

How to Protect Biometric Templates

Yagiz Sutcu^a, Qiming Li^b and Nasir Memon^b

^aPolytechnic University, Electrical & Computer Engineering Dept., Brooklyn, NY, USA;

^bPolytechnic University, Computer & Information Science Dept., Brooklyn, NY, USA

ABSTRACT

In addition to the inherent qualities that biometrics possess, powerful signal processing tools enabled widespread deployment of the biometric-based identification/verification systems. However, due to the nature of biometric data, well-established cryptographic tools (such as hashing, encryption, etc.) are not sufficient for solving one of the most important problems related to biometric systems, namely, template security. In this paper, we examine and show how to apply a recently proposed secure sketch scheme in order to protect the biometric templates. We consider face biometrics and study how the performance of the authentication scheme would be affected after the application of the secure sketch. We further study the trade-off between the performance of the scheme and the bound of the entropy loss from the secure sketch.

Keywords: Biometrics, template security, secure sketch, quantization, entropy loss

1. INTRODUCTION

Biometric authentication schemes have great potentials in building secure systems since biometric data of the users are bound tightly with their identities, and cannot be forgotten. Typically, a biometric authentication scheme consists of two phases. During the enrollment phase, a user, Alice, registers her biometric data with a trusted server. A biometric template is created for Alice, and is stored on some central server or a device carried by Alice (e.g., smartcard). During the authentication phase, Alice would provide another biometric sample, which is then compared with the template in the server or device, and Alice is authenticated if the new sample matches the template according to some matching function.

In this scenario, the biometric template of Alice would contain crucial information for successful authentication. Once revealed, Alice's template would potentially allow an attacker to obtain sufficient information to impersonate Alice. Hence it is important to prevent attackers from learning the biometric templates of the users. This is a very challenging issue because it is extremely difficult to build a server or a device that can never be compromised, and once compromised, the biometric templates cannot be revoked like passwords. This problem is intrinsically more difficult compared to traditional authentication systems based on passwords or certificates, where compromised user credentials can be easily revoked. Furthermore, since biometric data are prone to various noise during sampling and/or processing, we cannot use a cryptographic hash function to hide the biometric templates in the same way we hide passwords.

One method to hide biometric templates and allow robust authentication at the same time is to use recently proposed secure sketch schemes⁽¹⁾. In such a scheme, a sketch P is computed from the original template X , and when another sample Y of the same biometrics is obtained, X can be recovered from P and Y , if Y is similar to X according to some similarity measure. Such a sketch P is secure if it contains little information about the original biometric template X of the user. More precisely, the security of the scheme is measured by the average min-entropy of X given P , which is denoted as $\tilde{H}_\infty(X | P) = -\log(\mathbb{E}_{b \leftarrow P}[2^{-H_\infty(X|P=b)}])$, where the min-entropy $H_\infty(X) = -\log(\max_a \Pr[X = a])$. A general method is given in¹ to bound such entropy loss from above for any distribution of X , which is useful since the distributions of many biometrics are not known.

Further author information: (Send correspondence to Yagiz Sutcu)

Yagiz Sutcu: E-mail: ygzstc@yahoo.com, web: <http://isis.poly.edu>

Qiming Li: E-mail: qiming.li@ieee.org, web: <http://meta.poly.edu/liqm/>

Nasir Memon: E-mail: memon@poly.edu, web: <http://isis.poly.edu/memon/>

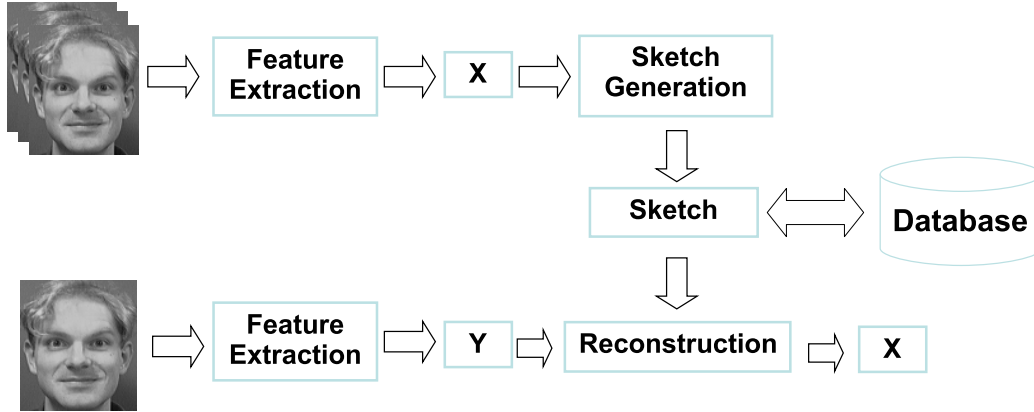


Figure 1. Sketch Generation and Template Reconstruction

In an authentication scheme, when Alice’s original biometric template X is reconstructed from P and Y , we can further extract a cryptographic key K from X , such that K is almost uniformly distributed, and use K to perform authentication similar to traditional methods. For example, we can treat the key K as the “password” of Alice. Generally speaking, if the min-entropy of the original X can be bounded from below, the entropy loss also gives an implicit lower bound on the length of the key K that we can extract from the biometric data. In other words, the less the entropy loss, the longer the key we can extract, and the more difficult it is for an attacker to impersonate Alice by guessing the value of the key K .

