a reliable performance.

As compared to the first proposed FuzzyHash on face biometrics (Teoh and Kim, 2007), this paper proposes an improved FuzzyHashing (FZH) technique which differs in the following aspects:

1.) We propose a new realization based on the FZH framework which possesses better performance and stronger security proven than that of the previous realization.
2.) We examine the technique in the context of fingerprint biometrics.

DEFINITION AND DESIGN REQUIREMENTS

The formal definition of FuzzyHash is outlined as follow.

DEFINITION 1: A (*k*, δ, *s*, **Gen**, **Ext**) – FuzzyHash template, *h* is constructed using a transformation function, $T : R^{n+m} \rightarrow \{0,1\}^k$ be a function to map a sequence *x* onto FZH based on user-specific information, *w*. Let *d* be a metric space on T with a distance function $\varepsilon : d \times d \rightarrow R$. The function T is called ε-robust to noise if and only if for all $x \in R^n$, there exists of $w \in R^m$ such that $\varepsilon > 0, \varepsilon(x, x') \leq \delta$, where **Gen**(*x*) = *w*, then **Ext**(*x*', *w*) = *h*. Specifically, **Gen** is a the procedure used to derive *w* from a set of biometric feature *x*, while **Ext** is the function allowing extraction of *h* based on the corresponding auxiliary information *w*, if *x*' is sufficiently close to *x*. With this formulation, only auxiliary information, *w* is required to be stored, instead of *x* to avoid the expose of original biometrics. The auxiliary info, *w* should not leak any information to enable the recovery of biometric data. Note that the key idea is to allow *h* to be generated on the fly when *x*' and *w* are presented.

Based on the definition above, the requirements for generating a FuzzyHash template are as follows:

1) **Reproduction.** For all possible intra-user variations, the same template, *h* should be reproducible whenever a similar biometric is presented.

2) **Revocability and Diversity.** Reissuable of the transformed template and no single template can be used in more than one application.

3) **One-way transformation.** It refers to non-invertibility of the feature domain transformation. i.e. the altered domain of the chosen transform must not be able to be inverted to original feature domain in whatever circumstances. In other words, x is irremissible even though h and w are presented.

4) **Randomness and uniformity.** For any input x, h should be approximately uniformly distributed among all possible 2^k template outputs. In other words, the entropy of each h generated as $H(x) = -\Sigma Pr(x)\log_2 Pr(x)$ with probability $Pr(X=x)$ for discrete random variable on a finite set $\{X=x_i | i = 1,\ldots,n\}$, with high randomness of maximum entropy ~1 for the binary case. The highly randomness implies an adversary could not perform a random statistical extraction of feature space patterns based on intercepting multiple FuzzyHash templates.

The rest of the paper is organized as follows: Section 3 outlines the design specification and requirements for a reliable template protection technique while Section 4 details the proposed algorithm. Experimental and security analysis are provided in Section 5, while section 6 and 7 presents the discussion and the concluding remarks, respectively.

REALIZATION

To realize the proposed FZH transformation function, we follow a three-step procedure. The steps are described as follows:
1) Random mixing of biometric feature vector f: $\square^n \rightarrow \square^p$, where $p<n$. This can be done through the BioPhasoring proposed in **Error! Reference source not found.**].
2) Discretize the output from f via a discretization technique **Error! Reference source not found.**], g: $\square^p \rightarrow \{0, 1\}^k$ where $p \leq k$.
3) Error correction code e: $\{0, 1\}^k \rightarrow \{0, 1\}^k$

In this context, T is then redefined as a composition function, ie. $T = f \circ g \circ e$. In this work, we also quantify the relation of step (2) and step (3) so that to optimize the best parameters for error correction.

BioPhasoring

The BioPhasoring comprises of two stages: (a) feature extraction and (b) random mixing. In this context, a 1-D feature vector $\boldsymbol{x}=\{x_i | i =1,\ldots,n\}$, with fixed length n is fist extracted from raw biometric data. Random mixing is carried out based on an iterated assimilation between the tokenized pseudo-random number (PRN) and the biometric feature. The PRN is derived from an external secret, such as a password or a token and *specific* to each user. The objectives of the proposed method are two folded: (1) To inject randomness into the biometric feature, and (2) To enable the changeability of biometric feature by altering the PRN.

