A Novel Privacy Enhancing Algorithm for Biometric System

Xuebing Zhou
xuebing.zhou@igd.fraunhofer.de
Christoph Busch
christoph.busch@igd.fraunhofer.de

Abstract: Biometrics provide the possibility of unique and convenient authentication. As its application areas grow rapidly, the problems, such as identity fraud and cross matching, threaten the security of biometric systems. Moreover, privacy concerns are associated with biometrics. In this paper we focus on privacy enhancing techniques for biometric systems, which can protect biometric information and enable using biometrics without exposure of privacy. A novel privacy enhancing algorithm is proposed. The algorithm is integrated in a 3D face recognition algorithm and tested using the FRGC database. By using the proposed algorithm, high security can be achieved as well as good verification performance.

1 Introduction

As application areas of biometrics widely broaden, security and privacy risks have attracted the attention not only of the biometric community. When applying biometric systems, private information of data subjects that is not relevant for authentication or verification purposes is available and retrievable. Among these private data can be gender, race, gene and even information about diseases. From the privacy legislation point especially in Europe, collection of such data is very critical. As people more and more appreciate the increased convenience of biometric systems, new kinds of associated security problems must not be ignored. In contrast to password or token- based authentication, biometrics utilize the unique physiological or physiological characteristics to authenticate a user's identity. Although these characteristics can not be forgotten or handed over, they can be faked. If a biometric sample or template of a data subject is obtained by an attacker, it is possible to create a fake biometric modality. Then, the biometric identity of the user can be compromised. And what is worse, biometric characteristics are unique and limited. Revocation and renewal are impossible. Additionally, the usage of the same biometric characteristic in different applications enables cross matching. These applications can be linked together for a potential attack. This way, a data collector can retrieve information about the user's activities within another application. Furthermore, storages of biometrics data becomes very critical. If a stored template is exposed, the biometric authentication in all applications using the same biometric modality is endangered.

In order to prevent the above mentioned security and privacy risks, privacy enhancing techniques are required. Their goal is to provide a generalized solution to protect stored biometric data. Privacy enhancing techniques convert biometric template into a secure reference with help of random variables. It is possible to derive numerous uncorrelated references from one single biometric characteristic. The resulting secure reference reveals very few information of the original biometric template; meanwhile, it is robust to biometric variation. Secure references are compared directly and recovering the original template is not necessary. Using such a technique, no information that would disclose any properties of the biometric characteristic is available within the system. Thus, the privacy of the data subject is preserved. The creation of a fake biometric characteristic from the stored reference is infeasible. Its diversity enables to generate distinct references for different applications and cross matching between different applications is prevented. Similarly the revocation or renewing of a template is possible.

Different approaches of privacy enhancing techniques exist. One of the ideas is to combine cryptography with error correction coding so that cryptographic hashing can be applied to noisy data such as biometric data [JW99, JM02, DRS04]. The realization of this idea strongly depends on the properties of extracted biometric features. For ordered features, where a number of components are stable, the fuzzy commitment scheme is proposed as shown in [JW99, DRS04]. In [Tul04]' [TAK$^+$05, LT03], the helper data scheme is introduced to construct the fuzzy commitment. The security of this schemes is proved in [TG04]. The algorithm was integrated in a 2D face recognition system using texture information [vdVKS$^+$SA], a fingerprint recognition system [VTDL03] and an ear identification system [TVI$^+$04]. For non-order features like minutiae of fingerprints, whose components vary and can not be described as a vector, the fuzzy vault scheme can be adopted [UJ04]. Another possible approach is cancelable biometrics, which utilizes "non-invertible" function as scrambling, morphing [RCB02, BCR04]. The employed function must satisfy the"non-invertibility" and authentication performance of the biometric system should not be reduced.

This paper is organized as follows. In section 2, a privacy enhancing algorithm is introduced. The important components such as binarization and error correction are elaborated. A security analysis is given. Section 3 shows the experimental results of integrating the algorithm in a 3D face recognition system. Finally, conclusions are given in section 4.

