A Dissimilarity-Based Approach for Biometric Fuzzy Vaults–Application to Handwritten Signature Images

George S. Eskander, Robert Sabourin, and Eric Granger

Laboratoire d'imagerie, de vision et d'intelligence artificielle
Ecole de technologie supérieure, Université du Québec, Montréal, Canada
geskander@livia.etsmtl.ca, robert.sabourin@etsmtl.ca, eric.granger@etsmtl.ca

Abstract. Bio-Cryptographic systems enforce authenticity of cryptographic applications like data encryption and digital signatures. Instead of simple user passwords, biometrics, such as, fingerprint and handwritten signatures, are employed to access the cryptographic secret keys. The Fuzzy Vault scheme (FV) is massively employed to produce bio-cryptographic systems, as it absorbs variability in biometric signals. However, the FV design problem is not well formulated in the literature, and different approaches are applied for the different biometric traits. In this paper, a generic FV design approach, that could be applied to different biometrics, is introduced. The FV decoding functionality is formulated as a simple classifier that operates in a dissimilarity representation space. A boosting feature selection (BFS) method is employed for optimizing this classifier. Application of the proposed approach to offline signature biometrics confirms its viability. Experimental results on the Brazilian signature database (that includes various forgeries) have shown FV recognition accuracy of 90% and system entropy of about 69-bits.

Keywords: Fuzzy Vault, Bio-Cryptography, Offline Signatures, Dissimilarity Representation.

1 Introduction

Cryptographic systems deal with the security of stored or transmitted information. For instance, data encryption methods provide a way for confidentiality, as only authorized persons, who possess the decryption key, can decrypt and understand the encrypted information. Also, the digital signature technology guarantee data integrity. A person digitally signs and transmits the information, by means of his signing key. A receiving party can verify that the received information are not changed during transmission.

The drawback of cryptography lies in its dependency on secret cryptographic keys, that if compromised, security of the whole system is compromised. Although the cryptographic keys are too long to be guessed by impostors, they are also too long to be memorized by the legitimate users. This problem is alleviated through storing the key in a secure place, e.g., a smart card, and a user retrieves his key by providing a simple password. Such token/password solution forms a weak point in a security system,

as whatever strong is the cryptographic key, overall system security is determined by the password length. Moreover, these authentication measures are not strongly associated with the user identity, so they cannot really distinguish between attackers and legitimate users. Any person who steals the password and the card can retrieve the cryptographic key and access the system.

Biometrics, which are human traits like fingerprint, iris, face, handwritten signatures, etc., are employed to alleviate the cryptographic key management problem. This new technology is known as bio-cryptography or crypto-biometrics, where cryptographic keys are secured by means of biometrics instead of the traditional passwords [1]. As biometrics are strongly associated with user identity, and it is less likely that they are stolen or forgotten, so they guarantee authenticity of the cryptographic systems users. However, design of the bio-cryptographic systems is challenging due to the fuzzy nature of the biometric traits. The intra-personal variability and inter-personal similarity of biometric signals lead to false rejection of authorized users and acceptance of unauthorized users, respectively.

The Fuzzy Vault Scheme (FV) [2] is employed to produce bio-cryptographic systems based on different biometric traits. This scheme locks the cryptographic key by means of locking features extracted from a biometric template. To unlock the FV, and retrieve the key, a genuine biometric sample is used to extract some unlocking features. The key can be unlocked only if the locking and unlocking features overlap substantially. As identical matching between the biometric template and the query sample is not required, so the FV absorbs some of the biometric variability.

In literature, different design approaches are proposed to design FV systems based on different biometric traits, and no generic approach is introduced. In this paper, a generic FV design approach is introduced. The functionality of the FV system is formulated as a simple classifier that operates in a dissimilarity space [3], where distances between a query sample and a FV constitute the classification space. Optimizing the parameters and input features of this classifier lead to accurate FV decoding. As a proof of concept, the proposed formulation is applied to the offline signature biometrics, and have shown promising results. The boosting feature selection method (BFS) [4] is employed here to optimize the FV model, while different optimization techniques may be applicable.

The rest of this paper is organized as follows. The next section presents a background on the FV scheme as applied to biometrics. The proposed dissimilarity-based formulation of the FV design problem is illustrated in section 3. Application of the proposed method to the offline signature biometrics, and the experimental results are presented and discussed in section 4.

