Non-invertible Biometric Encryption to Generate Cancelable Biometric Templates

Harkeerat Kaur and Pritee Khanna*

***Abstract*—Tremendous use of biometric based authentication systems has led to increased attacks on biometric data. It has given rise to several concerns regarding security and privacy of biometrics data. Biometrics are immutable and limited resources; and once compromised renders them insecure for further usage. To address these concerns Ratha et al. (2001) proposed the concept of cancelable biometrics, which transforms a biometric data and then uses it for storing and matching purposes. This work proposes a Hill cipher based technique to transform biometric signals. Experiments are performed on face and palmprint biometric modalities and important criteria like security, diversity, non-invertibility, and performance are thoroughly analyzed to showcase the effectiveness of the proposed approach.**

***Index Terms*—Cancelable biometrics, Random Projections, Hill Cipher, Non-invertible, Revocable.**

I. Introduction

Biometric based authentication systems are widely used for granting access in various commercial applications. The increased use of biometric applications have risked its security and given rise to many privacy concerns. Biometric are unique and limited to every individual; and once compromised are lost for entire lifetime. Database attacks are common means for digital theft of biometric identity, which is being used by attackers to make illegitimate access. Cross matching of biometric database to track individuals is another rising concern. Also, biometric reveals a lot of personal information, e.g., medical conditions. Templates protection schemes are proposed to address such security critical issues. Cancelable biometrics is one of the recent template protection schemes proposed by Ratha et al. in 2001 [1].

Cancelable biometric techniques repeatedly distort the biometric to generate their transformed versions which are later used for storing and matching purposes. Transformed biometric does not reveal any information about the original biometric and can be easily revoked in case of compromise by simply changing the transformation function/parameters. Also, different transformed versions of the same biometric can be generated for different applications, thus preventing cross matching attacks. The four important criteria to be fulfilled by any cancelable biometric technique are - *security*, *diversity*, *non-invertibility*, and *performance*.

The transformation techniques can be broadly classified into two categories - *non-invertible transforms* and *biometric salting*. Non-invertible transforms are one way surjective functions that map original biometric feature into a new subspace. The security of these transforms lie in the fact that they are non-invertible and the transformed biometric cannot be reverted to obtain the original even in case of an attack. Ratha et al. designed Cartesian, polar, and surface folding based non-invertible transforms for fingerprint data [2]. Non-invertible transforms are secure, but they tend to compromise the discriminability of biometric features.

Biometric salting blends biometric data with some user-specific auxiliary data to generate its transformed versions. Teoh et al. proposed BioHashing technique which projects biometric feature into a dimensionally reduced random subspace to generate its transformed version [3]. The projected features are binarized via thresholding to generate binary vectors called biocodes. However, BioHashing is invertible and performance degrades in stolen token scenario [4]. Teoh and Yaung and Lumini et al. suggested various improvements to enhance the security and non-invertibility of BioHashing by using multiple random projection and variable thresholding techniques[5], [6]. Biometric salting techniques effectively preserve the discriminability but are invertible.

It is essential to design a non-invertible technique that is able to discriminate transformed biometric features. This work proposes generation of cancelable templates based on a non-invertible encryption technique using Hill cipher algorithm. The biometric template is encrypted in such a way that it is not possible to decrypt it even if the encryption keys are available. The encrypted template is then used for storing and matching purposes. Matching is performed to evaluate the discriminability of the transformed templates and other aspects such as non-invertibility and diversity are also analyzed. The organization of the paper is as follows. Section II briefly discusses Hill cipher and the proposed approach. Experimental results are covered in Section III, and finally Section IV concludes the work.