2 A Novel Privacy Enhancing Algorithm

In this section, a novel privacy enhancing algorithm based on an error correction coding, polynomial and cryptographic function is proposed. The proposed algorithm can be adopted in any biometric recognition system, whose features can be described as vectors.

A component that helps to produce a random template is necessary. In the presented algorithm, a random number generator is integrated in the enrolment module so that a random secret code can be created. The created random code is used to set the coefficients of a polynomial in a Galois Field. The input biometric feature vector is also converted into a discrete codeword in the Galois Field. Each component of the codeword is in terms taken as input value for the polynomial and the corresponding output is calculated. In other words, the biometric feature vector is transformed into a codeword in the Galois Field and

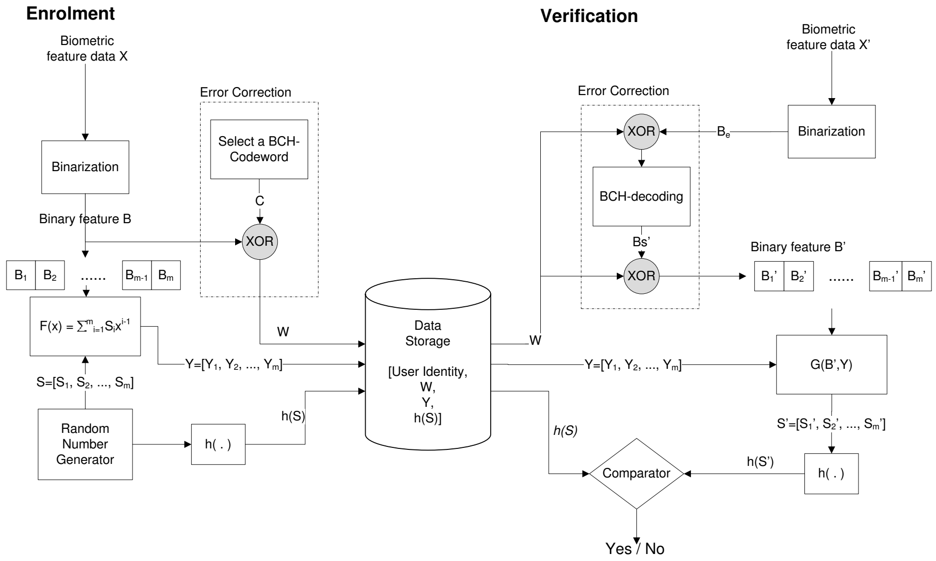

Figure 1: The block diagram of the privacy enhancing algorithm

projected in a new feature space using the polynomial. The hash value of the secret code and the output of the polynomial are stored as a template in the system. In order to realize the required robustness of the algorithm, an error correction coding method is used to correct errors, which may occur in measurement.

In figure 1, the block diagram of the algorithm is depicted. The enrolment process is shown as follows:

1. Provide a biometric feature vector X as input. Convert X into a binary feature vector B in the binarization process.

2. Randomly select a BCH-codeword C. Calculate the reference $W = B \otimes C$, which indicates the distance between B and the codeword C.

3. Divide the binary feature B into m blocks. Each block represents a k-bit long symbol, where the first l bits are reserved to indicate the original location of each bit blocks in B. And $l = ceil\,\{log_2(m)\}$.

4. Randomly generate a secret code S containing $k \times m$ bits and divide S into m blocks, $S = [S_1, S_2, \cdots, S_m]$.

5. Convert B and S into the codes in the Galois Field.

6. Generate a polynomial $F(x) = \sum_{i=1}^{m} S_i \cdot x^{i-1}$, where S are the coefficients of the polynomial.

7. Transform B with the polynomial $F(x)$ into $Y = [Y_1, Y_2, \cdots, Y_m]$, where

$$\begin{aligned} Y_i &= F(B_i) \\ &= S_1 + S_2 \cdot B_i + \cdots + S_m \cdot B_i^{m-1} \end{aligned} \qquad (1)$$

8. Store Y, W and the cryptographic hash $h(S)$ of S with the identity of the user.

In converting binary features into bit blocks, the location bits are used, which guarantees that the bit blocks differ from each other. The polynomial with the degree of m can only be reconstructed with at least m different support points.

