

Biometric Template Transformation: A Security Analysis

Abhishek Nagar^a Karthik Nandakumar^b and Anil K. Jain^{a,c}

^aMichigan State University, East Lansing, MI 48824, USA;

^bInstitute for Infocomm Research, A*STAR, Fusionopolis, Singapore;

^cDept. of Brain and Cognitive Eng., Korea University, Seoul 136-713, Korea

ABSTRACT

One of the critical steps in designing a secure biometric system is protecting the templates of the users that are stored either in a central database or on smart cards. If a biometric template is compromised, it leads to serious security and privacy threats because unlike passwords, it is not possible for a legitimate user to revoke his biometric identifiers and switch to another set of uncompromised identifiers. One methodology for biometric template protection is the template transformation approach, where the template, consisting of the features extracted from the biometric trait, is transformed using parameters derived from a user specific password or key. Only the transformed template is stored and matching is performed directly in the transformed domain. In this paper, we formally investigate the security strength of template transformation techniques and define six metrics that facilitate a holistic security evaluation. Furthermore, we analyze the security of two well-known template transformation techniques, namely, Biohashing and cancelable fingerprint templates based on the proposed metrics. Our analysis indicates that both these schemes are vulnerable to intrusion and linkage attacks because it is relatively easy to obtain either a close approximation of the original template (Biohashing) or a pre-image of the transformed template (cancelable fingerprints). We argue that the security strength of template transformation techniques must also consider the computational complexity of obtaining a complete pre-image of the transformed template in addition to the complexity of recovering the original biometric template.

Keywords: Biometrics, template security, template transformation, biohashing, cancelable templates

1. INTRODUCTION

Biometric recognition has a number of advantages over the traditional authentication mechanisms based on tokens (e.g., ID cards) or passwords. This is because of the inalienable and distinctive nature of the biometric traits. However, biometric systems are not fool-proof and a critical vulnerability that is unique to biometric systems is the compromise of the stored templates*. Stolen templates can be used by an adversary to create biometric spoofs^{1,2}(see Figure 1), which in turn can be used to gain illegitimate access to systems that employ the same biometric trait of the user. Even when spoof creation is difficult, a stolen template can be replayed to the biometric system in order to circumvent it (*intrusion* attack). Since biometric traits are supposed to be permanent and unique to an individual, stolen templates can also be used to link a user across databases (*linkage* attack) or glean additional information about the user such as race, gender and certain medical conditions.³ Unlike passwords, it is not possible for a legitimate user to revoke his biometric template and switch to another uncompromised template. Hence, ensuring the security of biometric templates is essential for gaining public trust and acceptance, which in turn will promote the widespread deployment of biometric systems.

A number of techniques have been designed to improve the security of biometric templates. The hardware-based approach involves designing a “closed” recognition system, where the biometric template never leaves a physically secure module such as a smart card or a hand-held device.⁴ The card or device may contain only the template and the matcher (Match-on-card) or the complete biometric system including sensor, feature extractor, template, and matcher (System-on-card). Such a device matches the input biometric trait with the template stored in the device and releases a key in case the authentication is successful.

Further author information: (Send correspondence to A. Nagar, nagarabh@cse.msu.edu, 1 517 285 3592)

*A template is a set of features extracted from the biometric trait. A template is stored in the biometric system database and is used for matching with the input biometric during an authentication attempt.

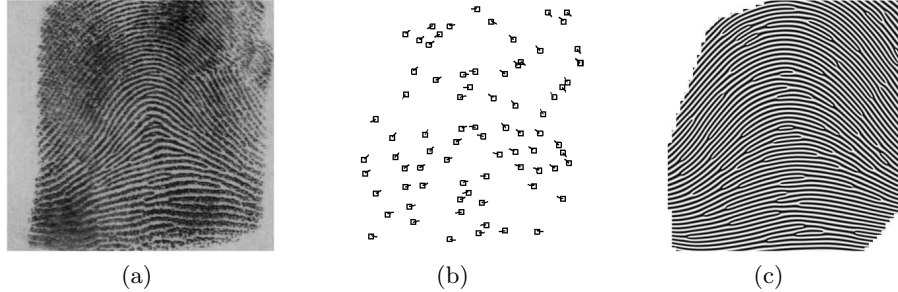


Figure 1. Reconstruction of a fingerprint image from its template (consisting of location and orientation of minutiae - points on a fingerprint where the ridges end or bifurcate). (a) Original fingerprint image, (b) Template consisting of minutiae extracted from the fingerprint image in (a), and (c) Fingerprint image reconstructed from the template in (b) using the technique proposed by Feng and Jain².

Software-based solutions for template protection store a modified version of the template that reveals as little information about the original biometric trait as possible and yet can be successfully used for authentication (see Figure 2). The proposed solutions can be classified into two main categories:⁵ (i) *Template or Feature Transformation*, and (ii) *Biometric Cryptosystem*. Template transformation techniques transform the biometric template based on parameters derived from external information such as user passwords or keys. During authentication, the same transformation function is applied to the query and matched with the stored template in the transformed domain. Biometric cryptosystems attempt to obtain error correcting information from the biometric features (with or without use of external key) that is known as *helper data*. The helper or auxiliary data does not reveal significant information about the biometric or the key. Error correcting codes are normally used in these systems to recover the enrolled biometric features or the key given the query biometric. These two approaches can also be combined to consolidate their advantages. One way to combine them is to use a transformed template as the input to a biometric cryptosystem (e.g., Nandakumar et al.⁶). Figure 2(c) shows the schematic diagram of such a hybrid biometric cryptosystem. Another possibility is to transform the helper data in a biometric cryptosystem using an homomorphic encryption scheme (e.g., Bringer and Chabanne⁷).

Both template transformation and biometric cryptosystems have their own advantages and limitations. Templates generated using the transformation approach are easily revocable (by changing the password or key). Since there are fewer restrictions on the matching algorithms that can be used in the transformed domain, it is possible to design sophisticated matchers that can robustly handle intra-user variations in the transformed biometric templates, thereby reducing the error rates of the biometric system. However, it is difficult to measure the security strength of template transformation techniques. On the other hand, biometric cryptosystems mostly rely on error correction coding theory. While this allows us to easily understand and evaluate the security strength in information-theoretic terms,⁸⁻¹¹ it restricts the use of any sophisticated matchers. Consequently, the matching performance of a biometric cryptosystem is limited by the error-correction capability of the code used and the only way to improve the performance is to extract invariant and discriminative features from the biometric trait with specific representation formats (e.g., fixed-length binary strings). Moreover, some biometric cryptosystems may be vulnerable to correlation attacks,¹² where multiple auxiliary data generated from the same biometric trait can be matched to extract the original biometric template, hence affecting the revocation capability.

Though template transformation techniques have some advantages compared to biometric cryptosystems, their practical applicability is hindered by the lack of formal security analysis. In this paper, we propose a set of measures that facilitate a holistic security evaluation of template transformation techniques. First, we review various template transformation techniques in Section 2, which is followed by the security analysis in Section 3. Sections 4 and 5 provide the security analysis of the well-known cancelable fingerprint template and Biohashing techniques based on the proposed metrics. Finally, we summarize our findings and conclusions in Section 6.