Although these schemes can be provably secure in terms of entropy loss, there are still gaps when they are applied to real world biometric templates. Most importantly, many biometric templates are not discrete, but are instead points in continuous domains (e.g., real numbers resulted from some signal processing techniques). In this case, it would be hard to define what the min-entropy of the original biometric template should be. Furthermore, to extract a discrete key from such a template, some kind of quantization would be necessary. However, since the formulation of secure sketch requires that the original X can be reconstructed exactly, the entropy loss could be arbitrarily high, which can be misleading.

Furthermore, the false accept and false reject ratios of the system are performance parameters that are crucial for the security and usability of the system. Basically, false acceptance rate (FAR) measures the ratio of impostors who are incorrectly accepted as legitimate users, and false rejection rate (FRR) measures the ratio of valid users who are rejected as impostors in a biometric system. Typically, these values (FAR and FRR) can be traded off against each other by changing some tunable threshold parameter in order to obtain the ROC curve of the system. But unfortunately they are not captured by the notion of secure sketch, basically because nothing is required when X and Y are *not* similar. For instance, when the false accept rate of the system is one per cent, an attacker is expected to succeed with probability of 0.01, regardless of the entropy loss of the secure sketch. However, we cannot have a system with arbitrarily low false accept rate, since the false reject rate could be too high for the system to be practical. From another point of view, the “min-entropy” of the original X is not a fixed value, but rather a parameter that is closely related to the false accept rate, which can be traded-off with the false reject rate.

In this paper, we examine an authentication scheme using face biometrics and show how to apply a known secure sketch scheme² on top of it to protect the biometric templates. We study how the performance of the system in terms of false accept and false reject rates would be affected after the application of the secure sketch. We further study the trade-off between the performance of the system and the bound of the entropy loss from the secure sketch. Our results can be used as a guideline in designing secure biometric authentication systems.

2. RELATED WORK

The fuzzy commitment scheme³ introduced by Juels and Wattenberg, which is based on binary error-correcting codes can be considered one of the earlier approaches very similar to the secure sketch schemes. In this scheme biometric templates are represented as binary strings where the similarity is measured by Hamming distance. Later, Juels and Sudan proposed the fuzzy vault scheme⁴ which considers sets of elements in a finite field with set difference as the distance function, and corrects errors by polynomial interpolation. Dodis et al.¹ further gives the notion of *fuzzy extractors*, where a “strong extractor” (such as pair-wise independent hash functions) is applied after the original X is reconstructed to obtain an almost uniform key. Constructions and rigorous analysis of secure sketch are given for three metrics: Hamming distance, set difference and edit distance. Secure sketch schemes for point sets⁵ are motivated by the typical similarity measure used for fingerprints, where each template consists of a set of points in 2-D space, and the similarity measure does not define a metric space.

Linnartz and Tuyls⁶ consider a similar problem for biometric authentication applications. They consider zero mean i.i.d. jointly Gaussian random vectors as biometric templates, and use mutual information as the measure of security against dishonest verifiers. Tuyls and Goseling⁷ consider a similar notion of security, and develop some general results when the distribution of the original is known and the verifier can be trusted. Later, Tuyls et al.⁸ further analyzed the same problem and presented some practical results along this line.

Reusability of the sketches⁹ is studied by Boyen et al. and shown that a sketch scheme that is provably secure may become insecure when multiple sketches of the same biometric data are compromised. Boyen et al. further study the security of secure sketch schemes under more general attacker models and techniques to achieve mutual authentication¹⁰ are proposed.

In addition to the above, there are many other approaches which address the same problems. The concept of cancelable biometrics¹¹ was first introduced by Ratha et al. and a case study for fingerprints¹² is presented recently. The underlying idea of the scheme is to store a distorted version of the template through the use of a fixed non-invertible transformation which is applied directly at the sensor. Although this approach gives the opportunity to cancel that template and corresponding transformation when the biometric data and/or transformations are compromised, it is not a trivial job to design such transformations.

Davida et al.¹³ were among the first to propose an off-line biometric authentication scheme based on error correcting codes (ECC) for iris. They suggested storing a signed form of biometric template in a portable storage device, like smartcard, instead of a database and matching the biometric data locally. However, despite the provable security of the algorithm, the error-correcting bits in the database leak some amount of information about biometric data of the user. Connie et al.,¹⁴ Teoh et al.¹⁵ and Jin et al.¹⁶ proposed similar threshold-based biometric hashing methods. Although high performance is achieved for faces, fingerprints and palmprints, Teoh et al.¹⁶ showed that when the tokenized random numbers are stolen and used by impostors, the system performance becomes unacceptable for real-world applications. Similarly, Tulyakov et al. presented another approach and proposed a set of symmetric hash functions¹⁷ for minutiae based fingerprint representations.