The verification process contains the following steps:

1. X' is a new biometric feature vector. Binarize X' into B_e.
2. Calculate $B' = B'_s \otimes W$, where B'_s is the corrected code of $B_e \oplus W$.
3. Divide binary feature B' into m k-bits-long blocks $[B'_1, B'_2, \cdots, B'_m]$ and convert B' into the Galois Field.
4. Find Y in the data storage and reconstruct the polynomial F with the set

   $$\{(B'_1, Y_1), (B'_2, Y_2), \cdots, (B'_m, Y_m)\}$$

   where $S' = G(B', Y)$.
5. Compare the hash $h(S')$ with the stored $h(S)$. If they are exactly the same, a positive response will be given, otherwise, a negative response.

The key components as well as the security of the algorithm are discussed in the following sections.

2.1 Error Correction

In the proposed scheme, the BCH-coding method is adopted to correct bit errors occurring in the verification. The BCH-code is the most popular and efficient linear block code. If the error probability of individual bits in a codeword is uniformly distributed, the BCH-code can achieve optimal coding rate. The error probability relies on the input biometric features and the binarization method. Dependent on the statistical characteristics of the bit errors, other coding methods might achieve a better biometric performance.

The stored reference W in the algorithm is an indicator showing the distance of the newly enrolled binarized feature to a BCH-codeword. During verification, a corrupted binarized feature B_e is obtained. In order to correct an error in B_e, the "corrupted" BCH-codeword can be estimated by $B_e \oplus W$. After the error correction, the corrected BCH-codeword B'_s is obtained. The corrected binarized feature can be calculated by translating B'_s with

W. If E represents the errors in the binarized feature B_e, equation 2 shows that the errors between B'_s and C are equal to E:

$$C \otimes B'_s = (B \otimes W) \otimes (B_e \otimes W) = B \otimes B_e = E \tag{2}$$

If the hamming weight of E is small enough and within the error correction capability of the BCH-code, the same binary feature vector as in the enrolment can be obtained and the user can be successfully verified.

2.2 Binarization

The binarization process converts a biometric feature into a binary form and strongly impacts the recognition performance of the algorithm. The details of the binarization process are described in [Zho07]. The resulting binary feature for different users should be uniformly distributed in order to be resistant to brute force attacks. Since a linear code is used in the algorithm, correcting one bit error requires at least two bits. Therefore, the binarization should make the resulting features very robust to intra class variations of the input biometric features. Using the algorithm, the interconnection of the secret code and biometric features is relaxed. The statistical independency of bits in the binary feature vector is not strictly required.

2.3 Security

As shown in section 2.1, W is supplementary data for the error correction step but little information about B can be retrieved from W. This can be proved as follows: Assuming that $H(B|W)$ is the conditional entropy of B for a given W. And B is statistically independent from the randomly selected codeword C, then $H(B, C) = H(B) + H(C)$. Since $W = B \oplus C$, the following equations are valid:

$$\begin{aligned} H(B,C,W) &= H(B,W) = H(B,C) = H(B) + H(C) && (3)\\ H(B,W) &= H(B) + H(W|B) \\ H(W|B) &= H(C) && (4) \end{aligned}$$

And the entropy of a variable is always greater or equal to its conditional entropy, namely $H(W) \geq H(W|B)$. Additionally W is a binary vector with the length of $N = (k-l) \times m$.

$$\begin{aligned} N &\geq H(W) \geq H(C) && (5)\\ H(B,W) &= H(W) + H(B|W) = H(B) + H(C) \\ H(B|W) &= H(B) + H(W|B) \\ H(B) &\geq H(B|W) \geq H(B) + H(C) - N && (6) \end{aligned}$$

If W is known, the uncertainty of B is equal to $H(B) + H(C) - N$. $H(B)$ is the entropy of the binary features, which is not greater than the entropy of the used biometric modality. $H(C)$ indicates the codeword space of C, which is dependent on the required error correction capability.