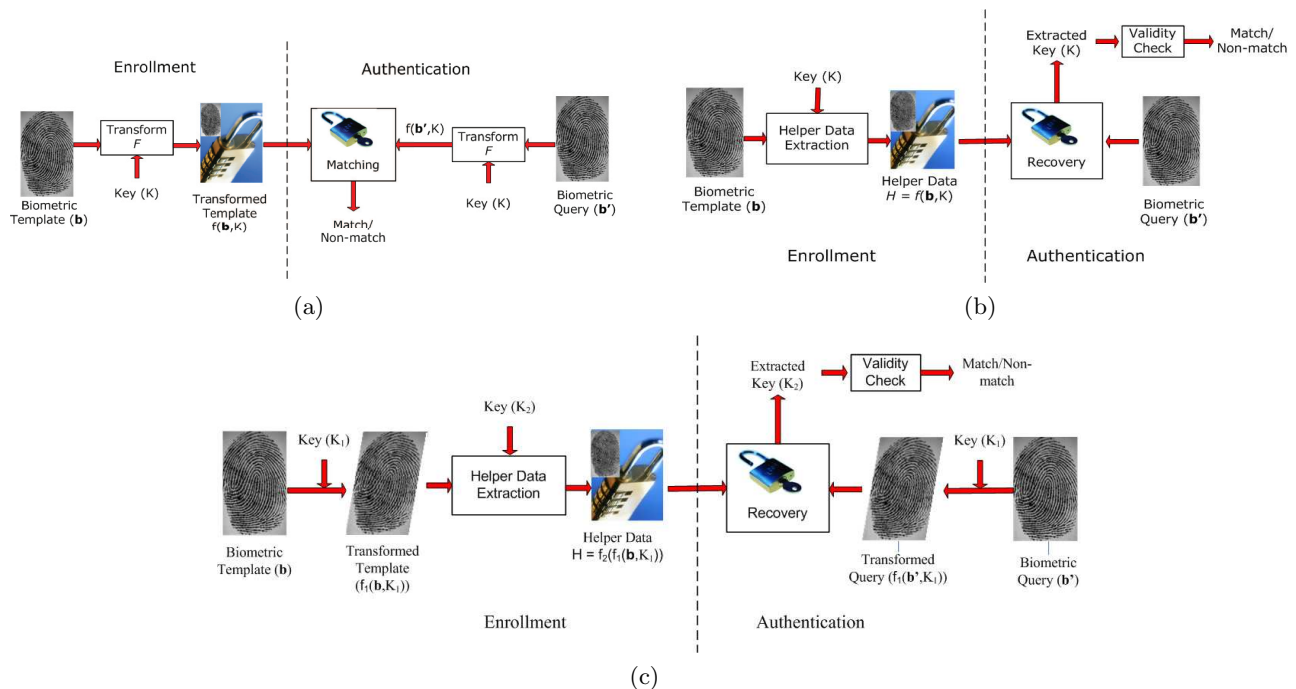


Figure 2. Schematic diagrams for (a) template transformation approach, (b) biometric cryptosystem, and (c) hybrid biometric cryptosystem. While these three template protection approaches are applicable to any biometric trait, fingerprints have been used here for illustration purpose.

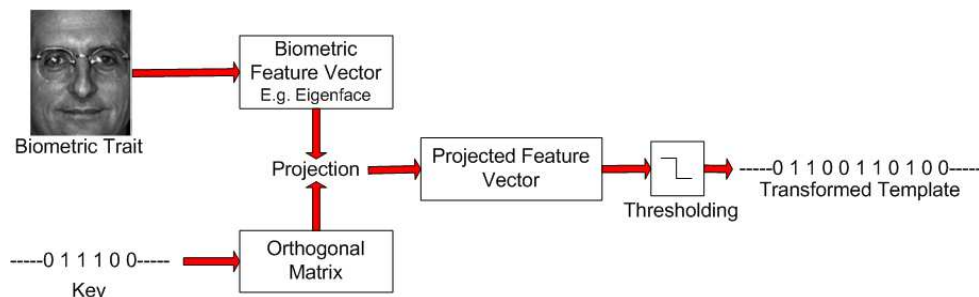


Figure 3. Schematic diagram of the Biohashing technique.

2. REVIEW OF BIOMETRIC TEMPLATE TRANSFORMATION

A number of template transformation techniques have been proposed (see Table 1), which can be classified into two main categories based on template representation used: (i) Vector based and (ii) Interest point based.

2.1 Vector based template transformation

In the vector based techniques, the biometric templates are represented as a vector and the dissimilarity between two vectors is usually computed as the Euclidean distance. One of the main requirements of a vector based template transformation function is the preservation of distances between the vectors after transformation. Biohashing¹³ is one such technique (see Figure 3), where the feature vector is transformed by multiplying it with an orthogonal transformation matrix and thresholding the individual elements. Due to increased inter-class variation and preservation of intra-class variation Biohashing significantly improves the matching performance. However, if the key is known to the adversary, the matching performance typically degrades due to the quantization of features and dimensionality reduction.

Table 1. List of different template transformation techniques available in literature and their characteristics.

Technique	Trait	Features	Transformation	Final representation
Biohashing, ^{13,14} PalmHash ¹⁵	Face, Palmprint, Fingerprint	Vector (Fisher Discriminant Features)	Random matrix multiplication	Vector
BioPhasor ¹⁶	Fingerprint	Vector (FingerCode)	Non-linear	Vector
Cancelable Face ¹⁷	Face	Vector (Face image)	Random matrix convolution	Vector
Robust Hash ¹⁸	Face	Vector (Singular Values of face image matrix)	Smooth multimodal function evaluation	Vector
Class Distribution Preserving Transformation ¹⁹	Face	Vector (Fisherface features)	Evaluation of distance of the feature vector from a set of points	Vector
Cancelable Iris ²⁰	Iris	Vector (Log-Gabor response)	Circular shift and combination, adding new pattern	Vector
Histogram of minutiae triangles ²¹	Fingerprint	Interest point	Hashing the histogram of minutiae triangle features	Vector
Symmetric Hash ²²	Fingerprint	Interest point (Minutiae as complex numbers)	Set of order invariant functions of minutiae	Minutiae map
Cancelable Fingerprints ²³	Fingerprint	Interest point (Minutiae map)	Image folding	Minutiae map
Alignment free cancelable fingerprint ²⁴	Fingerprint	Interest point (minutiae map, orientation field)	Transform minutiae according to surrounding orientation field	Minutiae map
Cuboid based Minutiae Aggregates ²⁵	Fingerprint	Interest point (Minutiae map)	Minutiae aggregate feature selection from random local regions	Vector

Another drawback of the Biohashing scheme is that it is easy to invert when the key is known to the adversary (see section 5). Inversion is the process of recovering the original biometric template from the transformed template and invertibility can be expressed in terms of the computational complexity and the number of guesses involved in recovering the original template. In some cases such as Biohashing, it is straightforward to directly recover the original biometric template (or a close approximation of it) when the key is known. However, in other cases like robust hashing¹⁸ and cancelable fingerprint templates,²³ it is either computationally hard to obtain the complete pre-image[†] of the transformed template or computationally difficult to identify the original biometric template from the pre-image due to the large size of the pre-image. Such schemes are considered to be difficult to invert (also sometimes loosely referred to as “non-invertible”).

An improvement of the Biohashing scheme is the BioPhasor¹⁶ technique, where the rows of the orthogonal transformation matrix are used as the imaginary part and added to the biometric vector to obtain a set of complex vectors. For each of these vectors, the argument of the values in them are averaged and quantized to form the final binary template. This transformation has been shown to better preserve the matching performance even if the password is known to the adversary. Although this scheme is claimed to be non-invertible, the complexity involved in inverting this transformation is not known. Savvides et al.¹⁷ showed that the distance between two Minimum Average Correlation Energy (MACE) filter outputs is preserved even when the face image is convolved using a random kernel matrix for template protection. However, this scheme is invertible given the knowledge of the convolution kernel and the specific MACE filters used. Sutcu et al.¹⁸ proposed a transformation technique, where each element of the input biometric vector is evaluated on a multi-modal polynomial. Due to the many-to-one nature of the transformation function induced by the multi-modality of the polynomials, it is difficult to invert the transformed template. Feng and Yuen²⁶ transformed the template by randomly selecting a set of vectors of the same dimension as the biometric feature vector and then storing the Euclidean distances of the biometric

[†]A pre-image of a transformed template is the collection of all the templates in the original domain that can generate the given transformed template.