Another quantization and ECC-based method for creating renewable binary templates for face recognition is proposed by Kevenaar et al.¹⁸ Although the quantization operation is non-invertible, high error correcting capability requirement for coding part and poor performance results make scheme impractical for real-world applications as mentioned by the authors. Similarly, for minutiae based fingerprint templates, Ang et al.¹⁹ proposed a key-based geometric transformation which can be canceled in case of compromise.

There have been a number of works on how to extract consistent keys from real biometric templates, which have quite different representations and similarity measures from the above theoretical works. Such biometric templates include handwritten online signatures,²⁰ iris patterns,²¹ voice features,²² and face biometrics.²³ Similarly, Vielhauer et al.²⁴ proposed a simple method to calculate biometric hash values using statistical features of online signatures. A key binding algorithm²⁵ is proposed by Soutar et al. and a face recognition scheme based on minimum average correlation energy filters²⁶ is proposed by Savvides et al. These works, however, do not have sufficiently rigorous treatment of the security, compared to well-established cryptographic techniques. Some of the works give analysis on the entropy of the biometrics, and approximated amount of efforts required by a brute-force attacker.

3. PRELIMINAIRES

In a recent work, we consider the problem of designing and analyzing secure sketch for biometric templates in continuous domains² and study how to design and analyze different quantization algorithms. Since it is very difficult to have a general algorithm to find the “optimal” quantizer, we instead examine the *relative entropy loss* for any given class of quantizers, which, for any given quantizer in that class, measures the number of additional bits we could have extracted if an optimal quantizer was used in the first place.

In this section, we will briefly summarize the basic concepts and definitions related to the quantization-based secure sketch scheme.

3.1. Entropy and Entropy Loss in Discrete Domain

In the case where X is discrete, we follow the definitions by Dodis et al.¹ They consider a variant of the *average min-entropy* of X given P , which is essentially the minimum strength of the key that can be consistently extracted from X when P is made public.

In particular, the min-entropy $H_\infty(A)$ of a discrete random variable A is defined as

$$H_\infty(A) = -\log(\max_a \Pr[A = a]) \quad (1)$$

Similarly, for two discrete random variables A and B , the average min-entropy of A given B is defined as

$$\tilde{H}_\infty(A | B) = -\log(\mathbb{E}_{b \leftarrow B}[2^{-H_\infty(A|B=b)}]) \quad (2)$$

For discrete X , the entropy loss of the sketch P is defined as $\mathcal{L} = H_\infty(X) - \tilde{H}_\infty(X|P)$. This definition is useful in the analysis, since for any ℓ -bit string B , we have $\tilde{H}_\infty(A | B) \geq H_\infty(A) - \ell$. For any secure sketch scheme for discrete X , let R be the randomness invested in constructing the sketch, it is not difficult to show that when R can be computed from X and P , we have

$$\mathcal{L} = H_\infty(X) - \tilde{H}_\infty(X | P) \leq |P| - H_\infty(R). \quad (3)$$

In other words, the entropy loss can be bounded from above by the difference between the size of P and the amount of randomness we invested in computing P . This allows us to conveniently find an upper bound of \mathcal{L} for any distribution of X , since it is independent of X .

3.2. Secure Sketch in Discrete Domain

Here we repeat the definitions of secure sketch and entropy loss in the discrete domain given by Dodis et al.¹ Let \mathcal{M} be a finite set of points with a *similarity* relation $\mathbf{S} \subseteq \mathcal{M} \times \mathcal{M}$. When $(X, Y) \in \mathbf{S}$, we say the Y is similar to X , or the pair (X, Y) is similar.

Definition 1: A sketch scheme in discrete domain is a tuple $(\mathcal{M}, \mathbf{S}, \text{ENC}, \text{DEC})$, where $\text{ENC} : \mathcal{M} \rightarrow \{0, 1\}^*$ is an encoder and $\text{DEC} : \mathcal{M} \times \{0, 1\}^* \rightarrow \mathcal{M}$ is a decoder such that for all $X, Y \in \mathcal{M}$, $\text{DEC}(Y, \text{ENC}(X)) = X$ if $(X, Y) \in \mathbf{S}$. The string $P = \text{ENC}(X)$ is the sketch, and is to be made public. We say that the scheme is \mathcal{L} -secure if for all random variables X over \mathcal{M} , the entropy loss of the sketch P is at most \mathcal{L} . That is, $H_\infty(X) - \tilde{H}_\infty(X | \text{ENC}(X)) \leq \mathcal{L}$.