2 Biometric Fuzzy Vaults

A FV scheme locks a cryptographic key K by means of a biometric template T. To unlock K, a biometric query sample Q is provided by the user. Figure 1 illustrates this locking/unlocking process. For key locking, K is split into $k+1$ strings and constitutes a coefficient vector $C = \{c_0, c_1, c_2, ..., c_k\}$. A polynomial p of degree k is encoded using C, where $p(x) = c_k x^k + c_{k-1} x^{k-1} + ... + c_1 x + c_0$. Then, a locking set $F^T = \{f_i^T\}_{i=1}^t$

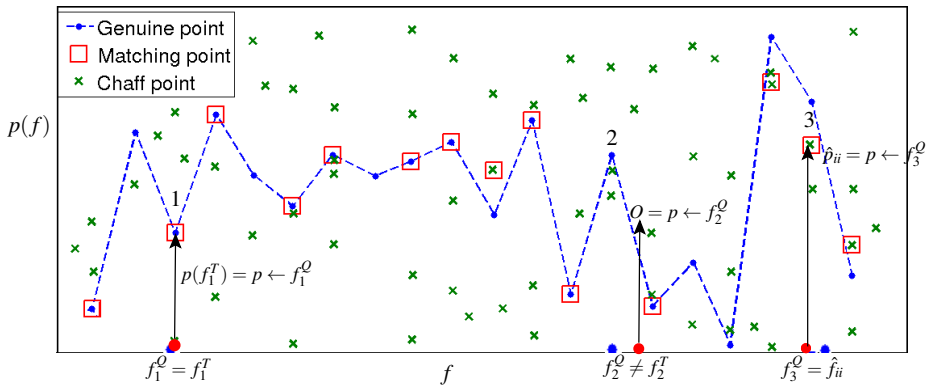

Fig. 1. Illustration of the FV locking/unlocking process.

is extracted from T. The polynomial is evaluated for all points in F^T and constitutes the set $p(F^T) = \{p(f_i^T)\}_{i=1}^t$. The points $(F^T, p(F^T))$ constitutes the genuine vault points.

It is known that, for a polynomial of degree k, only $k+1$ points on its curve are needed to reconstruct the polynomial equation. So, the genuine vault points can be used to reconstruct the polynomial p, and thereby the cryptographic key K. Hence, any person who accesses the genuine points can retrieve the key. Accordingly, to conceal these data from attackers, a set of z chaff (noise) points ($\hat{F} = \{\hat{f}_{ii}\}_{ii=1}^z, \hat{P} = \{\hat{p_{ii}}\}_{ii=1}^z$) are generated. Then, the chaff and genuine points are mixed to constitute the fuzzy vault V_T of length r points. Security of the vault relies on the amount of concealing chaffs. In case that an impostor accesses the vault data, he has to search for at least $k+1$ genuine points, out of $r = t + z$ points of the FV. This search task becomes infeasible with high number of chaff points z.

The proper way to unlock K from the vault V_T, by legitimate users, is to apply a biometric query sample Q. An unlocking set $F^Q = \{f_j^Q\}_{j=1}^t$ is extracted from Q. Then, the chaff points are filtered by matching items of F^Q against all items in V_T. In the ideal case, each feature encoded in F^Q locates the corresponding genuine feature encoded in F^T (e.g., point 1 in Figure 1). On the other hand, due to the fuzzy nature of biometrics, some elements of F^Q differ from their corresponding elements in F^T, and two types of errors might occur, namely erasures and noise. For the erasures case, f_i^Q does not match with any vault point, so it does not add any element to the matching set (e.g., point 2 in Figure 1). For the noise case, a feature f_i^Q might equate a chaff $\hat{f}_{ii}$, so that it adds a noise point $(\hat{f}_{ii}, \hat{p_{ii}})$ to the matching set (e.g., point 3 in Figure 1).

Finally, the resulting matching set is fed to a polynomial reconstruction algorithm, to reconstruct the encoded polynomial p. This process succeeds only if the matching set contains at least $k+1$ genuine points. However, even if enough genuine points exists, it is not possible to differentiate between the genuine and noise points. To overcome this, FV decoders employ error correction codes, like Reed-Solomon (R-S).