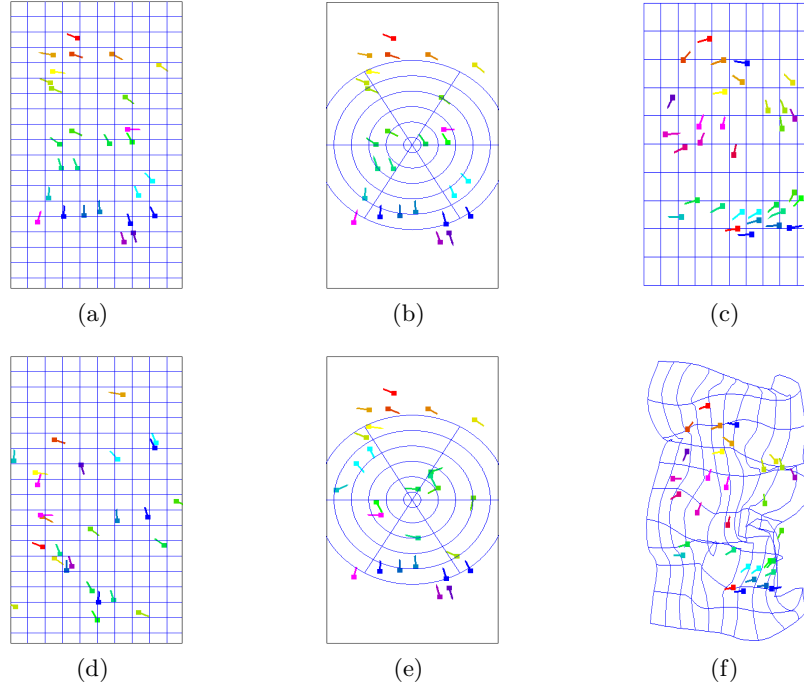


Figure 4. The original and transformed fingerprints for (a,d) Cartesian, (b,e) polar, and (c,f) Gaussian mixtures based transform .

vector from these vectors. This technique also uses the feature distribution of an individual user while designing the transform, which possibly leaks some additional information regarding the biometric vector. The complexity of inverting the template i.e. recovering the original biometric from the transformed template is expected to be greater than that of biohashing technique. Zuo et al.²⁰ proposed two template transformation schemes for iris images. In the first scheme called COMBO, the original iris template was tessellated into rectangles, rows were cyclically shifted and different rows were added to obtain the transformed template. In the second scheme called SALTING, the iris image or its binary representation was added to a randomly generated texture to obtain the transformed template. The COMBO approach is shown to be difficult to invert because of the addition of two different biometric features, which provides ambiguity about the component features.

2.2 Interest point based template transformation

Fingerprints are most commonly represented by a set of points, called minutiae. Hence, many fingerprint template transformation techniques are based on minutiae as the initial representation. Furthermore, to use the available minutiae-based fingerprint matchers in the transformed domain, it is desirable to have the final representation also in the form of a set of minutiae. To satisfy this criterion, Ratha et al.²³ proposed the use of cancelable fingerprint templates designed using three different minutiae transformation techniques, namely, cartesian, polar and functional (see Figure 4). In the cartesian transformation, the fingerprint is regularly tessellated into a set of rectangles and these rectangles are displaced according to the associated key. The polar transformation is similar to the cartesian transformation except that the fingerprint is divided into a number of shells and each shell is divided into sectors. Since the size of sectors is different for different shells, some restrictions are placed on the displacement of the sectors based on the password. In case of the functional transformation, two different functions are used: a mixture of 2D Gaussians and electric potential field in 2D charge distribution. These functions are evaluated at the minutiae locations to obtain the translation corresponding to that minutia.

All the three transformations proposed by Ratha et al.²³ are difficult to invert. This is due to the many-to-one nature of the transformation functions. However, these techniques lead to a reduction in the matching

performance due to an increase in the intra-user variations[‡]. Such transforms also require the fingerprints to be aligned before applying the transformation; misalignment can further increase intra-user variations. To avoid alignment, Lee et al.²⁴ proposed an alignment-free cancelable fingerprint transform. In this scheme, each minutiae is transformed according to the orientation field around that minutiae, which makes the relative translation of the minutiae invariant to the positioning of the finger. Tulyakov et al.²² use each minutia along with its two nearest neighbors to select one of the several symmetric functions available. The selected symmetric function is then evaluated on the three minutiae to obtain the coordinates of the transformed minutia.

Techniques have also been proposed to convert the minutiae based representation into a vector based representation. Farooq et al.²¹ select all minutiae triplets satisfying certain criteria and construct a histogram. Only those bins in the histogram with a single element are retained and the remaining bins are emptied to obtain the final binary feature vector. Cancelability is induced by flipping some of the bits and permuting the binary vector based on a specific key. The limitation of this approach is that it is easy to determine the unique triangles present in the fingerprint and their dimensions can be refined due to redundancy in the representation of each triangle. Furthermore, sides with similar length can be matched and combined to construct an approximate minutiae distribution. The complexity of such a procedure however might be high. Another scheme proposed by Sutcu et al.,²⁵ converts a set of minutiae into a vector based representation by counting the number of minutiae falling in certain specified rectangular regions. The configurations of rectangular regions can be changed in order to generate another template from the same biometric thereby inducing cancelability.

3. SECURITY ANALYSIS OF TEMPLATE TRANSFORMATION

We focus on the vulnerability of a template transformation scheme to intrusion and linkage attacks that can be staged using the knowledge of a stored template. Intrusion means gaining access to a biometric recognition system by presenting falsified authentication data to the system. Intrusion undermines one of the fundamental benefits of using a biometric system, which is non-repudiation. On the other hand, linkage attacks involve cross-matching across biometric systems to track the users covertly and this compromises the privacy of the user. Hence, it is important to analyze the probability of success of these two attacks in a biometric system.

We employ the following notation to describe the security metrics. Let \mathbf{b}_z and \mathbf{b}'_z represent the template and query biometric features of user z , respectively. Let f be the feature transformation function and f^{-1} be its inverse. Let f_β^{-1} denote the partial inverse transformation function, where β is the fraction of the original biometric template obtained by inverting the transformed template. Let K_z be a set of transformation parameters corresponding to user z and K'_z be a different set of transformation parameters for the same user. Let \mathcal{D}_O denote a distance function between the biometric features in the untransformed (original) domain and \mathcal{D}_T be a distance function between the biometric features in the transformed domain. The biometric system outputs a “match” decision if the distance between the template and query biometric features is less than a threshold ϵ .

3.1 System Usability

Security of a biometric recognition system affects the usability of the system as well. While considering the system security, it is important to measure any inconvenience incurred to the genuine users of the system as a result of the security techniques implemented. We measure the usability in terms of the false reject rate of the system. The false reject rate of the biometric system prior to the template transformation, FRR_O , is given by

$$FRR_O(\epsilon) = P\left(\mathcal{D}_O\left(\mathbf{b}_z, \mathbf{b}'_z\right) \geq \epsilon\right). \quad (1)$$

The false reject rate of the biometric system after the application of template transformation, FRR_T , is

$$FRR_T(\epsilon) = P\left(\mathcal{D}_T\left(f\left(\mathbf{b}_z, K_z\right), f\left(\mathbf{b}'_z, K'_z\right)\right) \geq \epsilon\right). \quad (2)$$

[‡]Intra user variation refers to changes in the template of the same user in different acquisitions of the biometric sample. Since the transformation functions are generally non-Euclidean, variations in minutiae position and orientation are escalated due to transformation, leading to high false reject rate.

FRR_O and FRR_T depend on the threshold ϵ and must be as low as possible to avoid inconvenience to the users. The threshold ϵ also controls the security and privacy of the system because the probability of success of an intrusion or linkage attack depends on ϵ .