Specifically, a BioPhasor for a kth user is given as

$$\alpha_{jk} = \frac{1}{n}\sum_{i=1}^{n} tan^{-1}\left(\frac{x_i^k}{r_{ij}^k}\right),\ j = 1,\ldots,p,\ p < n \qquad (1)$$

where $\mathbf{r}_j = \{r_i | i=1,\ldots,n\}$ is a PRN set independently drawn from $\mathbf{N}(0, 1)$. The output is a set of p values $\boldsymbol{\alpha} = \{\alpha_i | i=1,\ldots,p\}$ with range $[-\pi, \pi]$. Note that (1) is an underdetermined non-linear equation system. The inversion of (1) given $\boldsymbol{\alpha}$ and PRN is a NP-hard problem provided $p<n$. However, the performance will degrade if $p<<n$.

One potential issue of BioPhasoring is that if the genuine PRN is stolen by an adversary and uses it to claim to be the genuine user with his biometric feature; we shall consider whether the adversary can impersonate the genuine in this scenario (stolen-token scenario). In this case, the false accept rate will increase as compared to the normal setting (different-token scenario hereafter). We shall show experimentally in section 5.2 that BioPhasor is resistant to this threat.

Discretization

We follow the 2^N discretization algorithm as proposed in (Andrew, Toh and Yip, 2007) to divide the biometric sample space into equal-width intervals before converting each interval to a binary number. For a user j, the feature space is divided into 2^N segments based on the range of right (R) and left (L) boundaries of entire feature space. Note that the segment width, w varies according to $w = \arg\min_N \left(\left| \frac{R - L}{2^N} - 2\sigma_{ij} \right| \right)$ for feature element i of user j and σ_{ij} is the standard deviation of α_i of user j. In particular, the randomly transformed feature's ith element of user k, $\boldsymbol{\alpha} = \{\alpha_{ik} | i=1,\ldots,p\}$ are quantized into a binary representation based on a two-state decision of $-\pi$ and π, based on the mean of vector elements, $b_{jk} = \begin{cases} 0 & \text{if } 0 \le \alpha_{jk} < \pi \\ 1 & \text{if } -\pi < \alpha_{jk} \le 0 \end{cases}$.

During this transformation, the number of bit in each segment, γ_{ij} of each user can be determined from the divided feature space as the auxiliary information. To reduce the impact of too much information lost during the quantization process, we transform the feature set $\{\alpha_i | i=1,\ldots,p\}$ such that each transformed feature is discernible to separate the genuine user from potential impostors. Specifically, we transform α_i from the real space into the index space and follow by a gray encoding which produce a set of compact bit string of template. The use of gray encoding is to ensure that further states of genuine region. i.e., occurring with high probability from imposter would have higher hamming distance.

Error Correction Coding (ECC)

For error correction, we adopt Reed-Solomon code (RSB), which uses linear algebraic code for multiple error correction and is an important subclass of BCH codes. Generally, RSB belongs to the family of linear cyclic block codes and is designated as $\varepsilon(n_b, k_b, t_b)$ where $n_b \le 2^{m_b} - 1$ and $2t_b \le n_b - k_b$. k_b is the number of blocks after encoding, n_b represents the number of blocks before encoding, t_b the number of error blocks that can be corrected and m_b the bits number per block. We denote the ε parity as $\rho_b = n_b - k_b$.

EVALUATIONS

Database Setting

We use the fingerprint biometrics as the subjects of study. The proposed method is evaluated using datasets from FVC 2004, which was established with the aim of providing a benchmark to determine the state-of-the-art in fingerprint recognition applications (http://bias.csr.unibo.it/fvc2004/). The FVC 2004 comprises of four different databases (DB1-DB4) by using the following sensors/technologies:

1. DB1: optical sensor "V300" by CrossMatch.
2. DB2: optical sensor "U.are.U 4000" by Digital Persona.
3. DB3: thermal sweeping sensor "FingerChip FCD4B14CB" by Atmel.
4. DB4: synthetic fingerprint generation.