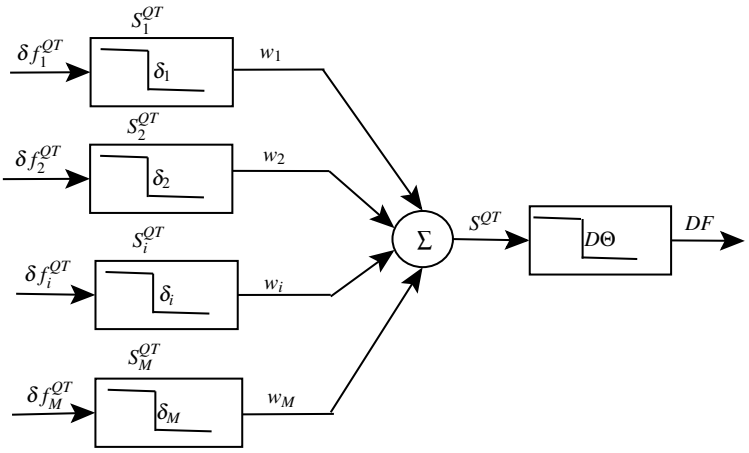

Fig. 2. Proposed formulation of the FV design problem.

The genuine set $(F^T, p(F^T))$ is considered as a code word of length t, that encodes a secret message of length $k+1$, where there are $t-k-1$ redundancy elements. During the decoding process, some noise is added to this code producing a corrupted version of it. The error correction codes can correct some of these errors and recover the secret message.

In literature, the FV design problem is addressed with different approaches. Generally, authors proposed methodologies to absorb the dissimilarities between template and query biometric signals, so that they are within the error correction capacity of the decoder. For instance, fingerprint-based FVs are proposed by Nandakumar et al., where query samples are aligned with the templates [5]. The FV design problem needs to be well formulated, so that a generic approach could be applied to the different biometrics.

3 Dissimilarity-Based Formulation of the FV System

The proposed approach relies on the concept of dissimilarity representation [3]. Instead of representing an object by a set of absolute measurements (features), distances between the objects are considered as features. This concept applies to the FV scheme. A query sample is classified as genuine or forgery, depending on its distance to the FV template.

Figure 2 illustrates the proposed formulation of the FV design problem. Assume $F = \{f_i\}_{i=1}^{M}$ is a vector of features extracted from a biometric signal. The FV decoder does not concern with the absolute feature values, but rather the difference between the locking and unlocking features. Accordingly, the feature representation is translated to a dissimilarity feature representation $\Delta F = \{\delta f_i^{QT}\}_{i=1}^{M}$, where $\delta f_i^{QT} = \|f_i^Q - f_i^T\|$ is the distance between the query and template signals, as measured by the feature f_i. In this new space, it is easier to locate a subset of features with

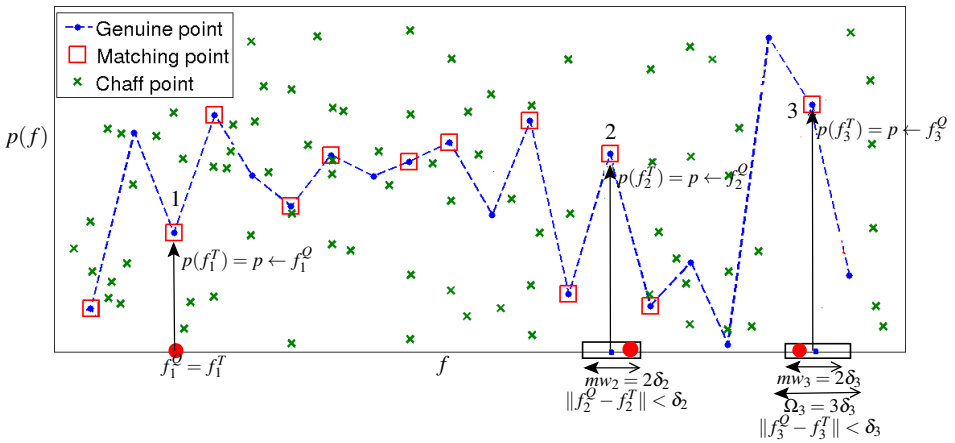

Fig. 3. Illustration of proposed dissimilarity-based FV locking/unlocking process.

similar values for intra-user samples, and that have dissimilar values when extracted from inter-personal samples. For traditional FV decoders, an unlocking feature locates a locking FV point, only if they are identical. Due to the variability of biometrics, it rarely happens that two feature instances are identical. To alleviate this, we modify the FV decoding functionality, so it consideres similar measurements as being identical. In order to define what the term "similar" means, we define $\Delta = \{\delta_i\}_{i=1}^{M}$, where δ_i is a threshold for a dissimilarity feature δf_i and $\delta f_i^{QT} < \delta_i$ only if Q is a genuine sample.