II. Template Transformation

A. Hill Cipher

Lester S. Hill invented Hill cipher, a polygraph substitution cipher, based on linear algebra in 1929. Hill cipher has many advantages in data encryption. It is simple (uses matrix multiplication) and resistant to frequency analysis. Also, it provides high speed and high throughput. Hill cipher works on the blocks of data by displaying them as a vector. Let P be a block plaintext data, which is to be encrypted and K is the key [7]. The ciphertext C is obtained by matrix multiplication (or projection) of P over K given as

$$\begin{bmatrix} C_1 \\ . \\ . \\ C_n \end{bmatrix} = \begin{bmatrix} K_{1,1} & . & . & . & K_{1,n} \\ . & . & . & . & . \\ . & . & . & . & . \\ K_{n,1} & . & . & . & K_{n,n} \end{bmatrix} * \begin{bmatrix} P_1 \\ . \\ . \\ P_n \end{bmatrix} \bmod N \quad (1)$$

H. Kaur (e-mail: harkeerat.kaur@iiitdmj.ac.in) and P. Khanna (phone: +91761-279-4222, e-mail: pkhanna@iiitdmj.ac.in) are with the Department of Computer Science and Engineering, PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, India.

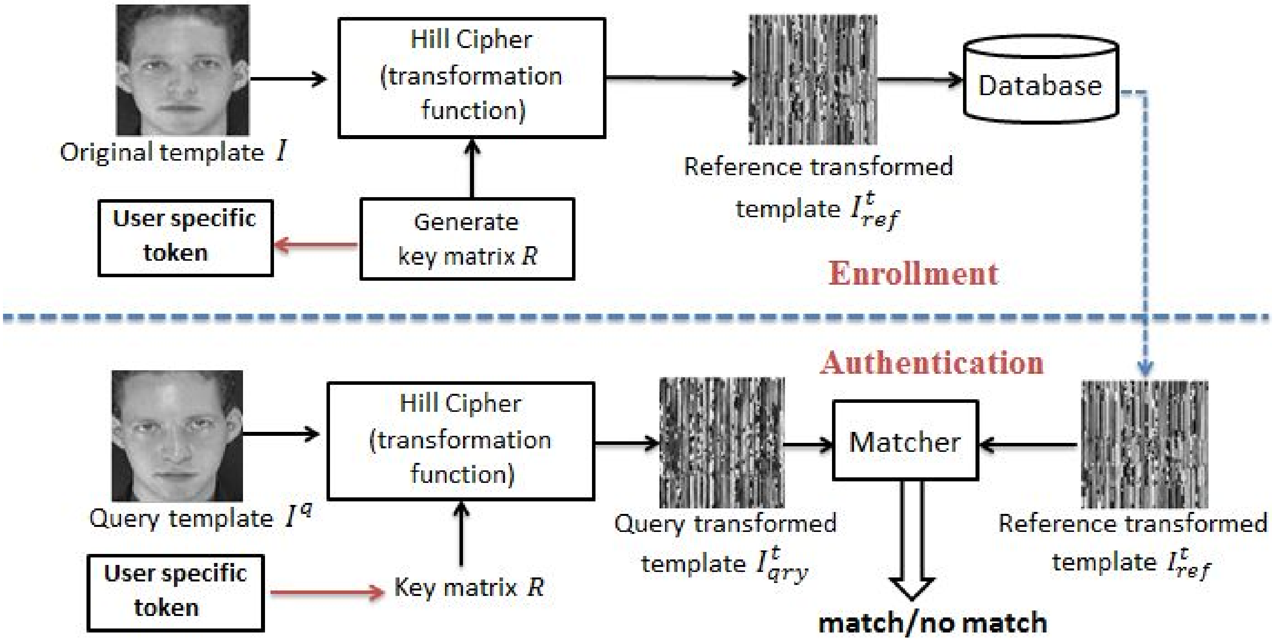
Hill Cipher (transformation function)
Original template I
Reference transformed template I^t_{ref}
Database
User specific token
Generate key matrix R
Enrollment
Authentication
Hill Cipher (transformation function)
Query template I^q
Query transformed template I^t_{qry}
Matcher
Reference transformed template I^t_{ref}
User specific token
Key matrix R
match/no match

Fig. 1. Enrollment and authentication processes for the proposed approach.