We call $\tilde{H}_\infty(X | P)$ the *left-over entropy*, which in essence measures the “strength” of the key that can be extracted from X given that P is made public. Note that in most cases, the ultimate goal is to maximize the left-over entropy for some particular distribution of X . However, in the discrete case, the min-entropy of X is fixed but can be difficult to analyze. Hence, entropy loss becomes an equivalent measure which is easier to quantify.

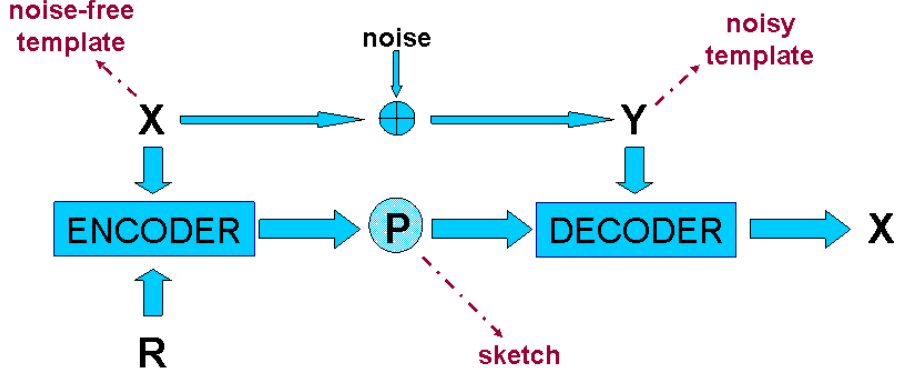


Figure 2. Sketch Generation and Reconstruction in Discrete Domain

3.3. Secure Sketch in Continuous Domain

To handle points in some continuous domain \mathcal{U} , we follow² and use a two-step approach. In particular, we quantize (discretize) the points such that they become points in a discrete domain \mathcal{M} . After that we apply known sketch scheme in discrete domain \mathcal{M} to construct the sketch. When a fresh measurement of the same biometrics is given, it is quantized using the same quantizer and the corresponding reconstruction algorithm in the discrete domain is used to recover the quantized version of the original data points.

More formally, let \mathcal{U} be a set that may be uncountable, and let \mathbf{S} be a similarity relation on \mathcal{U} , i.e., $\mathbf{S} \subseteq \mathcal{U} \times \mathcal{U}$. Let \mathcal{M} be a set of finite points, and let $\mathcal{Q} : \mathcal{U} \rightarrow \mathcal{M}$ be a function that maps points in \mathcal{U} to points in \mathcal{M} . We will refer to such a function \mathcal{Q} as a *quantizer*.

Definition 2: A quantization-based sketch scheme is (as defined in²) a tuple $(\mathcal{U}, \mathbf{S}, \mathcal{Q}, \mathcal{M}, \text{ENC}, \text{DEC})$, where $\text{ENC} : \mathcal{M} \rightarrow \{0, 1\}^*$ is an encoder and $\text{DEC} : \mathcal{M} \times \{0, 1\}^* \rightarrow \mathcal{M}$ is a decoder such that for all $X, Y \in \mathcal{U}$, $\text{DEC}(\mathcal{Q}(Y), \text{ENC}(\mathcal{Q}(X))) = \mathcal{Q}(X)$ if $(X, Y) \in \mathbf{S}$. The string $P = \text{ENC}(\mathcal{Q}(X))$ is the sketch. We say that the scheme is \mathcal{L} -secure in the quantized domain if for all random variable X over \mathcal{U} , the entropy loss of P is at most \mathcal{L} , i.e., $H_\infty(\mathcal{Q}(X)) - \tilde{H}_\infty(\mathcal{Q}(X) | \text{ENC}(\mathcal{Q}(X))) \leq \mathcal{L}$

It is worth to note that according to this definition, we only require the quantized original to be reconstructed. This, in some sense, avoids the problem of possible high entropy loss due to quantization. It is shown in² that when the quantization step (assuming scalar quantization) is close to the error that we want to tolerate, the resulting scheme would not be too much different in terms of left-over entropy from using the “optimal” quantization step, which may be difficult to find. Therefore, in this paper we will follow this principle, with some necessary deviation due to be nature of the biometrics in the real world.

4. QUANTIZATION-BASED SKETCH SCHEME FOR FACE BIOMETRICS

In this section, we describe our scheme to compute sketches from face images that allow us to extract consistent keys. Our main idea is as the following: Firstly, for a given image, we extract a feature vector V of size n (Section 4.1). Secondly, we discretize (quantize) the feature vector (Section 4.2) and finally, we apply a known sketch scheme to generate a sketch and to reconstruct the quantized feature vector (Section 4.3).