We only opt for DB1 and DB2 for experiments as they were acquired using the optical sensor, which is the most commonly used sensor in the market. Each DB1 and DB2 consists of 100 subjects and each subject contains 8 fingerprint impressions. We combine both these datasets and hence we have 200 images in total. Every impression undergoes a core-point detection by the method proposed in (Teoh, Ong and Ngo, 2003), after which a 128x128 square fingerprint image region being cropped with respect to this reference point. Five images from each subject are randomly selected to be the training samples for discretization (as well as parity generation) and the remainders are used for testing. A multichannel Gabor Filter (FingerCode) (Jain, Prabhakar, Hong and Pankanti, 2000) is used to extract the fingerprint features. The outcome is an ordered feature vector with length 376.

Generally, the performance of genuine distribution is based on the matching of different fingerprint in each class, which leading to 600 ((3x2)/2 attempts of each subject x 200) attempts. On the other hand, the imposter distribution is determined by the matching of each subject of a class against the all other subjects, where the processes are repeated for all subsequent templates for (200x199)/2 x 3 = 59,700 impostor attempts. For the stolen-token scenario described in section 4.1, we consider the worst-case scenario where the adversary always manages to obtain the genuine token. In other words, only one set of pseudo-random numbers (PRN) is mixed with all FingerCodes. imposter matching were evaluated under the assumption that imposters possess stolen PRN described above. To avoid statistical bias caused by the varying random matrices, the experiment is repeated run for 50 times and the results are averaged to obtain AER.

Performance

For the experiments, we first examine the effect of various p in equation (1) with n= 376. We vary from p = 50 until p = 350 with step size of 50. For these experiments, discretization is excluded and a normalized Euclidean distance is used as the matching metric.

Table 1 shows a performance comparison of the FingerCode with BioPhasor at various p values in terms of AER (Average Error Rate)=(FRR+FAR)/2. It is clearly seen that BioPhasor in different-token case outperforms FingerCode significantly, as what we had expected to see in user-specific token mixing algorithm (Andrew et al, 2006). However, we also consider the stolen-token scenario which is generic in real world applications. We observed that the AER performance of BioPhasor in stolen-token scenario is slightly poorer than that of the FingerCode.

Table 1: **Performances of BioPhasor for various *p* and FingerCode in AER (%)**

p	different- token	stolen-token
50	12.64	30.12
100	10.21	28.12
150	5.13	27.21
200	0.64	27.01
250	0.71	25.11
300	0.02	25.74
350	0	25.54
FingerCode	22.35	-

For the experiments in sequel, we examine the performance of discretized bit strings without ECC. We fix p = 200 for both DB1 and DB2. We take an average of the final length of bit string template, k by appending zeros within those lesser than k and discard those which are longer than k for all bit strings. In this work, k is set at 246. We will address later why this specific value has been chosen.

It can be seen from Table 2, the near separation of genuine and imposter distributions which imply a low AER especially in the stolen-token scenario (0.4%) after discretization. However, the left tail of imposter distribution in the stolen-token scenario as shown in Figure 1 is stretched to near zero τ (τ = 0.025). This indicates that there is a potential danger of over-correction of imposter bit string when ECC is applied, even though the amount of these imposter bit strings has been small.

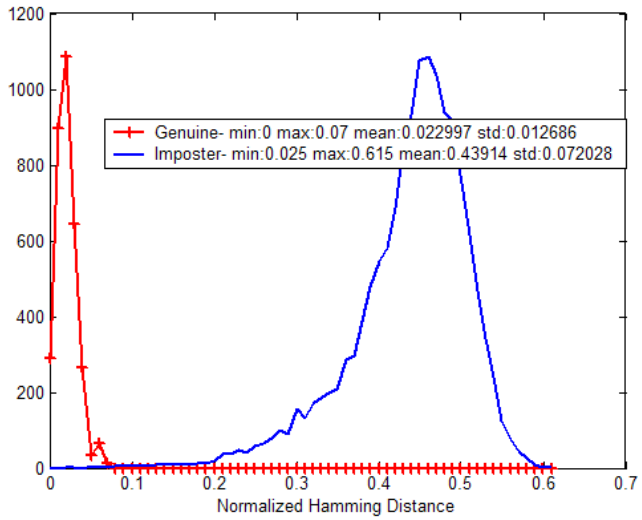

Figure 1: Genuine-imposter distribution of bit strings in stolen-token scenario before ECC.