where $N = 26$ for alphabets and $N = 256$ for gray level images. Eq. 1 can also be written as

$$C_1 = (K_{1,1} * P_1 + ... + K_{1,n} * P_n) \bmod N \quad (2)$$

$$C_n = (K_{n,1} * P_1 + ... + K_{n,n} * P_n) \bmod N \quad (3)$$

Inverse of the key matrix (K^{-1}) is required to decrypt the ciphertext given as

$$P = (K^{-1} * C) \bmod N \quad (4)$$

The first and foremost condition for the inverse to exist requires the key matrix K to be a square matrix. Also, the computed inverse matrix K^{-1} must be integral to recover the encrypted data losslessly. If K^{-1} contains fractional entities, then the original information cannot be recovered losslessly. The determinant of the encryption matrix K must not be zero and should also be relatively prime to N.

B. Proposed Approach

The proposed technique performs a one way hashing on the biometric template by making use of the non-invertibility of Hill cipher algorithm. Let, I be the biometric image signal of dimensions $k \times d$. Hill cipher is applied to encrypt the data by using orthonormal random matrices of dimensions $l \times k$ ($l \leq k$) as keys. The elements of orthonormal matrix R have fractional values between [-1,1]. The encrypted image I^{RP} is obtained as

$$I^{RP} = (R * I) \bmod N \quad (5)$$

Eq. 5 is similar to Eq. 1 for the Hill encryption algorithm, where I is the plaintext, I^{RP} is the ciphertext, and R is the key matrix. As the acquired signal I is a grayscale image, $N = 256$. The key matrix R is the orthonormal matrix generated from the tokenized psuedo-random number using Gram-Schmidt orthogonalization [3]. Projection of data on orthonormal random matrix is a distance preserving mapping also known as random projection [8]. Its key concepts are defined using *JL-Lemma* [9], [10].

Fig. 1 shows the block diagram for the proposed approach. During enrollment, every user is assigned a different user specific key projection matrix R. Any random key matrix of appropriate dimensions can be used for encryption. The biometric is transformed using the proposed approach and the transformed version is stored in the database as reference template I^t_{ref}. At authentication, the query template is distorted in the similar manner and using the same user specific key, I^t_{qry}. Finally, matching is performed between the transformed reference I^t_{ref} and query template I^t_{qry} to declare a match or non-match. In case of a compromise, the projection matrix R can be changed to easily generate a new transformed template.

III. Experimental Results

A. Performance Evaluation

Effectiveness of the proposed approach is tested on two biometric modalities, namely face and palmprint. ORL, Extended Yale Face Database B, and Indian face database are three standard face databases that capture expression and orientation variations. ORL contains 40 subjects with 10 samples per subject [11]. YALE is an illumination variant database for 38 subjects from which 10 samples having uniform light variations are selected per subject [12]. Indian face database contains 61 subjects from IIT Kanpur with 11 samples per subject [13]. For each database a training and testing database is constructed by randomly selecting any 3 and 7 images respectively out of the available samples. For palmprints, CASIA and PolyU databases are used. CASIA contains 5,239 palmprint image samples for 602 subjects [14]. PolyU contains 600 samples for 100 subjects, with 6 samples per subject [15]. Training and testing set for each palmprint database is constructed by randomly selecting 2 and 4 sample images, respectively.

Performance is evaluated as the recognition accuracy when system uses transformed templates instead of original ones