4.1. Template Representation

We assume that we can extract a feature vector of size n from each biometric sample. Therefore,

$$V_i = [v_{i1} \ v_{i2} \ \dots \ v_{in}]^T \quad (4)$$

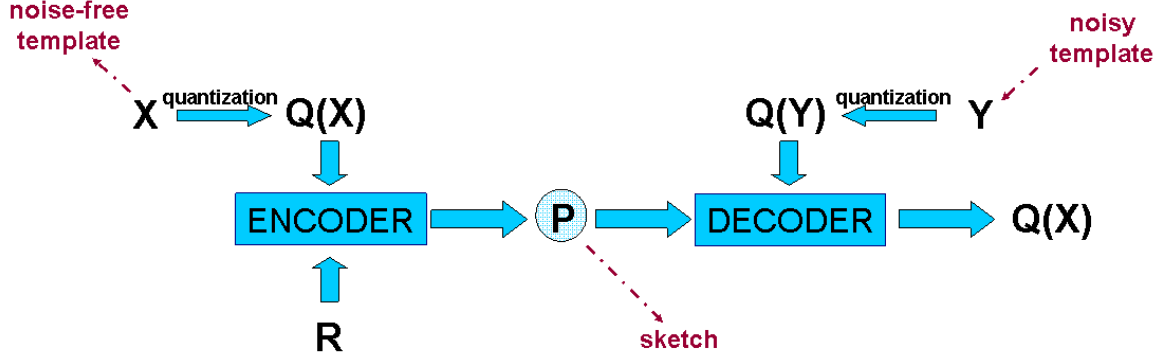


Figure 3. Sketch Generation and Reconstruction in Continuous Domain

represents the n -dimensional feature vector of i -th user of the system where each coefficient $v_{ij} \in R$ is a real number. These coefficients can be extracted from certain transformations on the raw measurement. For example, we can apply singular value decomposition and take the n most significant coefficients (Section 5).

In addition, we also assume that the value of each coefficient v_{ij} can vary within a certain *range*, which is going to be determined through experiments on the data set. In other words, we consider the j -th coefficient for the i -th user to be always associated with a range, which is defined by a *midpoint* \bar{v}_{ij} and a *size* δ_{ij} . Therefore, the template for the i -th user consists of two vectors. The first is the list of n midpoints $\bar{v}_{i1}, \dots, \bar{v}_{in}$, and the other is the list of range sizes for each coefficients $\delta_{i1}, \dots, \delta_{in}$.

In the simplest case, for the i -th user in the system, we can consider a sample $V_i = [v_{i1} \ v_{i2} \ \dots \ v_{in}]^T$ as authentic if

$$\bar{v}_{ij} - \delta_{ij} \leq v_j \leq \bar{v}_{ij} + \delta_{ij} \quad (5)$$

for all $j = 1, \dots, n$.

4.2. Quantization and Codebook

The most important step for generating a sketch for the biometric template is to discretize every component of the feature vector such that we can apply a sketch scheme for discrete domains. For this purpose, we employ a straightforward method, which uses a scalar quantizer for each of the coefficients to map it to a discrete domain.

First, we determine global ranges of each and every component of the feature vectors from the training data set obtained during enrollment phase. Let these values be MN_j and MX_j . More formally:

$$MN_j = \min_i(mn_{ij}) \quad (6)$$

and

$$MX_j = \max_i(mx_{ij}) \quad (7)$$

Next, the discrete domain \mathcal{C}_j for the j -th component is computed by quantizing the overall user range by the quantization step δ_j . That is,

$$\mathcal{C}_j = \{MN_j - r_j, MN_j - r_j + \delta_j, MN_j - r_j + 2\delta_j, \dots, MN_j - r_j + L_j\delta_j\} \quad (8)$$

where L_j is appropriately chosen integer which satisfies $MN_j - r_j + L_j\delta_j \geq MX_j$ and r_j is a positive random number.

In this way, for the j -th component of the i -th user, a range of midpoint \bar{v}_{ij} and size δ_{ij} can be translated to a discrete range where the discrete midpoint is quantization of \bar{v}_{ij} in \mathcal{C}_j , and the discrete range size d_{ij} is given by

$$d_{ij} = \lceil \frac{\delta_{ij}}{\delta_j} \rceil \quad (9)$$

Finally, the codebook C_j^i for the j -th component of the i -th user is a subset of \mathcal{C}_j , and can be determined by choosing one point out of every $2d_{ij} + 1$ consecutive points in \mathcal{C}_j .

In this setup, δ_j 's are simply determined as a function of the minimum range size of each component of the feature vector observed in overall user space. That is,

$$\delta_j = \alpha \min_i(\delta_{ij}) \quad (10)$$

where α is a parameter which can take values in $(0, 1]$.

It is worth noting that, in the above formulation, the quantization step δ_j can be determined in many different ways. However, it is reasonable to assume that, δ_j should be related to some statistics of the range of the feature components, namely δ_{ij} 's.

4.3. Sketch Generation and Template Reconstruction

During enrollment, the biometric data of each user are acquired and feature vectors are extracted several times as a part of training process. Then the variation (i.e., the midpoint and range size) of each feature vector component is estimated by analyzing the training data set. Next, we construct a codebook for each component of each user as in Section 4.2.