3.2 Security Evaluation Measures for Intrusion Threats

First, we consider the case of an impostor presenting his/her own biometric trait in order to get authenticated. In this scenario, the adversary does not expend any effort to guess the biometric features of the user that he/she is trying to impersonate, so the probability of a successful intrusion mainly depends on the inter-user variability of the biometric features. This kind of attack is known as a zero-effort attack and the intrusion success probabilities are known as false accept rates. The false accept rate of the biometric system prior to the template transformation (FAR_O) is given by

$$FAR_O(\epsilon) = P\left(\mathcal{D}_O\left(\mathbf{b}_i, \mathbf{b}'_j\right) < \epsilon\right), \text{ where } i \neq j. \quad (3)$$

A plot of FAR_O versus $(1 - FRR_O)$ for various values of ϵ gives the receiver operating characteristic (ROC_{orig}) curve of the biometric system prior to template transformation.

After template transformation, the impostor has to present the biometric features along with a set of transformation parameters for authentication. Hence, there are two possibilities. Suppose that the impostor does not know the transformation parameters of the specific user that he is trying to impersonate. The false accept rate with unknown transformation parameters (K) is given by

$$FAR_{UK}(\epsilon) = P\left(\mathcal{D}_T\left(f\left(\mathbf{b}_i, K_i\right), f\left(\mathbf{b}'_j, K_j\right)\right) < \epsilon\right), \text{ where } i \neq j. \quad (4)$$

and a plot of FAR_{UK} versus $(1 - FRR_T)$ gives the receiver operating characteristic (ROC_{diff}) curve of the biometric system after template transformation for unknown transformation parameters.

If the impostor somehow knows the transformation parameters of the genuine user that he/she is trying to impersonate, the false accept rate with known transformation parameters (K) is

$$FAR_{KK}(\epsilon) = P\left(\mathcal{D}_T\left(f\left(\mathbf{b}_i, K_i\right), f\left(\mathbf{b}'_j, K_i\right)\right) < \epsilon\right), \text{ where } i \neq j \quad (5)$$

and a plot of FAR_{KK} versus $(1 - FRR_T)$ gives the receiver operating characteristic (ROC_{same}) curve of the biometric system after template transformation for known transformation parameters. A comparison of ROC_{orig} and ROC_{same} will indicate the degradation in the matching performance due to the template transformation.

Besides the false accept rates, two other intrusion probabilities must be considered. First we consider the case when the stored (transformed) template and the transformation parameters are available to the adversary. The goal of the adversary is to gain illegitimate access to the biometric system. In this case, the adversary will try to recover either a fraction (β) or the complete biometric template and then replay the inverted template along with the transformation parameters to gain access fraudulently. The probability of success of such an attack is called the Intrusion Rate due to Inversion for the Same biometric system ($IRIS$) and is defined as

$$IRIS(\beta, \epsilon) = P\left(\mathcal{D}_T\left(f\left(f_{\beta}^{-1}\left(f\left(\mathbf{b}_i, K_i\right), K_i\right), K_i\right), f\left(\mathbf{b}_i, K_i\right)\right) < \epsilon\right). \quad (6)$$

The value of $IRIS(\beta, \epsilon)$ is usually 1 if a transformation is easy to invert or an element in the pre-image of the transformed template can be obtained (as in the case of many-to-one transformations). $IRIS(\beta, \epsilon)$ will be low when it is difficult to obtain the complete pre-image of the transformed template.

Next, we consider the case when the stored (transformed) template and the transformation parameters are available to the adversary who wants to impersonate the same user in a different biometric system that employs the same biometric trait. We also assume that the adversary has knowledge of the transformation parameters of the second system. In this case, the adversary will try to recover either a fraction (β) or the complete biometric

template and then replay the inverted template along with the transformation parameters of the second system to gain access fraudulently. The probability of success of such an attack is referred to as the Intrusion Rate due to Inversion for a Different biometric system (*IRID*) and is defined as

$$IRID(\beta, \epsilon) = P\left(\mathcal{D}_T\left(f\left(f_\beta^{-1}(f(\mathbf{b}_i, K_i), K_i), K'_i\right), f(\mathbf{b}'_i, K'_i)\right) < \epsilon\right). \quad (7)$$

Finally, we also need to consider the effort spent by the adversary to invert a transformed template. Let $E(\beta)$ denote the effort required in terms of the number of guesses required (expressed in bits) to recover a fraction β of the original biometric template from the transformed template. The plot of β versus $E(\beta)$ is called the coverage-effort curve (C-E curve).²⁷ The coverage-effort curve is a quantitative measure to evaluate the invertibility of a biometric template, provided it is possible for the adversary to check whether the recovered template is a true template. The C-E curve relates the probability of success of intrusion attacks due to inversion (*IRIS* and *IRID*) and difficulty in inverting a transformed biometric template.

3.3 Security Evaluation Measures for Linkage Threats

In order to link two different templates generated from the same biometric trait of a user with different sets of transformation parameters, the adversary may either directly match the transformed templates or he can first invert the templates and then match the inverted templates. Suppose that both sets of transformation parameters, which were used to generate the two templates, are known to the adversary. The cross match rates can be defined in the transformed (CMR_T) and original (CMR_O) feature domains as follows.

$$CMR_T(\epsilon) = P\left(\mathcal{D}_T\left(f(\mathbf{b}_i, K_i), f(\mathbf{b}'_i, K'_i)\right) < \epsilon\right), \text{ and} \quad (8)$$

$$CMR_O(\beta, \epsilon) = P\left(\mathcal{D}_O\left(f_\beta^{-1}(f(\mathbf{b}_i, K_i), K_i), f_\beta^{-1}(f(\mathbf{b}'_i, K'_i), K'_i)\right) < \epsilon\right). \quad (9)$$

In case of linkage attack in the untransformed domain, the failure rate or the False Cross Match Rate of the attacker is given by

$$FCMR_O(\beta, \epsilon) = P\left(\mathcal{D}_O\left(f_\beta^{-1}(f(\mathbf{b}_i, K_i), K_i), f_\beta^{-1}(f(\mathbf{b}'_j, K'_j), K'_j)\right) < \epsilon\right), \text{ where } i \neq j. \quad (10)$$

A plot of $CMR_O(\beta, \epsilon)$ versus $FCMR_O(\beta, \epsilon)$ provides the receiver operating characteristic (ROC_{inv}) curve for the linkage attack in the original domain.

The complexity of cross-matching BioPhasors is difficult to estimate, however, inversion of Biohashing, and cancelable face is computationally easy and is expected to generate a close approximation to the original template. In order to link templates secured using cancelable fingerprint templates approach, one can overlay all the pre-images of minutiae in the transformed template and then match.^{28,29} Note that in this case the matcher should not penalize the non-matching minutiae. Similar techniques can also be used to link templates encrypted using the robust hashing approach. In case of histogram of minutiae triplets, it is easy to obtain the original histogram, which can be easily matched. Symmetric hashing, cancelable iris, CDP, and cuboid based minutiae aggregates are not straight forward to invert and link.

Comprehensive security evaluation of a template transformation scheme entails analysis of the intrusion and linkage probabilities and their effect on the system usability measured in terms of FRR_T . In order to measure the probability of system intrusion, we have defined FAR_{UK} , FAR_{KK} , *IRIS*, and *IRID*. FAR_{UK} and FAR_{KK} analyze the attacks staged by an adversary by presenting an arbitrary biometric template whereas *IRIS* and *IRID* analyze the attacks when the attacker steals a transformed template, inverts it and then uses it for intruding the system. Linkage probabilities can be measured in terms of CMR_O , where the templates are linked in the original domain (after inversion), and CMR_T , where the templates are linked in the transformed domain.

