An Iris-Based Interval-Mapping Scheme for Biometric Key Generation

Christian Rathgeb and Andreas Uhl
Department of Computer Sciences
University of Salzburg
5020 Salzburg, Austria
{crathgeb, uhl}@cosy.sbg.ac.at

Abstract

In order to increase security in common key management systems, the majority of so-called Biometric Cryptosystems aim at coupling cryptographic systems with biometric recognition systems. As a result these systems produce cryptographic keys, dependent on biometric information, denoted Biometric Keys.

In this work a generic approach to producing cryptographic keys out of biometric data is presented, by applying so-called Interval-Mapping techniques. The proposed scheme is adapted to iris, which has not yet been examined according to this technique, as well as on-line signatures to demonstrate that this approach is generic. Performance results show that the approach pays off.

1. Introduction

Focusing on practical cryptosystems any kind of crucial data is encrypted and decrypted applying algorithms well-known to be secure [14]. Regardless to applied algorithms, security depends on the secrecy of keys, revealing a security leakage when speaking of key management systems. Since cryptographic keys are hard to remember, these are stored on tokens (e.g.: smartcard) and released based on alternative PIN- or password-based authentication [15]. Exposing the weakest link of key management systems, (in)security of passwords is extended to security of encrypted data.

In order to enhance the reliability of any kind of key-release mechanism, biometrics are introduced, emerging a new area of research, namely *biometric cryptosystems*. Substantial security benefits are achieved, since it is significantly more difficult to forge, copy, share and distribute biometrics with as much ease as passwords and PINs [6]. Additionally, convenient consistency with respect to key management, an equal level of security (one account is no easier to break than any other), is provided, in contrast to user-selected password or PINs which may be chosen weakly.

Though different types of biometric cryptosystems[1] exist, most of these aim at generating cryptographic keys, dependent on biometric data, denoted *biometric keys*. Systems generating such keys are classified with respect to the coupling of biometric algorithms and cryptosystems [15]. Loose coupling corresponds to so-called *key-release schemes*, in which PIN- or password-based authentication is replaced by a biometric recognition algorithm. Within key-release schemes keys and biometric templates are stored separately, which does not conform to requirements of high security applications. *Key-generation* and *key-binding schemes* refer to generating/binding cryptographic keys from/with biometric data. Through tight coupling the secret key is bound to biometric information and stored templates do not reveal information, neither about the key, nor about biometric data.

Generic biometric recognition algorithms perform threshold-based matching where a certain degree of similarity between biometric measurements suffices to authenticate users. While biometric algorithms handle variances by setting appropriate thresholds, key generation schemes must overcome biometric variance in order to generate hundred percent correct keys. Over the past few years several approaches have been proposed in order to produce correct cryptographic keys out of noisy biometric data.

In this work a distinct group of biometric key generation systems is examined, namely so-called "interval-mapping schemes". In the concept of interval-mapping schemes adequate intervals are set up and extracted biometric features are mapped into these in order to create biometric keys or hashes. By introducing simple polygonal chains instead of Gaussian functions a generic and much simpler technique for constructing intervals which associate biometric features with previously chosen binary codewords is presented. Besides online signatures the technique is applied to iris biometrics for the first time, which does not represent an apparent area of application according to interval-mapping schemes.

This paper is organized as follows: first literature concerning so-called interval-mapping schemes is reviewed in detail (Sect. 2). Subsequently, the theoretical basis of the proposed scheme is examined (Sect. 3). Applied biometric algorithms are described (Sect. 4) and experimental results of the presented approach based on these algorithm are

[1] Different approaches exist, with different aims, falling under the category of biometric cryptosystems, however, all these approaches share the common purpose of securing cryptographic and biometric systems.

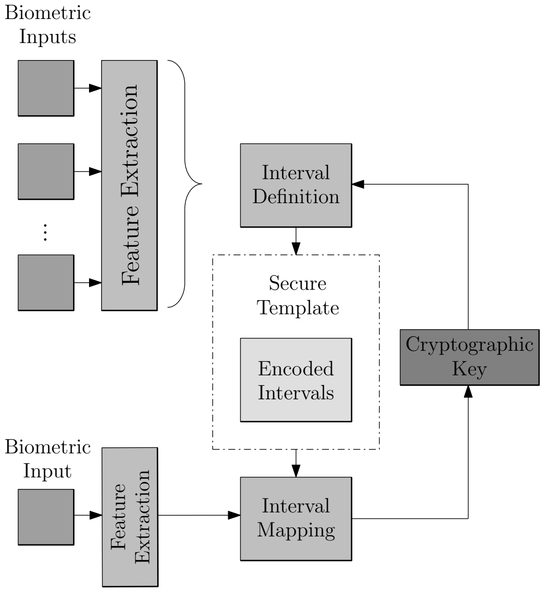

Figure 1. The basic operation mode of an Interval-Mapping scheme: features are extracted out of several input samples, intervals are defined and encoded. Through the according interval-mapping algorithm features are mapped into intervals and a key is returned.

shown (Sect. 5). Finally a conclusion is given (Sect. 6).