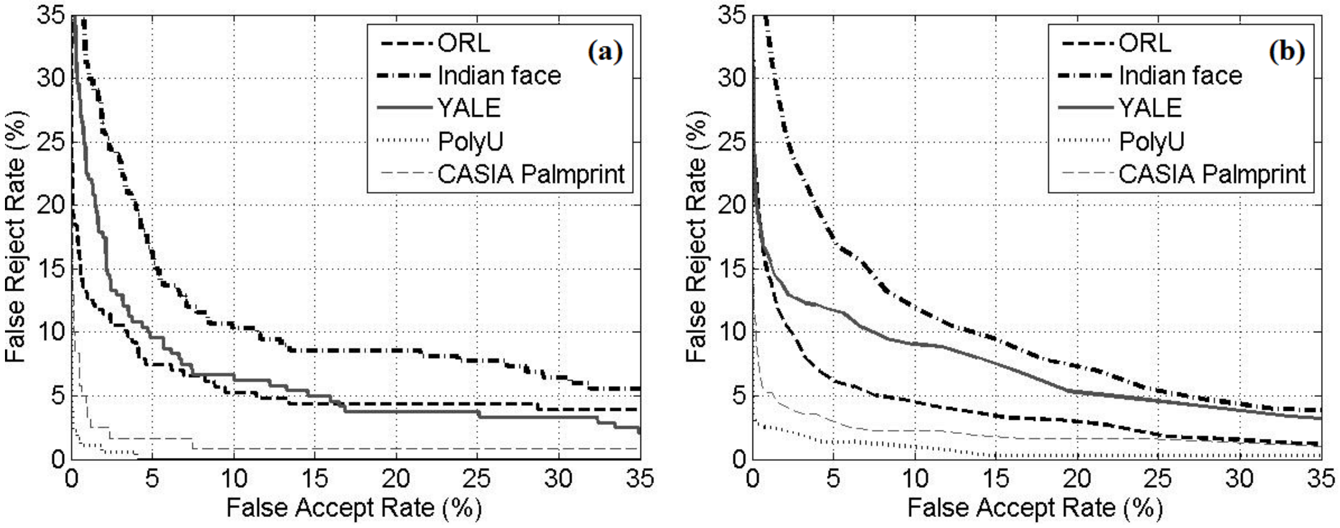

Fig. 2. ROC curves for matching performance (a) original domain (b) transformed domain.

TABLE I
MATCHING PERFORMANCE FOR FACE AND PALMPRINT TEMPLATES

Modality	Database	Original (Untransformed)		Transformed (Worst Case)	
		EER	DI	EER	DI
Face	ORL	5.42%	3.133	7.12%	3.347
	Indian Face	9.31%	2.398	11.13%	2.312
	YALE	7.59%	2.612	9.95%	2.521
Palmprint	PolyU	0.62%	7.569	1.02%	8.235
	CASIA	2.34%	5.083	2.97%	4.023

for matching purposes. The two standardized indices used for measuring performance are Equal Error Rate (EER) and Decidability Index (DI). While EER determines the probability of false acceptance and rejection, DI measures separability of genuine and impostor population [16]. A combination of lower EER and higher DI values indicates good performance. The template transformation depends upon user-specific secret key. Ideally, each user is assigned a different secret key R to generate transformed templates. In case of ideal scenario, when each user is assigned a different key R, the matching performance results in nearly 0% EERs. It is due to the increased inter-user variations obtained by projecting each user on a different random subspace.

However, in real life it is quite easy for one to lose its user-specific key which can be used by an attacker to intrude the system. To test the performance of the proposed approach in stolen key scenario, each user is assigned the same secret key R to generate transformed templates. The condition is similar to public-key encryption where each template is transformed on the same scale and hence, it is expected that the performance should be similar to a conventional non-cancelable system. Therefore, matching is performed between original untransformed templates and transformed templates using Linear Discriminant Analysis and the results are reported in Table I. It can be observed that the matching performance of the proposed approach in stolen key scenario is comparable to conventional system that operates on original templates. The ROC curves are shown in Fig. 2. The DI values are also sufficiently high which indicate good separability between genuine and imposter populations in the transformed domain.