4. SECURITY OF CANCELABLE FINGERPRINT TEMPLATES

We choose cancelable fingerprint templates as an example for security evaluation because though the scheme is difficult to invert, a pre-image computation technique is available in the literature.²⁷ We evaluate the security strength of the mixture of Gaussians based transformation function, which is claimed to have the best performance among all the transforms evaluated by Ratha et al.²³ The mixture of Gaussians used to obtain the transformation function is given by

$$f(\vec{x}) = \sum_{i=1}^N t_i \pi_i e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)\Sigma_i^{-1}(\vec{x}-\vec{\mu}_i)'}, \quad (11)$$

where N is the number of mixture components, and π_i, t_i, μ_i , and Σ_i correspond to the mixing probabilities, the signs (+ or -), mean vectors, and covariance matrices of the different components, respectively. Here, \vec{x} is a vector representation of a minutia point consisting of only the x and y coordinates of the minutiae. In our experiments, N is set to 24 and Σ_i is taken to be a diagonal matrix with each diagonal entry equal to 50^2 for each component. The remaining parameters are determined using the user specific key. These parameters are the same as those used by Ratha et al.²³

The transformation of each minutia is represented as direction of minutia translation (denoted by ϕ_ψ), magnitude of minutia translation (denoted by ϕ_d) and difference in minutia direction (denoted by ϕ_θ). The three components of the transformation can be obtained as:

$$\phi_d(\vec{x}) = \gamma \{1 + f(\vec{x})\}, \quad \phi_\psi(\vec{x}) = \arctan\left(\frac{g'_y(\vec{x})}{g'_x(\vec{x})}\right) + \alpha_\psi, \quad \phi_\theta(\vec{x}) = \arctan\left(\frac{f'_y(\vec{x})}{f'_x(\vec{x})}\right) + \alpha_\theta, \quad (12)$$

where $f'_y(\cdot), f'_x(\cdot), g'_y(\cdot), g'_x(\cdot)$ are the x and y derivatives of two mixture of Gaussians f and g , and $\alpha_\psi, \alpha_\theta \in [0, 360)$ is a random offset in direction; γ is used to manipulate the overall translation of minutiae.

We evaluate the performance of the above template transformation technique using the publicly available FVC 2002 database 2. It consists of 100 different fingers and 8 impressions per finger, each captured at a resolution of 569 dpi. The size of images is 296×560 . We evaluate two different instances of the mixture of Gaussians based transformation with the values of γ being 30 (Trans-1) and 60 (Trans-2) respectively. Their respective transformation functions are shown in Figure 5. Figure 6 shows the evaluation measures described in Section 3 corresponding to these two instances of the mixture of Gaussians.

Figure 6(a) shows that the matching performance after transformation is significantly degraded compared to the original minutiae and the amount of degradation increases with γ . Also, the matching performance is lower when the attacker knows the key as shown by the ROC_{same} plots. As seen from Figures 6(b) and 6(c), the reduction in performance is mainly due to an increase in the FRR, which is primarily due to misalignment. Note

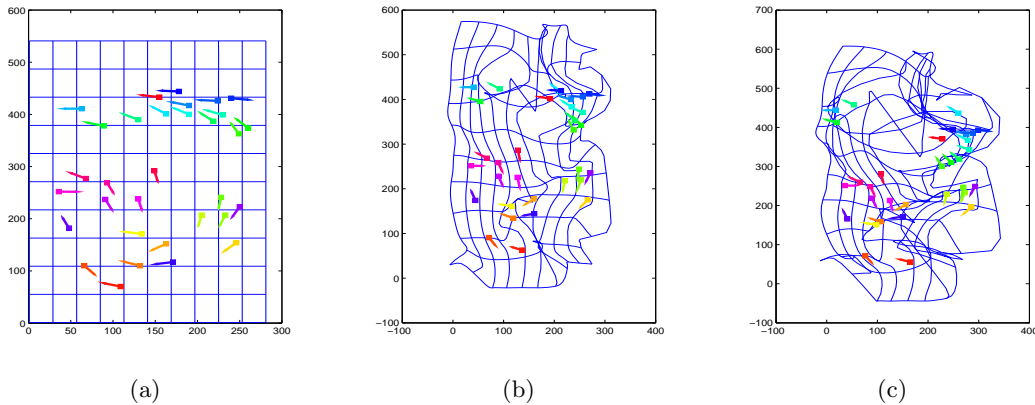


Figure 5. Minutiae transformation (a) minutiae distribution in the original image, (b) minutiae transformed according to mixture of Gaussians, where γ is 30, and (c) transformed minutiae when the value of γ is 60.

that the minutiae must be pre-aligned before applying the transformation function. We use the high curvature points³⁰ in the fingerprint for pre-alignment, which cannot be extracted reliably in partial fingerprints.

Figure 6(d) depicts the feasibility of intruding a different biometric system that employs the same fingerprint using the template inverted from the current system. At an operating threshold (ϵ) of 950, IRID for Trans-2 is around 51% when the attacker expends zero effort in inverting the template, i.e., $E(\beta) = 0$ or $\beta = E^{-1}(0)$. In other words, when the attacker just replays the most likely minutiae set from the pre-image of the transformed template without spending any effort on identifying the original minutiae from the pre-image, there is a 51% chance that he will succeed in intruding the system. A completely inverted template will further increase the intrusion rate to 54%. This value is even higher (64% for zero effort and 65% for full inversion) in the case of Trans-1. Note that the chances of intrusion increased only by 1% for Trans-1, while it increased by around 3% in case of Trans-2. This can be explained by the C-E curve shown in Figure 6(e); attacker can recover only 87% of minutiae without any effort in the case of Trans-2, whereas in the case of Trans-1 he can recover around 94%. Note that the value of $IRID(1, \epsilon)$ is upper bounded by $(1 - FRR_T(\epsilon))$, which corresponds to case where the