To evaluate the success rate of an exact reproduction of FZH whenever auxiliary information such as the number bits per segment γ_{ij}, parity ρ and genuine fingerprint are presented. The FAR and FRR are redefined to elaborate the scenario of FuzzyHash production. Here, FRR_{FH} is defined as the error rate when a genuine user's h failed to be extracted, and the FAR_{FH} is defined as the error rate that a legitimate user's h is produced from an imposter's fingerprint. Note that for the stolen-token scenario, FAR_{FH} is the only indicator as we do not allow an adversary to generate the genuine-user's FH by using his own biometric and the stolen PRN.

Table 2: Genuine-imposter distribution statistics and AER(%) of bit strings before ECC

	Genuine (mean, std, min, max)	**Imposter (mean, std, min, max)**	**AER (%)**
diff-token	(0.02, 0.013, 0, 0.07)	(0.49, 0.035, 0.358, 0.627)	0
stolen-token	(0.02, 0.013, 0, 0.07)	(0.44, 0.07, 0.03, 0.62)	0.4

In the experiments, we try m_b = 5 and m_b= 6, hence $n_b = 2^5$-1= 31 and $n_b = 2^6$-1= 63 blocks.

Table 3 shows the list of possible ε codes that can be used. From Table 2, the maximum value of genuine scores is about 0.07. We may choose a suitable ε (n_b, k_b, t_b) where $t_b/k_b \geq 0.07$. From

Table 3, we notice that many codes can be selected since they all fulfill the condition, ie. $t_b/k_b \geq$ 0.07. However, we should choose the one which offers a sufficiently long bit string $m_b k_b$, as well as good performance. In this case, a moderate t_b/k_b such as RS(63, 41, 10) could be a good candidate. From this code, the length of h is large, ie. k=246 (41x6) bits and hence performs well. On the other hand, we should not use a high t_b/k_b, such as ε (31, 11, 10) and ε (63, 21, 20) as they introduce high FAR_{FH} and short bit string.

We should point out that the zero error rate obtained is based on the FVC 2004 subsets DB1 and DB2, and we do not claim that the proposed technique is perfect with error free performance in all circumstances. Typically, a high t_b/k_b is required to correct the errors if the genuine distribution is large. However, we do not know the genuine-imposter distribution a priori and thus an accurate choice of t_b/k_b is difficult to find. However, our proposed technique relaxes the ECC selection hassle to a certain extent as it provides a near complete separation of genuine-imposter distribution, even in the stolen-token scenario.

Table 3: Performances of FuzzyHashing in FAR and FRR with m_b=5 and m_b=6.

n_b	k_b	t_b	t_b/k_b	Bit length, k	FRR_{FH} (%)	FAR_{FH} (%)	FAR_{FH}(%) (stolen - token)
31	11	10	0.91	55	0	0.12	2.73
	15	8	0.53	75	0	0	0.01
	21	4	0.19	105	0	0	0
	27	2	0.07	135	2.8	0	0
63	21	20	0.95	126	0	0.26	3.64
	31	16	0.51	186	0	0	0
	41	**10**	**0.24**	**246**	**0**	**0**	**0**
	51	6	0.11	306	0.02	0	0

One-way Transformations

The examination of the one-way transformation of FZH lies in two important dispositions: (1) irreversible random extraction of biometric information via BioPhasor, f: $P^n \rightarrow P^p$ where $p<n$ and (2) transformation from real-valued BioPhasor features to binary domain, g: $P^p \rightarrow \{0, 1\}^k$ where $p \leq k$. The overall effect of FZH is a one-way transformation of the real-space biometric vector into binary-space hashes without compromising the biometrics itself. In summary, BioPhasoring, f: $P^n \rightarrow P^p$ has been shown irreversible for $p<n$ and hence it is non-commutative. The net effect of FZH transformation, T: P^n x $P^n \rightarrow \{0, 1\}^k$ is a one-way transformation function based on the product principle of Shannon (Shannon, C. E., 1949), which stated that a systematic cascading of different types of ciphers in single cryptosystems will increase the cipher strength provided that the product ciphers are associative but not commutative.

Randomness Requirements

To examine the randomness of the h, we empirically calculate the entropy of each bit string H(b) from the combined FVC2004 DB1 and DB2 dataset. Figure 2 depicts the average entropy distribution of 600 (200x3) bit string with k = 246. We have the empirical entropy bound of about [0.99 ± 0.05]x246=[231 246] bits. Compared to its ideal entropy of 246 bits, the degradation is not significant.