B. Non-invertibility Analysis

An essential criteria of cancelability is the security of transformed templates. It requires that even if the transformed template, transformation function, and key are known to the attacker; he/she should not be able to learn any information about the original biometric template. The proposed approach performs Hill encryption using orthonormal key matrices R. Restricting R to orthonormal set of matrices is for two particular reasons. Firstly, according to the *JL-Lemma* the pair-wise distances between points of original image are preserved after projection on the subspace R and even after modulus N hashing. Thus the discriminating capability is not lost by the transformed template. Another property possessed by orthonormal matrices is the existence of inverse $R^{-1} = R^T$[17]. Although R is not chosen as a square matrix to prevent invertiblity of Hill cipher. However, if R is a square matrix, its inverse will always posses fractional values between [-1,1] and determinant 1. Both the conditions does not allow a lossless recovery of information on decryption. Such a selection of key matrix helps in achieving non-invertibility, thus being the second reason. Fig. 3 depicts some samples of original, encrypted, and decrypted images. The decrypted images are very noisy and do not reveal original biometric.

C. Diversity Analysis

Diversity is the ability to generate new transformed template by changing the transformation parameter (random projection matrix R). Transformed templates generated from the same biometric sample by changing parameters are diverse if their mutual information content is low, i.e., they should not correlate. The correlation between any two transformed templates is calculated as

$$C_r(T_1, T_2) = \frac{\sum\sum(T_1 - \bar{T_1})(T_2 - \bar{T_2})}{\sqrt{(T_1 - \bar{T_1})^2 + (T_2 - \bar{T_2})^2}}, \qquad (6)$$

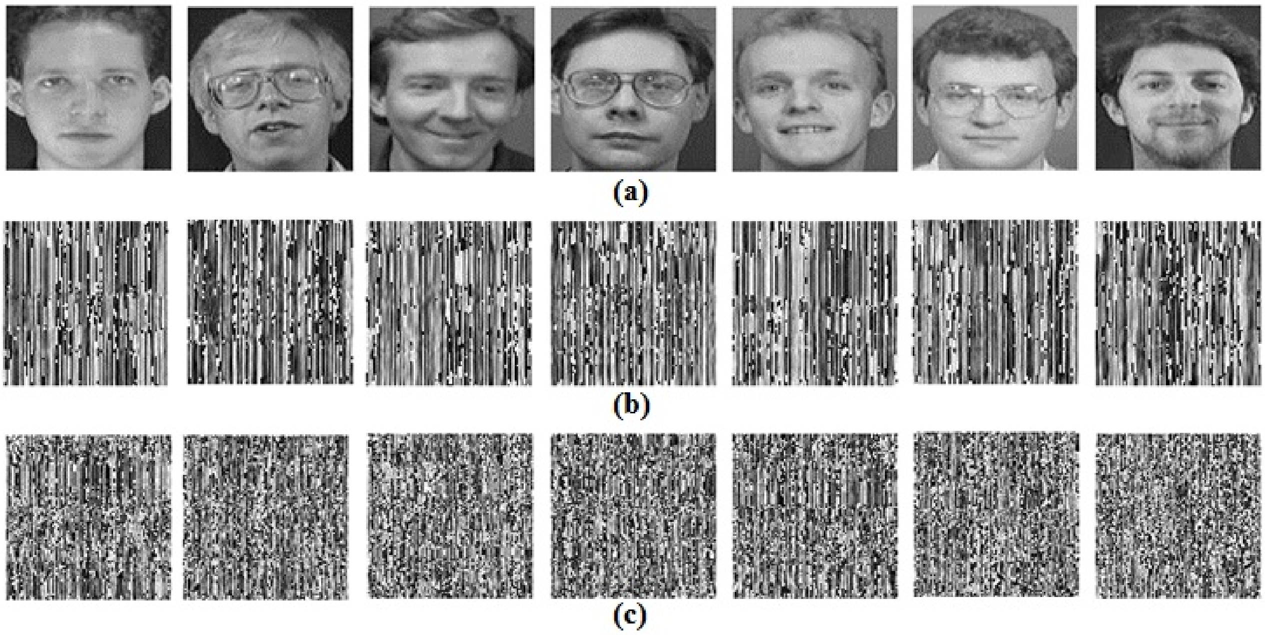

Fig. 3. Image samples (a) original (b) encrypted and (c) recovered.