The vector Δ is used during the FV unlocking process, for matching unlocking features with the FV points adaptively, based on the expected feature variability. An unlocking element f_i^Q is considered matching a FV element f_i^T if they lie in a matching window $m_i = 2 \times \delta_i$. Figure 3 illustrates the viability of the adaptive matching a method. On contrary to the strict matching shown in Figure 1, point 2 could be filtered from the chaff points as $\delta f_2^{QT} < \delta_2$. Also, the dissimilarity thresholds vector Δ, is used during the FV locking phase, for adaptive chaff generation. The chaff points are generated so that they have equal separation space Ω. This separation space is computed for each feature f_i, so that $\Omega_i = 3 \times \delta_i$. By this method, it is less likely that an unlocking element f_i^Q equates a chaff element f_{ii}. For instance, point 3 in Figure 3 is filtered, and f_3^Q did not collide with the chaff f_{ii}, as the chaff are generated outside the matching window w_3.

Accordingly, similarity of the samples Q and T, as measured by a feature f_i is given by:

$$S_i^{QT} = \begin{cases} 1 & if\ (\delta f_i^Q < \delta_i) \\ 0 & otherwise \end{cases} \tag{1}$$

The most t discriminative features are selected to encode the FVs, and the overall similarity score between the biometric query samples and its FV template is given by:

$$S^{QT} = \sum_{i=1}^{M} (w_i \times S_i). \tag{2}$$

where $w_i = 1$, only if the feature f_i is selected for FV encoding.

Define the decoding threshold $D\Theta$ as the minimum number of correct elements of a code of length t, that are needed to recover the encoded word of length k. Also, define the error correction capacity ε as the maximum correctable errors, where $\varepsilon = t - D\Theta$. Consider the Berlekamp-Massey algorithm, as a specific R-S error correction code employed for FV decoding. For this code, the $\varepsilon = (t-k-1)/2$, and $D\Theta = (t+k+1)/2$. Accordingly, the functionality of the FV decoder can be formulated as:

$$DF = sign(S^{QT} - D\Theta). \tag{3}$$

where DF is positive if the FV is decoded successfully, and it is negative otherwise. It is obvious that, good design of a FV decoder implies that DF is positive only for genuine query samples. This functionality relies on the elements that produce the similarity score S^{QT} and on the parameters that control the decoding threshold $D\Theta$. These elements and parameters have to be optimized for accurate FV decoding accuracy.

4 Application to the Offline Signature Biometrics

The proposed FV design approach is applied to the offline handwritten signatures, where digitized signature images are used to lock/unlock the FVs. In the preliminary version of this work [6], only the feature selection task is employed. In this paper, the feature dissimilarity is modeled and used to implement the proposed adaptive chaff generation and matching methods.

Although the proposed model of the FV system can be optimized by employing different supervised learning or optimization methods, we employed a Boosting feature selection (BFS) algorithm [3] for feature selection and feature dissimilarity modeling. BFS facilitates selection of a single feature at every boosting iteration, while learning its decision threshold. In case that BFS is employed in a dissimilarity representation space, the algorithm learns the dissimilarity threshold δ_i for every feature feature f_i. To this end, the enrolling signature images are first represented as vectors in a high dimensional feature space. This representation is projected on a dissimilarity representation space, where pairwise feature distances are computed. Boosting feature selection is employed in this space, producing a compact space with the intra-personal distances are minimized and the inter-personal distances are maximized. The feature dissimilarity thresholds are modeled in this space. This method is originally introduced by Rivard et al., to develop writer-independent offline signature verification

Table 1. Performance of the proposed FV implementation

Measure %	Previous work[6] (Only ESC)	Proposed Method (ESC+DPDF)	
		Strict matching	Adaptive matching
FRR	25	28	11.53
FAR_{random}	3	2	2.05
FAR_{simple}	7	3	2.39
$FAR_{simulated}$	36	22	24.38
AER	17.75	13.75	10.08

Table 2. Impact of Chaff Quantity on the FV Performance

Chaff separation (Ω)	Without chaff	Fixed separation				Adaptive separation	
		0.2	0.10	0.05	0.025	2δ	3δ
No. of FV points (r)	20	200	400	800	1600	1768	1528
No. of chaff points (z)	0	180	380	780	1580	1748	1508
Security	0-bits	45-bits	52-bits	60-bits	68-bits	69-bits	68-bits
FRR	5.25	11.53	28.94	55.53	75.81	7.03	6.13
FAR_{random}	2.74	2.05	1.06	0.58	0.31	2.40	2.31
FAR_{simple}	3.49	2.39	1.58	0.88	0.49	2.89	3.26
$FAR_{simulated}$	33.14	24.38	15.63	8.15	3.42	29.77	31.06
AER_{all}	11.15	10.08	11.80	16.28	20.00	10.52	10.69

systems [7]. Recently, we adapted these systems for specific users by tuning their universal representations for specific users [8]. In this paper, training of such user-specific verification systems is employed in the dissimilarity representation space, as a wrapper to select discriminative FV locking features, and to learn the dissimilarity thresholds. This process is out of the scope of this paper, and more details are found in [9].