Therefore, the sketch P_i for user i is a vector

$$P_i = [p_{i1} \ p_{i2} \ \dots \ p_{ik}]^T. \quad (11)$$

For each p_{ij} we have

$$p_{ij} = Q_j^i(\bar{v}_{ij}) - \bar{v}_{ij} \quad (12)$$

where $Q_j^i(\bar{v}_{ij})$ is the codeword in C_j^i that is closest to \bar{v}_{ij} .

During authentication, biometric data of the i -th user is taken and corresponding feature vector is computed. Let us denote this noisy feature vector as $\tilde{V}_i = [\tilde{v}_{i1} \ \tilde{v}_{i2} \ \dots \ \tilde{v}_{in}]^T$. Then the decoder takes \tilde{V}_i and P_i and calculates $Q_j^i(\tilde{v}_{ij}) - p_{ij}$ for $j = 1, \dots, n$. Reconstruction of the original biometric will be successful if

$$-d_{ij} \leq Q_j^i(\tilde{v}_{ij}) - Q_j^i(\bar{v}_{ij}) < d_{ij} \quad (13)$$

where d_{ij} is the user specific error tolerance bound for the j -th component. It is not difficult to see that, $Q_j^i(\tilde{v}_{ij}) - p_{ij} = Q_j^i(\tilde{v}_{ij}) - Q_j^i(\bar{v}_{ij}) + \bar{v}_{ij}$ and the errors up to the some preset threshold value will be corrected successfully.

4.4. Security

As mentioned earlier, $\tilde{H}_\infty(X | P)$ is called the *left-over entropy*, which measures the “strength” of the key that can be extracted from X given that P is made public and in most cases, the ultimate goal is to maximize the left-over entropy for some particular distribution of the biometric data considered. However, in the discrete case, the min-entropy is fixed but can be difficult to analyze and entropy loss becomes an equivalent measure which is easier to quantify.

For this construction, in order to estimate the left-over entropy, firstly, we tried to estimate the min-entropy of V ($H_\infty(V)$) assuming that the components of the feature vector are independent. Therefore, the min-entropy of each component are estimated independently and the total min-entropy of the feature vector V is calculated as the summation of the individual min-entropies of the components. That is,

$$H_\infty(V) = \sum_{i=1}^n H_\infty(v_i) \quad (14)$$

To estimate $H_\infty(v_i)$, we simply considered the distribution of the feature vector component v_i over all user space and analyzed the histogram of that distribution while setting the bin size to the quantization step size δ_i

of that component. The number of elements in the most likely bin gives a rough estimate of the min-entropy of the feature vector component i .

The (component-wise) entropy loss in the quantized domain can simply be bounded by

$$\mathcal{L}(P) \leq \sum_{i=1}^n \mathcal{L}(p_i) \tag{15}$$

where $\mathcal{L}(p_i)$ is the entropy loss of the sketch for the component i of the feature vector representation of the biometric data. This can be conveniently bounded by the size of the sketch. That is,

$$\mathcal{L}(p_i) \leq |p_i| = \log(2 \lceil \frac{\delta_{ij}}{\delta_j} \rceil + 1). \tag{16}$$

We note that the “entropy loss” is a worst case bound, which states that there exists an input distribution that will give such amount of information leakage, but not necessarily the distribution for the particular biometric data. In other words, the entropy loss is an upper bound of the information leakage and the estimation of entropy loss may not be accurate in reality. Whether the attacker can really gain that much information needs to be further studied for particular distributions. We consider it an open question to bound the “exact information leakage” of the sketch.

5. EXPERIMENTS AND RESULTS

Face images are one of the widely used biometrics for recognition and authentication purposes. In our experiments, we use the Essex Faces 94 face database (E94 database),²⁷ which is publicly available and essentially created for face recognition related research studies. The E94 database contains images of 152 distinct subjects, with 20 different images for each subject where the size of each JPEG image is 180x200. We first transformed these images to 8-bit gray level images and then use these gray level images in our experiments. For each subject, we randomly chose the 12 out of 20 samples for training and the remaining 8 sample face images are used for validation. Sample images from the E94 database are given in Figure 4.

Due to the established properties of singular values, many face image based biometric recognition systems proposed in recent years²⁸⁻³⁰ are used singular values as features. Therefore, we also used them for testing our scheme. In our simulations, only first 20 singular values of the images are considered. However, it should be noted that the essence of the technique is not specific to face image data and can be applied to any type of ordered biometric features.

During the enrollment stage, feature vectors (simply the first 20 singular values of the training face images) are calculated and then the variation of each feature vector component is estimated by analyzing the training data set. At the authentication stage, biometric data of the user is taken and corresponding feature vector which will be queried is created. Then, this noisy biometric data is mapped to the set of closest codewords of the corresponding codebooks (as explained in Section IV) and checked for legitimacy.