TABLE II
CORRELATION INDEX VALUES FOR DIFFERENT DATABASES

	Face			Palmprint	
Database	ORL	Indian face	YALE	PolyU	CASIA
CI (%)	15.9%	14.1%	12.3%	11.5%	14.6%

where $\bar{T_1}$,$\bar{T_2}$ represents the mean of templates T_1, T_2, respectively. Here, a set of ten different transformed templates is generated for each database and correlation C_r is calculated between each pair of transformed templates. The mean of C_r values over a database is defined as correlation index (CI), which determines the percentage of mutual information content. Table II provides percentage of mutual information content for different modalities and databases. It is observed from Table II that the values are low, which indicate low mutual information content and good diversity.

IV. CONCLUSION

The proposed approach is tested for face and palmprint modalities to analyze its performance on matching, non-invertibility, and diversity. It is found that the proposed approach delivers good matching performance for both stolen token and legitimate key scenarios. To test non-invertibility, the templates are decrypted and the recovered templates are found to be noisy which justifies the security of the approach. Revocability and diversity analysis indicate that new transformed templates can be generated by assigning a different random projection matrix R to a user.

REFERENCES

[1] N. K. Ratha, J. H. Connell, and R. M. Bolle, “Enhancing security and privacy in biometrics-based authentication systems,” *IBM systems Journal*, vol. 40, no. 3, pp. 614–634, 2001.

[2] N. Ratha, J. Connell, R. M. Bolle, and S. Chikkerur, “Cancelable biometrics: A case study in fingerprints,” in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 4. IEEE, 2006, pp. 370–373.

[3] A. T. B. Jin, D. N. C. Ling, and A. Goh, “Biohashing: two factor authentication featuring fingerprint data and tokenised random number,” *Pattern recognition*, vol. 37, no. 11, pp. 2245–2255, 2004.

[4] P. Lacharme, E. Cherrier, and C. Rosenberger, “Preimage attack on biohashing,” in *International Conference on Security and Cryptography (SECRYPT)*, 2013.

[5] A. Teoh and C. T. Yuang, “Cancelable biometrics realization with multispace random projections,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 5, pp. 1096–1106, 2007.

[6] A. Lumini and L. Nanni, “An improved biohashing for human authentication,” *Pattern recognition*, vol. 40, no. 3, pp. 1057–1065, 2007.

[7] S. William and W. Stallings, *Cryptography and Network Security, 4/E*. Pearson Education India, 2006.

[8] E. Bingham and H. Mannila, “Random projection in dimensionality reduction: applications to image and text data,” in *Proc. of 7th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2001, pp. 245–250.

[9] J. Matoušek, “On variants of the johnson–lindenstrauss lemma,” *Random Structures & Algorithms*, vol. 33, no. 2, pp. 142–156, 2008.

[10] S. Dasgupta, “Learning mixtures of gaussians,” in *Foundations of Computer Science, 1999. 40th Annual Symposium on*. IEEE, 1999, pp. 634–644.

[11] ORL face database, *AT&T Laboratories Cambridge*, http://www.cl.cam.ac.uk/.

[12] Yale face database, *Center for computational Vision and Control at Yale University*, http://cvc. yale. edu/projects/yalefaces/yalefa/.

[13] The Indian face database, *IIT Kanpur*, http://vis-www.cs.umas.edu/.

[14] CASIA palmprint database, *Biometrics Ideal Test*, http://biometrics.idealtest.org/downloadDB/.

[15] PolyU palmprint database, *The Hong Kong Polytechnic University*, http://www4.comp.polyu.edu.hk/biometrics/.

[16] C. Lobrano, R. Tronci, G. Giacinto, and F. Roli, “A score decidability index for dynamic score combination,” in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 69–72.

[17] P. Indyk and R. Motwani, “Approximate nearest neighbors: towards removing the curse of dimensionality,” in *Proc. of 30th annual ACM symposium on Theory of computing*. ACM, 1998, pp. 604–613.