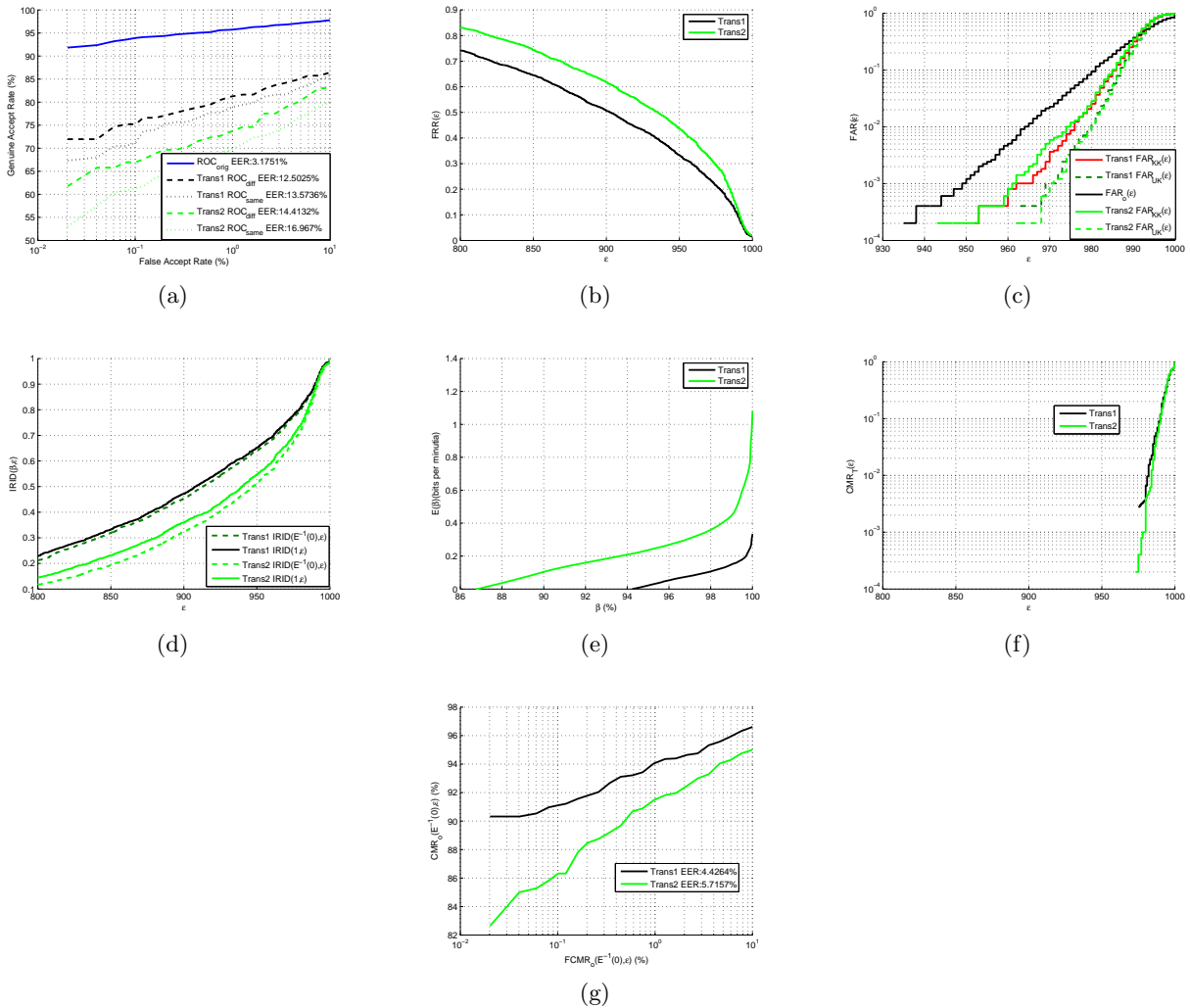


Figure 6. Evaluation measures for the mixture of Gaussian template transformation. (a) ROC_{orig} , ROC_{diff} , ROC_{same} , (b) $FRR_T(\epsilon)$ (c) $FAR_{UK}(\epsilon)$ and $FAR_{KK}(\epsilon)$, (d) $IRID(\beta, \epsilon)$ for two different values of β , (e) the C-E curve, (f) $CMR_T(\epsilon, \beta)$ for $\beta = 1$, and (g) ROC_{inv} . Neurotechnology Verifinger 4.2 is used to perform minutiae matching. These evaluations correspond to two transformations, Trans-1 and Trans-2, where γ equals 30 and 60, respectively.

Transform	FRR_T (%)	FAR_{UK} (%)	FAR_{KK} (%)	$IRID(E^{-1}(0), \epsilon)$	CMR_T (%)
Trans-1	33	0.02	0.02	64	0
Trans-2	44	0.02	0.02	51	0

Table 2. Values of FRR, FAR_{UK} , FAR_{kk} , $IRID(E^{-1}(0), \epsilon)$, and CMR_T for the cancelable fingerprint template scheme corresponding to a threshold (ϵ) of 950.

attacker is able to exactly recover the original fingerprint.

Figure 6(f) shows the feasibility of successfully cross-matching two templates obtained from the same biometric trait but transformed using different transformation parameters. While both Trans-1 and Trans-2 have a zero cross match rate at $\epsilon = 950$, Trans-1 usually has slightly higher CMR_T than Trans-2. Figure 6(g) shows the ROC_{inv} corresponding for $\beta = E^{-1}(0)$. It shows that at a False Cross-match Rate of 0.1%, the chance of correctly linking the templates from two different systems is 91.5% for Trans-2 and 94% for Trans-1.

Table 2 tabulates the values of five security metrics (FRR_T , FAR_{UK} , FAR_{KK} , $IRID$, and CMR_T) for Trans-1 and Trans-2 at a threshold of $\epsilon = 950$. It is quite clear that while Trans-2 is more secure than Trans-1, it is less usable than Trans-1 because of the higher false reject rate, which demonstrates the trade-off between security and usability that is a commonly encountered problem in biometric template protection. Moreover, our analysis also shows that in order to prevent intrusion into other biometric systems that use the same trait and to mitigate linkage threats, it is not enough to design the transformation function such that it is computationally hard to recover the original template, but it must also be computationally difficult to obtain the pre-image of a transformed template. This issue is usually not given adequate attention in the literature.²³

5. SECURITY OF BIOHASHING SCHEME

Biohashing is a vector based template protection technique that is used to secure different biometric traits such as fingerprints,¹⁴ face,¹³ palm,¹⁵ etc. In a typical Biohashing scheme, the input biometric trait is represented as a vector of real numbers, say $\mathbf{x} \in \mathbb{R}^n$. This representation is then converted to a binary vector $\mathbf{b} = [b_1, b_2, \dots, b_m]$ using the transformation matrix $M \in \mathbb{R}^{m \times n}$ and the thresholds $\delta_i, i = 1, \dots, m$. The Biohash features are obtained as:

$$b_i = \begin{cases} 0 & \text{if } \sum_{j=1}^n M_{ij}x_j < \delta_i \\ 1 & \text{otherwise} \end{cases} \quad (13)$$

In our experiments, we use the FERET face database that contains 14,051 images. From these we select a subset of 500 subjects with two frontal images per subject. We align the images using the eye locations and crop a segment of size 100×125 from each image. Eigenface³¹ features are used to represent the face images in our experiments. We use top 100 Eigenface features in order to extract 80 bits using the Biohashing technique.

We now propose a method to recover a close approximation to the original biometric features given the Biohash features (\mathbf{b}) and the transformation parameters, i.e., M and $\delta_i, i = 1, \dots, m$. This problem can be formulated as an optimization problem as follows:

$$\text{argmin } \|\mathbf{x} - \mathbf{a}\|_2, \text{ subject to } \sum_{j=1}^n M_{ij}x_j < \delta_i, \text{ if } b_i = 0 \text{ and } \sum_{j=1}^n M_{ij}x_j > \delta_i \text{ if } b_i = 1, \quad (14)$$

where \mathbf{x} is the original biometric feature vector that is to be estimated, \mathbf{b} is the vector of binary Biohash features and \mathbf{a} is one of the unrelated biometric feature vectors from a database. We use the lsqmin function available in the MATLAB optimization toolbox to obtain a solution to this problem. The above problem is solved for t different values of \mathbf{a} in order to obtain $\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^t$. The final estimate of \mathbf{x} , $\hat{\mathbf{x}}$, is obtained as

$$\hat{\mathbf{x}} = \frac{\sum_{i=1}^t \mathbf{x}^i / d_i^2}{\sum_{i=1}^t 1/d_i^2}, \quad (15)$$

FRR_T (%)	FAR_{UK} (%)	FAR_{kk} (%)	$IRID(E^{-1}(0), \epsilon)$	CMR_T (%)
9	0.02	5	50	0

Table 3. Values of FRR, FAR_{UK} , FAR_{KK} , $IRID(E^{-1}(0), \epsilon)$, and CMR_T for the Biohashing technique corresponding to a threshold (ϵ) of 20.

where d_i is the Hamming distance between Biohash features corresponding to \mathbf{x}^i and \mathbf{a}^i . \mathbf{a}^i 's are chosen such that hamming distance between Biohash features corresponding to a_i and b_i is less than certain threshold. Figure 7(b) shows an example of a face image reconstructed from the Eigenface features ($\hat{\mathbf{x}}$) that are estimated by inverting the Biohash template (\mathbf{b}) using equations (14) and (15). We observe that many distinctive features in the original face image (Figure 7(a)) are also present in the reconstructed image, which demonstrates the effectiveness of our inversion algorithm. Figure 8 shows the evaluation of Biohashing technique with respect to



Figure 7. Inversion of a Biohash template. (a) Original face image from the FERET database (after alignment and cropping), (b) face image reconstructed from the Eigenface features ($\hat{\mathbf{x}}$) that are estimated by inverting the Biohash template (\mathbf{b}) using equations (14) and (15).

the different evaluation criteria proposed in Section 3 except the C-E curve, which is not directly applicable to Biohashing. Figure 8(a) shows the three ROC curves i.e. ROC_{orig} , ROC_{same} , and ROC_{diff} . In contrast to the cancelable fingerprints technique, the ROC_{diff} of Biohashing shows significantly better performance than ROC_{orig} , whereas ROC_{same} has lower matching performance compared to ROC_{orig} . This is because Biohashing uses the external information (key or password) to significantly alter the distribution of the biometric features and increase the inter-user separation. However, this advantage is lost when the key is known to the adversary. On the other hand, the cancelable fingerprints scheme attempts to retain the fingerprint minutiae distribution, so that a traditional minutiae matcher can still be applied to match the transformed minutiae sets. At the operating threshold of 20, the $IRID(E^{-1}(0), \epsilon)$ value is around 0.5, implying that the attacker has 50% success rate in intruding into a different system using the same biometric trait. The cross match rate in the transformed domain (CMR_T) is almost zero at the operating threshold of 20. As expected, the CMR_T follows FAR_{UK} closely. With respect to cross matching in the original domain, the CMR_O is around 82% at 10% $FCMR_O$ as shown in Figure 8(f). Table 3 lists the results values of FRR, FAR_{UK} , FAR_{KK} , $IRID(E^{-1}(0), \epsilon)$, and CMR_T corresponding to the operating threshold of 20.