In the preliminary version of this work [6], we employed Extended-Shadow-Code (ESC) features. Here, we investigate a multi-type feature extraction approach, where Directional Probability Density Function (DPDF) is also employed. The Brazilian database is used for proof-of-concept simulations. The FV system is tested for 60 users, with 40 query samples per user (10 genuine signatures, 10 random, 10 simple, and 10 simulated forgeries). The Average Error Rate (AER) is employed for performance evaluation, where $AER = (FRR + FAR_{random} + FAR_{simple} + FAR_{simulated})/4$. For more details about feature extractions, experimental DB and testing protocol, see [7].

Table 1 reports the performance of the signature-based FVs. In [6], only ESC features are employed. While when the DPDF features are added, the performance is enhanced as AER is reduced from 17.75% to 13.75%. Applying the adaptive matching method, enhanced the performance as AER is reduced from 13.75% to 10.08%.

Table 2 shows the impact of adaptive chaff generation method. It is clear that the FRR is low when no chaff points are generated, while this implementation is not secure. The traditional chaff generation method, is to generate chaff points with fixed separation space between them. In such case, there is a trade-off between security and

robustness. For instance, with small separation, e.g., 0.025, there are 40 FV points generated with the same index (1 genuine + 39 chaff points). In this case, a high number of chaffs (1580) is generated, while system entropy is 68-bits and $FRR = 75\%$. On the other hand, by applying the adaptive chaff generation method, high number of chaffs could be generated (1508), with minimal impact on the system robustness ($FRR = 6\%$).

5 Conclusions and Future Work

In this paper, the FV design problem is formulated based on the dissimilarity representation concept. The proposed formulation facilitates selection of FV encoding features, from large number of feature extractions. Features are translated to a dissimilarity representation space, so their dissimilarities can be modeled. Features with discriminative dissimilarities are selected for FV encoding, and their modeled dissimilarities controls the chaff generation and matching processes. The method is applied to the offline signature biometrics, where boosting feature selection algorithm is employed for training. The proposed method, however, is general and future work will address different biometrics, feature selection and optimization techniques. Also, all of the model parameters maybe optimized for higher FV decoding performance.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada and BancTec Inc.

References

1. U. Uludag, S. Pankanti, S. Prabhakar and A.K. Jain., Biometric Cryptosystems: Issues and Challenges. *Proceedings of the IEEE*, vol.92, issue.6, pp.948-960, 2004.
2. A. Juels and M. Sudan., A Fuzzy Vault scheme. *In proc. IEEE int. Symp. Inf. Theory*, Switzerland, pp.408, 2002.
3. Elzbieta Pekalska , Robert P.W. Duin. Dissimilarity representations allow for building good classifiers. *PR Letters*, vol.23, no.8, pp.161-166, 2002.
4. K. Tieu and P. Viola., Boosting image retrieval. *International Journal of Computer Vision*, vol.56, no.1, pp.17-36, 2004.
5. K. Nandakumar A. K. Jain and S. Pankanti., Fingerprint based Fuzzy Vault: Implementation and Performance. *IEEE TIFS*, vol.2, no.4, pp.744-757, 2007.
6. Eskander, G.S., Sabourin, R. and Granger, E., Signature based Fuzzy Vaults with boosted feature selection. *SSCI-CIBIM 2011*, pp.131-138, Paris, 2011.
7. Rivard, D, Granger, E and Sabourin, R., Multi-Feature extraction and selection in writer-independent offline signature verification. *IJDAR*, vol.16, no.1, pp.83-103, 2013.
8. Eskander, G.S., Sabourin, R. and Granger, E., Adaptation of writer-independent systems for offline signature verification. *ICFHR-2012*, pp.432-437, Bari, Italy, 2012.
9. Eskander, G.S., Sabourin, R. and Granger, E., On the Dissimilarity Representation and Prototype Selection for Signature-Based Bio-Cryptographic Systems. *SIMBAD2013*, York, UK, 3-5 July 2013, accepted for publication.