The stored reference Y is the polynomial projection of B with the secret code S. Retrieving B over Y needs the exact information of S, and vice versa. The security of the hash function is dependent on the length of the secret code. The complexity of a successful brute force attack on the secret is $2^{k \times m}$.

3 Experimental Results

The proposed algorithm has been implemented within a 3D face recognition system. Face recognition has very high user acceptance. Additionally 3D face recognition adopts the rich geometric information of face surfaces and is robust to light and pose variation. In comparison to 2D face recognition, 3D facial information is more difficult to obtain. Measuring 3D information requires the cooperation of the capture subject and is therefore resistant to spoofing attacks. In the experiment, a histogram-based algorithm is used (for more details about the feature extraction algorithm see [ZSBF08]). The system has been tested using the face recognition grand challenge (FRGC) database version 2 [FPea05]. The database consists of 4007 range images acquired from 477 users.

The biometric features extracted with the histogram-based algorithm contains 476 components. 300 users were chosen randomly to find the binarization threshold. In figure 2, the false match rate (FMR) and the false non-match rate (FNMR) curves are depicted. After the binarization, the curve of FMR shifts to the right and the robustness reduces slightly. In figure 3 the receiver operating characteristics (ROC) curves shows an overview of the recognition performance change. The dashed ROC curve of the binarized features is below the solid one representing the real valued features. It indicates also a slight performance degradation.

In the experiment of the privacy enhancing algorithm, N_{enrol} samples of a data subject, who has more than N_{enrol} samples, are chosen randomly for the enrolement. The remaining samples serve as verification samples for the performance evaluation. The experiment is repeated 5 times for each N_{enrol} in order to get reliable results. The 255 most reliable bits are selected from the features of each data subject, since the length of the BCH-code is $2^N - 1$, where N is a positive integer. The polynomial in Galois Field of (2^4) is used. The length of the secrete code is 368. For operational points corresponding to different error correction code configurations are tested. The simulations are displayed in table 1.

Increasing error correction capability improves the FNMR and thus the robustness of the algorithm, however it reduces the FMR, the discriminative power. The performance of $N_{enrol} = 5$ is better than $N_{enrol} = 3$, which indicates that enlarging the number of enroled samples can improve the recognition results.

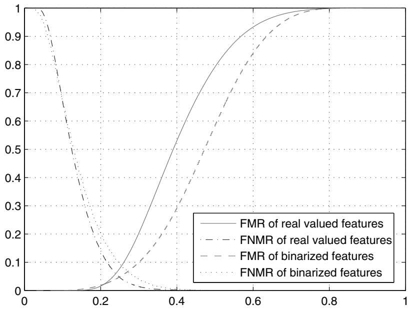

Figure 2: The FMR and FNMR curves of the real valued features and binarized features

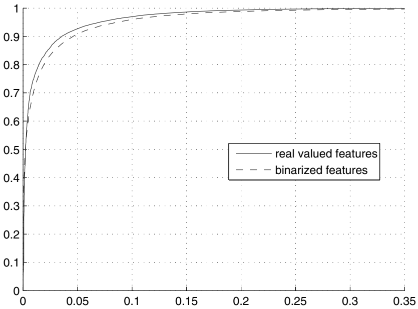

Figure 3: The ROC curves of the real valued features and binarized features

BCH ($c/s/e$)	BER	$N_{enrol} = 3$	$N_{enrol} = 5$
255/107/22	8.6%	FNMR=1.36%; FMR=20%	FNMR=1.78%; FMR=15%
255/91/25	9.8%	FNMR=1.71%; FMR=16.5%	FNMR=2.2%; FMR=12.5%
255/79/27	10.5%	FNMR=1.98%; FMR=15.0%	FNMR=2.57%; FMR=11.3%
255/63/30	11.7%	FNMR=2.42%; FMR=12.3%	FNMR=3.09%; FMR=9.6%

Table 1: Examples of possible BCH codes and the corresponding FNMR and FMR, where c is the length of codeword, s is the length of message and e is the number of correctable bit errors.