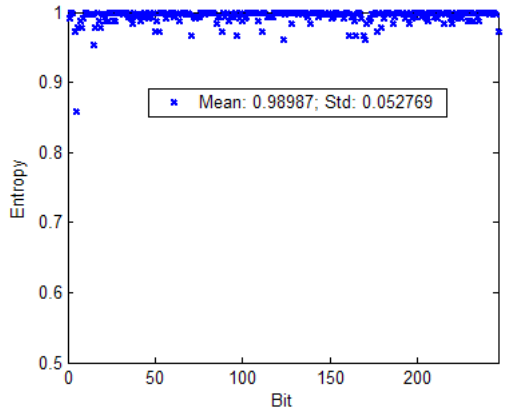

Figure 2: Average entropy distribution of h

We next examine the case when an adversary manages to acquire the stored information of discretization such as the segment bits number γ_{ij}, where the information leakage due to γ_{ij}. In other words, H(h |γ_{ij}) is considered. If this is a leakage, then H(h|γ_{ij})<H(h) where H(h) is the empirical entropy of h. However, no information is leaked from γ_{ij} since we do not store the statistics of feature elements, which may reveal the probable location of the genuine segment. Hence H(h |γ_{ij}) = H(h) ≈ 231 bits (the lower computed entropy bound).

We also consider the entropy lost due to parity checksum, ρ, H(h |ρ). According to section 4.3, we have H(h |ρ) = H(h) - $m_b\rho_b$. Thus, H(h |ρ) = 231 – 6(31-21) = 171 bits. Nevertheless, even if an adversary accesses ρ, the length of bit string is still sufficiently large (171 bits) to prevent a brute-force attack.

Diversity

Our scheme requires the users to enroll their biometric data for verification process. Biometrics, unlike any other types of identifiers, is unique for an individual as it contains sensitive information about that person. Thus, the sharing of biometric data across multiple applications represents a serious threat to privacy. In this context, the proposed method appears to be a promising solution to this problem as we allow multiple sets of independent external PRN to be tied with the same biometric data to yield a set of independent FuzzyHash for different applications. The template diversity provides more practical security system mechanism as any PRN could be securely revoked without affecting the others.

According to Teoh, Ngo and Goh (2007), the normalized hamming distance calculated from the comparisons of two uncorrelated binary bit strings with length k can be interpreted as a binomial distribution which has the functional form $f(x)=\frac{k!}{\lambda!(k-\lambda)!}0.5^k$, with expectation $\pi=0.5$ and standard deviation $0.5/\sqrt{k}$ where $x = \lambda/k$, is the fraction of bits that happen to agree when two uncorrelated bit strings are compared. Therefore both theoretical expectation and standard deviation values will be 0.5 and $0.5/\sqrt{(246)} = 0.032$ (k=246), respectively.

As shown in Figure 3, the empirical distribution closely resemblances the theoretical fractional Binomial distribution with expectation 0.49 and standard deviation 0.030. This implies that the generated FuzzyHash templates are nearly independent to each other, even when the same biometric data has been used. This is important to ensure that the *FuzzyHash* is revocable if it is compromised.

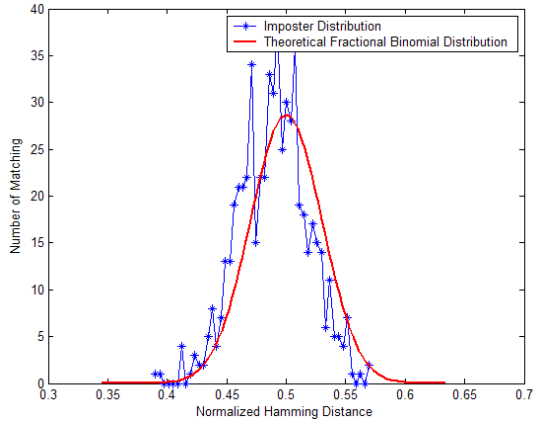

Figure 3: Diversity Property Evaluation

DISCUSSIONS

In general, there are some similarities shared between Biometric Encryption (BE) and FuzzyHashing (FZH): both do not use biometric templates and an exact matching is required. Identifiers in BE is externally derived and bounded, therefore it is revocable whenever an old one is compromised, whereas FZH depends on the changing of auxiliary information, i.e. user-specific pseudo-random numbers in BioPhasoring. Even though this characteristic is not obvious, it is testified by the strong diversity property of FuzzyHash. While FZH does not require a binding process, the binding process in BE poses the risk of recovering the biometric information if the binding is weak, such as the analysis reported in (Boult, Scheirer and

Woodworth, 2007). In FZH, the chain structure ensures that the transformation is one-way and the risk of recovery of biometric information is extremely low.