In order to evaluate the performance of the proposed scheme, 8 test data for every user is used to generate 152x8=1216 genuine authentication attempts and 151x152x8=183616 impostor authentication attempts (8 attempts by 151 remaining users for every user in the system). To be able to obtain the complete ROC curve, we calculated FAR and FRR values by varying the quantization step size, δ_i . However, it should be noted that, once d_{ij} values are calculated, they are fixed and did not changed during the determination of the ROC curve.

As already mentioned earlier, the quantization step δ_j can be determined in many different ways depending on operational constraints (such as the noise level which needs to be tolerated) and also depending on the data set considered. Here, we considered a straightforward approach and set the quantization step to be a fraction of the minimum range observed over the whole data set (i.e., $\delta_j = \alpha \min_i(\delta_{ij})$).

Figure 5 shows the effect of α on the performance of the scheme for 3 different values of α . As can be seen from Figure 5, small values of α seem to improve the performance of the scheme. However, it is easy to observe that decreasing α to below 0.5 has no significant effect on the performance.



Figure 4. Some examples from E94 database

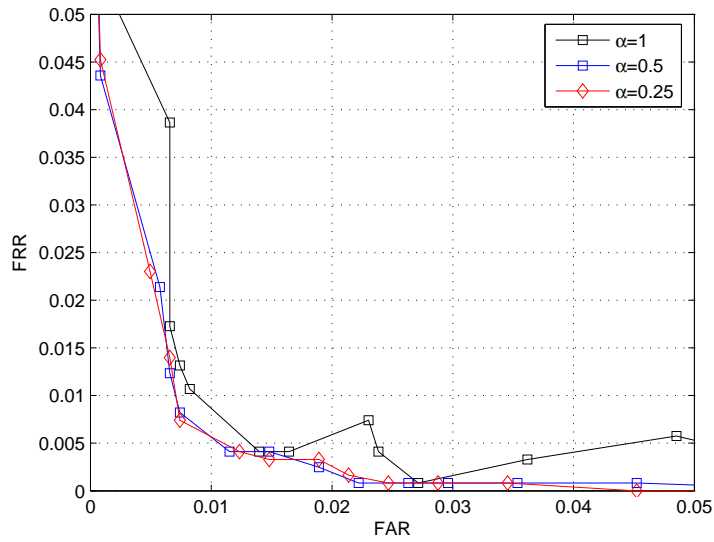


Figure 5. Comparison of the performance for different values of α

As mentioned in Section 4.4, we estimate the min-entropy of the original biometrics by computing the min-entropy of the quantized feature vectors, assuming that all components are independent, and using quantization steps that are equal the minimum errors to be tolerated for each component (i.e., $\alpha = 1$). Under this setting, the min-entropy of the feature vectors is estimated to be about 108 bits.

On the other hand, the entropy loss of the scheme can be calculated by the size of the sketch. Since the errors to be tolerated are quite different for different users even for the same component, the resulting entropy loss is much larger than the theoretically achievable $n \log 3$. From the experiments, the average size of the sketch is about 73 bits, which gives a guarantee of 35 bits in the left-over entropy. Nevertheless, as we noted earlier, this is just a lower bound for the left-over entropy, and the exact security requires further investigation.

Another possible way to obtain better left-over entropy is to reduce the value of α . As we observed earlier, this would not gain much advantage in performance. Furthermore, having a smaller α might actually decrease the left-over entropy. This can happen for certain distribution of X where decreasing α does not increase the min-entropy of quantized X , but increases the information leakage (because of the now larger sketch size). Therefore, we would recommend using $\alpha = 1$.

6. CONCLUSIONS

In this paper we study the problem of secure storage of biometric templates in biometric-based authentication systems. We examine a recently proposed cryptographic primitive called secure sketch and identify several practical issues when we apply known theoretical results to real biometrics. We give a concrete construction of secure sketch for face biometrics, and we illustrate the subtleties and difficulties in applying theoretical bounds. We show various trade-offs among different parameters of the system.

In particular, we note that the security measure in terms of entropy loss may not be sufficient since FAR and FRR should also be taken into consideration of a practical system. Furthermore, entropy loss alone could be just too large to be meaningful, or sometimes becomes misleading, especially when the original biometrics are represented in continuous domains.

We consider it as a challenging open problem to find a general and accurate way to compute the min-entropy (or any quantitative means that measures the success probability of smart attackers) of biometric data, and to determine the exact information leakage of the sketches. It seems that, at least in some cases, known theoretical results become not very useful and the exact security of the system needs to be further investigated.