4 Conclusion

In this paper, the security and especially privacy leakage of biometric systems are discussed. To efficiently preserve privacy and to improve the security of biometric systems a novel privacy enhancing algorithm is proposed. Its scheme and the individual components are discussed. The properties and requirements of the key components of the algorithm, binarization and error correction coding, are introduced. The security of the proposed algorithm is analyzed. The algorithm was implemented in a 3D face recognition system. The results of the conducted experiment show that a high security level is achieved with reasonable false match and false non-match rates.

References

[BCR04] R. Bolle, J. H. Connell, and N.K. Ratha. System and method for distorting a biometric for transactions with enhanced security and privacy. US 6836554 B1, Dec 2004.

[DRS04] Y. Dodi, L. Reyzin, and A. Smith. Fuzzy Extrators: How to generate strong secret keys from biometrics and other noisy data. In LNCS 3027, editor, *In Advances in cryptology - Eurocrypt'04*, pages 523–540, 2004.

[FPea05] Patrick J. Flynn, P.Jonathon Phillips, and et al. Overview of the face recognition grand challenge. In *IEEE conference on computer vision and pattern recognition*, 2005.

[JM02] A. Juels and M.Sudan. A fuzzy vault scheme. In *IEEE International Symposium on Information Theory*, http://biometrics.cse.msu.edu/uludag-jain-fuzzy-fp.pdf, 2002.

[JW99] A. Juels and M. Wattenberg. A fuzzy commitment scheme. In *6th ACM Conference on Computer and Communications Security*, pages pp. 28–36., http://www.rsasecurity.com/rsalabs/node.asp?id=2048, 1999.

[LT03] J. P. Linnartz and P. Tuyls. New shiedling functions to enhance privacy and prevent misuse of biometric templates. In *4th international conference on audio- and video-based biometric person authentication*, 2003.

[RCB02] N.K. Ratha, J.H. Connell, and R. Bolle. Enhancing security and privacy of biometric-based authentication systems. *IBM Systems Journal*, 40:No. 3, 2002.

[TAK$^+$05] Pim Tuyls, A.H. M Akkermans, T.A.M. Kevenaar, G.J.Schrijen, A.M.Bazen, and R.N.J.Veldhuis. Practical biometric authentication with template protection. In *Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 436–446, 2005.

[TG04] P. Tuyls and J. Goseling. Capacity and examples of template protecting biometric authentication systems. In LNCS, editor, *Biometric authentication workshop (BioAW 2004)*, number 3087, pages 158–170, Prague, 2004.

[Tul04] P. Tulys. Privacy protection of biometric templates: cryptography on noisy data. In *Revue HF (Rev. HF) ISSN 0035-3248*, no3, pages 55–64, 2004.

[TVI$^+$04] P. Tuyls, E. Verbitskiy, T. Ignatenko, D. Schobben, and T. H. Akkermans. Privacy protected biometric templates: ear identification. In *Proceeding of SPIE*, volume 5404, pages 176–182, April 2004.

[UJ04] U. Uludag and A. Jain. Fuzzy fingerprint vault. In *Workshop: Biometrics: Challenges Arising from Theory to Practice*, citeseer.ist.psu.edu/uludag04fuzzy.html, August 2004.

[vdVKS$^+$SA] Michiel van der Veen, Tom Kevenaar, Geert-Jan Schrijen, Ton H. Akkermans, and Zuo Fei. Face biometrics with renewable templates. In *Security, Steganography, and Watermarking of Multimedia Contents VIII*, 1519 January 2006, San Jose, California, USA.

[VTDL03] E. Verbitskiy, P. Tuyls, D. Denteneer, and J.-P. Linnartz. Reliable Biometric Authentication with Privacy Protection. *24th Benelux Symp. on Info. Theory*, 2003.

[Zho07] Xuebing Zhou. Template Protection and its Implementation in 3D Face Recognition Systems. In *Biometric Technology for Human Identification IV*, Orlando, Florida, USA, 2007.

[ZSBF08] Xuebing Zhou, Helmut Seibert, Christoph Busch, and Wolfgang Funk. A 3D Face Recognition Algorithm Using Histogram-based Features. In *Eurographics Workshop on 3D Object Retrieval*, pages 65–71, Crete, Greece, 2008.