Our research works bears some resemblance to the works by Davida et al. (1998) in the sense that the techniques follow the **Gen**()/**Ext**() process. Davida et al. applied ECC to the Iris Code directly but we do it after BioPhasoring and discretization, with substantive security and stability enhancement.

Table 4 compares the proposed technique with various biometric template protection techniques discussed in Section 2. We do not include theoretical works such as those in Davida et al (1998) and Juels and Sudan (2002) which did not present an empirical validation. Since most of the techniques are biometric form-factor dependent, we present only the best reported results, including the techniques and database adopted. From Table 4, it is observed that our technique achieves the best performance in comparison to others.

In our technique, discretization facilitates "fastening" of intra-variation of feature vectors and at the same time repulses inter-class feature vectors via user-specific stored information, γ_{ij}. As a net effect, intra-class variations are suppressed while the inter-class variations are enhanced, and this brings a good separation of genuine-imposter distributions. Subsequently, ECC reduces the errors further. With small or no overlapping in genuine-imposter distribution, ECC can correct up to zero errors as demonstrated in the experiments. Therefore the proposed method is potentially applicable to other form of biometrics as long as the biometric feature is presented in a fixed length feature vector. It is noted that γ_{ij} is a user-specific data derived from biometric features; hence it is useless if an adversary snatches this piece of information without having the biometrics features, and the same concept is applied to parity checksum in ECC.

Table 4: Comparisons of various template protection techniques with FZH method

Biometrics	Technique	Bit length	Database	FRR	FAR
Fingerprint (Tuyls, 2005)	Gabor Filter + Reliable component extraction + Error Correction	76	FVC2000, 110 subjects	5.4%	5.2%
		40	Homemade, 500 subjects	5.4%	3.5%
Face (Kevenaar et al, 2005)	Gabor Filter + Reliable component extraction + Error Correction	58	Caltech, 24 subjects	3.5%	0%
		58	FERET, 237 subjects	35%	0%
Iris (Hao et al, 2006)	Concatenated Error Correction	140	Homemade, 70 subjects	0.41%	0%
Fingerprint (Nandakumar et al, 2007)	Random transformation + Fuzzy Vault	58 ~ 70	FVC2002, 100 subjects	10%	0%
			MSU-DBI 160	19.4%	0%

			subjects		
Proposed Method	BioPhasoring + Discretization	171	FVC 2004, 200 subjects	0%	0%

CONCLUSION

A secure biometric template protection technique – FuzzyHashing (FZH) was proposed in this paper, which consists of biometric transformation based on Bio-Phasoring, discretization and ECC. The design requirements of FuzzyHash include randomness, reproduction, diversity and one-way transformation. FZH comprise of two important components: auxiliary info extraction during enrollment and FuzzyHash transformation. The FuzzyHash can be repeatedly generated by using the user's biometric input and the auxiliary information. We presented a realization of FuzzyHash in terms of a composite function using fingerprint biometrics. Our experiments and security analysis on the fingerprint biometrics using a combination of FVC 2004 DB1 and DB2 databases suggest that FZH is highly feasible in practice.

REFERENCES

Boult, T. E., Scheirer, W. J., & Woodworth, R. (2007). Fingerprint Revocable Biotokens: Accuracy and Security Analysis. *IEEE Conf. of Computer Vision and Pattern Recognition*, 1–8.

Cavoukian, A., & Stoianov, A. (2007). Biometric Encryption: A Positive-Sum Technology that Achieves Strong Authentication, Security and Privacy. Retrieved from http://www.ipc.on.ca/index.asp?navid=67&fid1=31.

Chong, S. C., Teoh, A. B. J., & Ngo, D. C. L. (2005). Tokenised Discretisation In Iris Verification. *IEICE Electron*, *11*(2), 349–355.