It is evident from Figure 8(d) that biometric templates from one database can be inverted and used to compromise other systems using the same biometric trait. This is due to the contiguous nature of the pre-image of Biohash. We propose a modification to the original Biohashing scheme, which leads to a non-contiguous pre-image and thus is less vulnerable. The only difference between the modified and the original Biohashing scheme is the binarization procedure. In the original Biohashing technique, binarization is performed by first obtaining the median (δ) of each transformed feature and then thresholding the transformed features using this value. Instead, in the modified technique, each feature is thresholded at three different values: λ^{th} -, 50^{th} -, and $(100 - \lambda)^{th}$ - percentiles leading to four quanta for each feature. While the first and third quanta are represented as a 1, the other two quanta are represented as a 0. Note that $\lambda = 0$ leads to the original Biohashing technique.

Figure 9 shows the ROC_{diff} corresponding to the modified technique for $\lambda \in \{2, 5, 10\}$. While there is certain reduction in the matching performance, it is now difficult to invert the template. The probability of guessing the correct quanta in each dimension is $p_\lambda = \max(\lambda/50, 1 - \lambda/50)$ given that one always chooses the larger quanta.

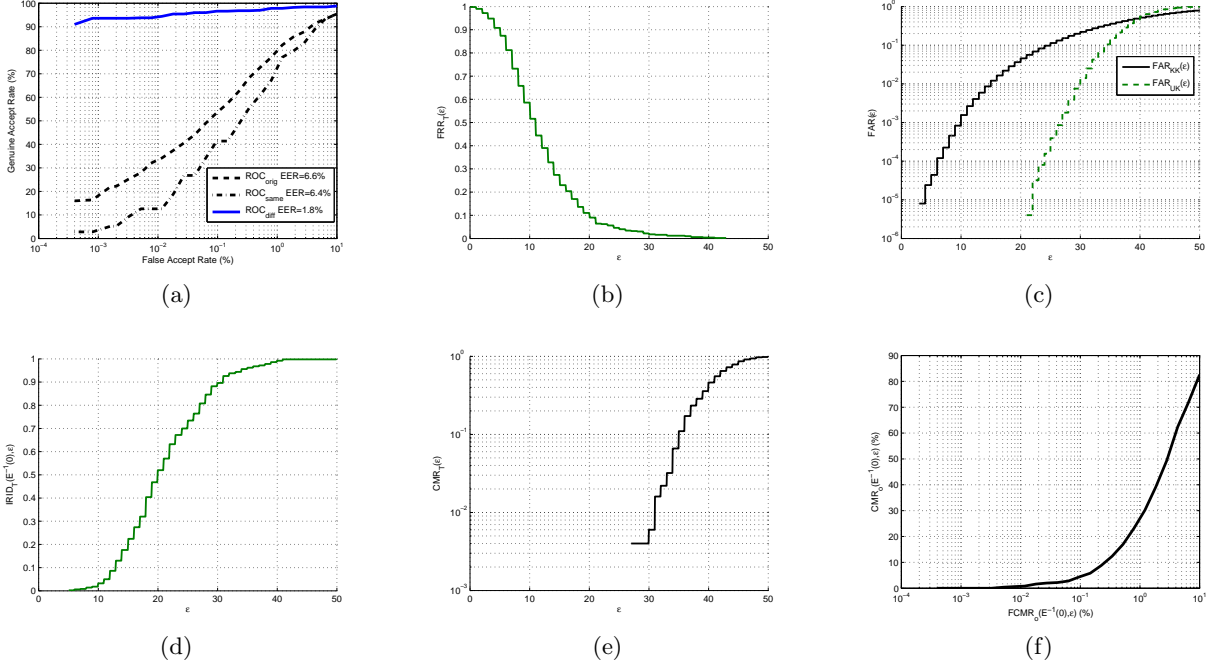


Figure 8. Evaluation of Biohashing technique. (a) ROC_{orig} , ROC_{diff} , ROC_{same} , (b) $FRR_T(\epsilon)$ (c) $FAR_{UK}(\epsilon)$ and $FAT_{KK}(\epsilon)$, (d) $IRID(\beta, \epsilon)$ for two different values of β , (e) $CMR_T(\epsilon, \beta)$ for $\beta = 1$, and (f) ROC_{inv} for the Biohashing technique based on 500 subjects from FERET database. In this experiment, 100 Eigenface features were extracted and 80 bits/template were extracted using Biohashing. The value of t used here is 100.

Thus if there are p Eigenface dimensions to be guessed using m Biohash bits, the probability of identifying the correct quantum in which the non-quantized Biohash values fall is p_λ^m . The security, in terms of bits, for guessing this is $-\log_2(p_\lambda^m)$. In case $m = 80$, the security corresponding to $\lambda = 2, 5$, and 10 are 4.7 bits, 12.1 bits, and 25.8 bits respectively. However, in order to increase the security, m can be increased. In case $m = 400$, the security corresponding to $\lambda = 2, 5$, and 10 is 23.6 bits, 60.8 bits, and 128.8 bits respectively. ROC_{diff} corresponding to the modified Biohashing scheme for different values of λ and $m=80$ and 400 are shown in Figure 9. The matching performance of the Biohashing scheme reduces as λ is increased. However, increasing the number of dimensions improves the security as well as the matching performance in case the impostor does not know the key.

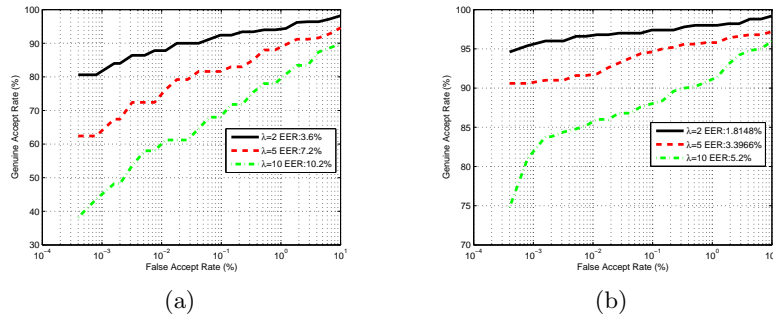


Figure 9. ROC_{diff} corresponding to the modified technique (a) ROC_{diff} for $\lambda \in \{2, 5, 10\}$ corresponding to the case when number of dimensions of PCA retained in 100 and number of bits extracted using Biohashing technique is 80, and (b) shows the ROC_{diff} for $\lambda \in \{2, 5, 10\}$ corresponding to the case when number of dimensions of PCA retained is 500 and the number of bits extracted using Biohashing technique is 400.

6. CONCLUSIONS

When a user's biometric template information falls into the hands of an adversary, it can seriously undermine the security (intrusion threats) of the biometric system and privacy (linkage threats) of the user. Hence, biometric template protection is a critical problem that needs to be addressed to enhance the public acceptance of biometric technology. Considering the recent surge in the number of techniques being developed for protecting the biometric templates, it is essential to develop a set of measures which can evaluate the strength of these techniques. One of the well known approaches for template protection is the template or feature transformation technique. Compared to biometric cryptosystems, template transformation schemes have certain advantages like easy revocability and flexibility in the matcher design. But these advantages are stymied by the lack of a thorough security analysis of these techniques.