REFERENCES

1. Y. Dodis, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," in *Eurocrypt, LNCS 3027*, pp. 523–540, Springer-Verlag, 2004.
2. Q. Li, Y. Sutcu, and N. Memon, "Secure sketch for biometric templates," in *Asiacrypt*, (Shanghai, China), December 2006.
3. A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *Proc. ACM Conf. on Computer and Communications Security*, pp. 28–36, 1999.
4. A. Juels and M. Sudan, "A fuzzy vault scheme," in *IEEE Intl. Symp. on Information Theory*, 2002.
5. E.-C. Chang and Q. Li, "Hiding secret points amidst chaff," in *Eurocrypt*, 2006.
6. J.-P. M. G. Linnartz and P. Tuyls, "New shielding functions to enhance privacy and prevent misuse of biometric templates," in *AVBPA 2003*, pp. 393–402, 2003.
7. P. Tuyls and J. Goseling, "Capacity and examples of template-protecting biometric authentication systems," in *ECCV Workshop BioAW*, pp. 158–170, 2004.
8. P. Tuyls, A. Akkermans, T. Kevenaer, G. Schrijen, A. Bazen, and R. Veldhuis, "Practical biometric authentication with template protection," in *AVBPA*, pp. 436–446, 2005.
9. X. Boyen, "Reusable cryptographic fuzzy extractors," in *Proceedings of the 11th ACM conference on Computer and Communications Security*, pp. 82–91, ACM Press, 2004.
10. X. Boyen, Y. Dodis, J. Katz, R. Ostrovsky, and A. Smith, "Secure remote authentication using biometric data," in *Eurocrypt*, 2005.
11. N. Ratha, J. Connell, and R. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Systems Journal* **40**(3), pp. 614–634, 2001.
12. N. Ratha, J. Connell, and R. Bolle, "Cancelable biometrics: A case study in fingerprints," in *18th International Conference on Pattern Recognition (ICPR 2006)*, 2006.
13. G. Davida, Y. Frankel, and B. Matt, "On enabling secure applications through off-line biometric identification," in *Proc. IEEE Symp. on Security and Privacy*, pp. 148–157, 1998.
14. T. Connie, A. Teoh, M. Goh, and D. Ngo, "Palmhashing: a novel approach for cancelable biometrics," *Information Processing Letters* **93**, pp. 614–634, 2005.
15. A. Teoh, D. Ngo, and A. Goh, "Personalised cryptographic key generation based on facehashing," *Computers and Security* **23**, pp. 606–614, 2004.

16. A. B. Teoh and D. C. Ngo, "Cancellable biometrics featuring with tokenized random number," *Pattern Recognition Letters* **26**, pp. 1454–1460, 2005.
17. S. Tulyakov, F. Farooq, and V. Govindaraju, "Symmetric hash functions for fingerprint minutiae," in *Lecture Notes in Computer Science, LNCS*, 2005.
18. T. Kevenaer, G. Schrijen, M. V. der Veen, A. Akkermans, and F. Zuo, "Face recognition with renewable and privacy preserving binary templates," *Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, pp. 21–26, 2005.
19. R. Ang, R. Safavi-Naini, and L. McAven, "Cancelable key-based fingerprint templates," in *Lecture Notes in Computer Science, LNCS*, 2005.
20. F. Hao and C. Chan, "Private key generation from on-line handwritten signatures," *Information Management and Computer Security* **10**(2), 2002.
21. F. Hao, R. Anderson, and J. Daugman, "Combining cryptography with biometrics effectively," Tech. Rep. UCAM-CL-TR-640, University of Cambridge, 2005.
22. F. Monrose, M. Reiter, Q. Li, and S. Wetzel, "Cryptographic key generation from voice," in *IEEE Symp. on Security and Privacy*, 2001.
23. Y. Sutcu, T. Sencar, and N. Memon, "A secure biometric authentication scheme based on robust hashing," in *ACM MM-SEC Workshop*, 2005.
24. C. Vielhauer, R. Steinmetz, and A. Mayerhoefer, "Biometric hash based on statistical features of online signatures," *IEEE International Conference on Pattern Recognition (ICPR)*, 2002.
25. C. Soutar, D. Roberge, S. Stojanov, R. Gilroy, and B. V. Kumar, "Biometric encryption using image processing," in *SPIE, Optical Security and Counterfeit Deterrence Techniques II*, **3314**, 1998.
26. M. Savvides, B. V. Kumar, and P. Khosla, "Cancelable biometric filters for face recognition," *Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004* **3**, pp. 922–925, 2004.
27. "The essex faces94 database, available at <http://cswww.essex.ac.uk/mv/allfaces/index.html>."
28. Z. Hong, "Algebraic feature extraction of image for recognition," *Pattern Recognition* **24**, pp. 211–219, 1991.
29. Y.-Q. Cheng, "Human face recognition method based on the statistical model of small sample size," *SPIE Proceedings of the Intell. Robots and Comput. Vision* **1607**, pp. 85–95, 1991.
30. W.Y.-Hong, T.T.-Niu, and Z. Yong, "Face identification based on singular value decomposition and data fusion," *Chinese J. Comput. (in Chinese)* **23**, 2000.