Connie, T., Teoh, A. B. J., Goh, M. K. O., & Ngo, D. C. L. (2004). PalmHashing: A Novel Approach to Cancelable Biometrics. *Information Processing Letter*, *93*(1), 1–5.

Davida, G., Frankel, Y., & Matt, B. J. (1998). On enabling secure applications through off-line biometric authentication. *In Proc. IEEE Symp. Privacy and Security*, 148–157.

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

Jain, A. K., Prabhakar, S., Hong, L., & Pankanti, S. (2000). Filterbank-based fingerprint matching. *IEEE Trans. Image Processing*, *9*(5), 846–859.

Joy, T. M., & Thomas, J. A. (1991). Elements of Information Theory (2nd Edition). *John Wiley & Sons Inc*.

Juels, A., & Sudan, M. (2002). A Fuzzy Vault Scheme. *In Proceedings of IEEE International Symposium on Information Theory*.

Juels, A., & Wattenberg, M. (1999). A Fuzzy Commitment Scheme. *ACM Conference on Computer and Communications Security*, 28–36.

Kevenaar, T. A. M., Schrijen, G. J., van der Veen, M., Akkermans, A. H. M., & Zuo, F. (2005). Face recognition with renewable and privacy preserving binary templates. *Fourth IEEE Workshop on AutoID*, 21–26.

Nandakumar, K., Nagar, A., & Jain, A. K. (2007). Hardening Fingerprint Fuzzy Vault Using Password. *Lecture Notes in Computer Science*, *4642*, 927–937.

Purser, M. (1995). Introduction to Error-Correcting Codes. *Artech House, Boston*.

Ratha, N. K., Chikkerur, S., Connel, J. H., & Bolle, R. M. (2007). Generating cancelable fingerprint templates. *IEEE Trans Pattern Anal Mach Intell*, *29*(4), 561–572.

Ratha, N., Connell, J., & Bolle, R. M. (2001). Enhancing security and privacy in biometrics-based authentication systems. *IBM Syst. J.*, *40*(3), 614–634.

Savvides, M., Vijaya Kumar, B. V. K., & Khosla, P. K. (2004). Cancelable Biometrics Filters for Face Recognition. *Int. Conf. of Pattern Recognition*, *3*, 922–925.

Shannon, C. E. (1949). A mathematical theory of secrecy systems. *Bell System Technical Journal*, *28*(1949), 623–656.

Teoh, A. B. J., & Chong, T. Y. (2007). Cancellable Biometrics Realization with Multispace Random Projections. *IEEE T System, Man and Cybernetic B*, *37*(5), 1096–1106.

Teoh, A. B. J., & Kim, J. (2007). FuzzyHash: A Secure Biometric Template Protection Technique. *Frontiers in the Convergence of Bioscience and Information Technologies (FBIT 2007), Jeju Island, Korea*.

Teoh, A. B. J., & Ngo, D. C. L. (2005). Cancellable biometerics featuring with tokenised random number. *Pattern Recognition Letter*, *26*(10), 1454–1460.

Teoh, A. B. J., Goh, A., & Ngo, D. C. L. (2006). Random Multispace Quantisation as an Analytic Mechanism for BioHashing of Biometric and Random Identity Inputs. *IEEE Transactions on PAMI*, *28*(12), 1892–1901.

Teoh, A. B. J., Ngo, D. C. L., & Goh, A. (2004). Biohashing: Two Factor Authentication Featuring Fingerprint Data and Tokenised Random Number. *Pattern Recognition*, *37*(11), 2245–2255.

Teoh, A. B. J., Ong, T. S., & Ngo, D. C. L. (2003). Automatic Fingerprint Center Point Determination, *Lecture Notes in Artificial Intelligence*, *2903*, 633–640.

Teoh, A. B. J., Toh, K. A., & Yip, W. K. (2007). 2^N Discretisation of BioPhasor in Cancellable Biometrics. *Lecture Notes in Computer Science*, 435–444.

Tuyls, P., Akkermans, H. M., Kevenaar, A. M., Greert, J. S., Asker, M. B., & Veldhuis, N. J. (2005). Practical Biometric Authentication with Template Protection. *AVBPA 2005*, 436–446.