We have proposed six different measures to evaluate the security strength of template transformation schemes. Based on these measures, we analyze the security of two-well known transformation techniques, namely, cancelable fingerprints and Biohashing. Our analysis shows that both these techniques are vulnerable to intrusion and linkage attacks, as indicated by their high *IRIS*, *IRID* and *CMR_O* values. In particular, the vulnerability of the Biohashing scheme is due to the relative ease with which an impostor can invert the transformed template to obtain a close approximation to the original biometric template. Hence, we propose a modification to the Biohashing scheme that can address this limitation, though at the expense of a marginal reduction in the matching performance.

In the case of cancelable fingerprint template scheme, the vulnerabilities arise because an impostor can easily obtain the pre-image of the transformed template. Even though it is computationally hard to recover the original template from the pre-image, the pre-image itself is sufficient to carry out linkage and intrusion attacks. Therefore, for enhanced template security, we argue that the non-invertibility of a transformation function must also be measured in terms of the complexity of obtaining the complete pre-image of a transformed template, rather than simply analyzing the complexity of recovering the original template. However, proving the computational hardness of this problem is not easy because it may be possible to design greedy algorithms that can perform the inversion efficiently.

Our experiments also highlight the well-known tradeoff between the security and usability. In this context, hybrid biometric cryptosystems may have an edge because the complementary strengths of template transformation and biometric cryptosystems can be leveraged to improve both the security and usability of a biometric system. As future work, we plan to investigate the effect of various improvements proposed in the Biohashing³² technique on the security analysis. Also we shall investigate the other template transformation techniques such as CDP transform.

REFERENCES

- [1] Cappelli, R., Lumini, A., Maio, D., and Maltoni, D., "Fingerprint Image Reconstruction From Standard Templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(9), 1489–1503 (2007).
- [2] Feng, J. and Jain, A., "FM model based fingerprint reconstruction from minutiae template," in [*International conference on Biometrics (ICB)*], (2009).
- [3] Mordini, E. and Massari, S., "Body, biometrics and identity," *Bioethics* **22**(9), 488–498 (2008).
- [4] Maltoni, D., Maio, D., Jain, A. K., and Prabhakar, S., [*Handbook of Fingerprint Recognition*], Springer-Verlag (2009).
- [5] Jain, A., Nandakumar, K., and Nagar, A., "Biometric template security," *EURASIP Journal on Advances in Signal Processing* **2008**, 1–17 (2008).
- [6] Nandakumar, K., Nagar, A., and Jain, A. K., "Hardening Fingerprint Fuzzy Vault Using Password," in [*Proceedings of Second International Conference on Biometrics*], 927–937 (August 2007).
- [7] Bringer, J. and Chabanne, H., "An authentication protocol with encrypted biometric data," in [*Proceedings of the Progress in Cryptology, AFRICACRYPT*], 109–124 (2008).
- [8] Dodis, Y., Ostrovsky, R., Reyzin, L., and Smith, A., "Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data," Tech. Rep. 235, Cryptology ePrint Archive (February 2006). A preliminary version of this work appeared in EUROCRYPT 2004.

- [9] Lai, L., Ho, S., and Poor, H., "Privacy-security tradeoffs in biometric security systems," in [*Proceedings of the 46th Annual Allerton Conference on Communication, Control, and Computing*], 23–26 (2008).
- [10] Ballard, L., Kamara, S., and Reiter, M., "The practical subtleties of biometric key generation," in [*Proceedings of The 17th Annual USENIX Security Symposium*], 61–74 (2008).
- [11] Golic, J. and Baltatu, M., "Entropy analysis and new constructions of biometric key generation systems," *IEEE Transactions on Information Theory* **54**(5), 2026–2040 (2008).
- [12] Scheirer, W. J. and Boulton, T. E., "Cracking fuzzy vaults and biometric encryption," in [*Proceedings of Biometrics Symposium*], (september 2007).
- [13] Teoh, A. B. J., Goh, A., and Ngo, D. C. L., "Random Multispace Quantization as an Analytic Mechanism for BioHashing of Biometric and Random Identity Inputs," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**, 1892–1901 (December 2006).
- [14] Jin, A. T. B., Ling, D. N. C., and Goh, A., "Biohashing: two factor authentication featuring fingerprint data and tokenised random number," *Pattern Recognition* **37**(11), 2245–2255 (2004).
- [15] Connie, T., Teoh, A. B. J., Goh, M., and Ngo, D. C. L., "PalmHashing: a novel approach for cancelable biometrics," *Information Processing Letters* **93**(1), 1–5 (2005).
- [16] Teoh, A. B. J., Toh, K.-A., and Yip, W. K., " 2^N Discretisation of BioPhasor in Cancellable Biometrics," in [*Proceedings of Second International Conference on Biometrics*], 435–444 (August 2007).
- [17] Savvides, M. and Vijaya Kumar, B. V. K., "Cancellable Biometric Filters for Face Recognition," in [*Proceedings of IEEE International Conference Pattern Recognition*], **3**, 922–925 (August 2004).
- [18] Sutcu, Y., Sencar, H. T., and Memon, N., "A Secure Biometric Authentication Scheme Based on Robust Hashing," in [*Proceedings of ACM Multimedia and Security Workshop*], 111–116 (August 2005).
- [19] Feng, Y. C., Yuen, P., and Jain, A., "A hybrid approach for face template protection," in [*Proceedings of SPIE Conference of Biometric Technology for Human Identification*], **6944** (2008).
- [20] Zuo, J., Ratha, N. K., and Connell, J. H., "Cancelable iris biometric," in [*Proceedings of the 19th International IAPR Conference on Pattern Recognition (ICPR 2008)*], 1–4 (2008).
- [21] Farooq, F., Bolle, R., Jea, T., and Ratha, N., "Anonymous and revocable fingerprint recognition," in [*Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*], (June 2007).
- [22] Tulyakov, S., Farooq, F., Mansukhani, P., and Govindaraju, V., "Symmetric hash functions for secure fingerprint biometric systems," *Pattern Recognition Letters* **28**(16), 2427–2436 (2007).
- [23] Ratha, N. K., Chikkerur, S., Connell, J. H., and Bolle, R. M., "Generating Cancelable Fingerprint Templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**, 561–572 (April 2007).
- [24] Lee, C., Choi, J.-Y., Toh, K.-A., and Lee, S., "Alignment-free cancelable fingerprint templates based on local minutiae information," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **37**(4), 980–992 (2007).
- [25] Sutcu, Y., Rane, S., Yedidia, J., Draper, S., and Vetro, A., "Feature extraction for a slepian-wolf biometric system using ldpc codes," in [*Proceedings of the IEEE International Symposium on Information Theory*], (July 2008).
- [26] Feng, Y. C., Yuen, P. C., and Jain, A. K., "A hybrid approach for generating secure and discriminating face template," *IEEE Transactions on Information Forensics and Security* (2010). To appear.
- [27] Nagar, A. and Jain, A. K., "On the Security of Non-Invertible Fingerprint Template Transforms," in [*Proceedings of IEEE Workshop on Information Forensics and Security*], (December 2009).
- [28] Quan, F., Fei, S., Anni, C., and Feifei, Z., "Cracking cancelable fingerprint template of ratha," in [*International Symposium on Computer Science and Computational Technology.*], **2**, 572–575 (2008).
- [29] Shin, S. W., Lee, M.-K., Moon, D., and Moon, K., "Dictionary attack on functional transform-based cancelable fingerprint templates," *ETRI Journal* **31**(5), 628–630 (2009).
- [30] Nandakumar, K., Jain, A. K., and Pankanti, S., "Fingerprint-based Fuzzy Vault: Implementation and Performance," *IEEE Transactions on Information Forensics and Security* **2**, 744–757 (December 2007).
- [31] Turk, M. and Pentland, A., "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience* **3**(1), 71–86 (1991).
- [32] Nanni, L. and Lumini, A., "Random subspace for an improved biohashing for face authentication," *Pattern Recognition Letters* **29**(3), 295–300 (2008).