2. Related Work

Speaking of biometric cryptosystems a manageable amount of literature has been published. While some schemes aim at binding constant features with cryptographic keys [12, 5] others overcome biometric variance by means of error correcting codes [3, 8, 7]. Additionally, approaches have been made to increase entropy of existing keys [9] as well as securing biometric templates [11].

2.1. Interval-Mapping Schemes

One group of biometric key-generation systems aims at defining intervals for extracted biometric features, in order to deal with biometric variance. Intervals are encoded using one or several bits such that a concatenation of biometric features, mapped to intervals, produces a cryptographic key/hash (see Fig. 1). User-specific templates, which are stored in a database, consist of appropriate intervals including fake intervals to hide information.

Feng and Wah [4] proposed a scheme for hash generation from online signatures. At time of enrollment user boundaries are created for each extracted feature, defined as $(\overline{T}-b\cdot std_T, \overline{T}+b\cdot std_T)$, where $\overline{T}$ is the mean of ten feature values, std_T is the standard deviation of these values and b is a parameter to be adjusted. Subsequently a so-called database boundary, defined by the minimum and maximum value of all user boundaries, is divided into several intervals, each associated with an integer, forming one template, including all user-specific boundaries. During authentication features are fitted into intervals and feature codes are returned. The authors report a FRR of 28% and a FAR of 1.2% for the generation of 40-bit hashes. The applied test-set comprises a total number of 750 registered persons of a presumably self-acquired database.

Vielhauer *et al.* [16] proposed a hash generation algorithm based on online signatures as well. During enrollment a two-column interval matrix is generated for each user, where the i-th line consists of $\triangle I_i$ and Ω_i. The first value, $\triangle I_i$, is computed based on two intervals. The first interval, the initial interval, is defined as $[I_{InitLow}..I_{InitHigh}] = [MIN(n_{ij})..MAX(n_{ij})]$ where n_{ij} is the value of the i-th feature of the j-th sample, $j \leq 5$. The second interval, the extended interval, is defined as $[I_{Low}, ..., I_{High}] = [I_{InitLow} \cdot (1-t_i)..I_{InitHigh} \cdot (1+t_i)]$ where t_i is the tolerance factor. The second value of each column, Ω_i defines the interval offset for the i-th feature. The hash values are generated by interval mapping of every single feature against the interval matrix. Test results were evaluated on a test set of 10 subjects including skilled forgeries. As performance measurement a zero FAR and a FRR of 7.05% are reported.

Sutcu *et al.* [13] proposed another interval-based system for face biometrics. For each user several samples are processed and intervals are defined for each feature during enrollment. In order to encode intervals, single Gaussian functions are fitted to intervals, such that the evaluation of these functions reveals a hash code. Additionally, a number of fake Gaussian functions are added to hide information. The system was tested using the ORL database of faces which consists of 10 different images of 40 distinct subjects. As experimental results a FRR of 55.7% and a FAR of 3.3% are reported.

While behavioral biometric characteristics, such as online signatures, often produce feature vectors of real numbers, suitable to be used in interval-mapping schemes, most iris recognition algorithms are adopted to Daugman's approach [2] generating binary iris-codes where information lies within single bits. Apparently, interpreting bit streams of binary iris codes as decimal features does not serve a purpose of extracting distinct feature values since single bit errors may cause fatal discrepancies in the decimal value of a binary stream. The investigation of using binary feature vectors for interval-mapping schemes, for example, by applying Gray-Code to binary streams of iris codes is part of future work.

3. Construction of Interval-Mapping Scheme

The proposed scheme in based on a basic Interval-Mapping approach. Biometric recognition algorithms, which provide a n-dimensional feature vector $(f_1, f_2, ..., f_n)$ are used to set up intervals, which are encoded through an adequate number of bits, in order to generate cryptographic keys, long enough to be applied in cryptographic algorithms. In the following subchapters the theoretical basis of enrollment as well as authentication is described in detail.

3.1 Enrollment

During enrollment firstly feature vectors of user k, denoted by $(f_1^k, f_2^k, ..., f_n^k)$, are aligned, dependent on the range of all features processed during enrollment such that,

$$\begin{aligned} \hat{f_i^k} &= l \cdot \frac{f_i^k - f_{i-}}{f_{i+} - f_{i-}}\,, \qquad (1) \\ f_{i-} := \min(f_i^k)\ \forall k\,, &\qquad f_{i+} := \max(f_i^k)\ \forall k. \end{aligned}$$

where $l \in \mathbb{N}$ is a predefined parameter. Thereby features are mapped into the interval $[0, l]$. Subsequently, the mean value, $\overline{f_i^k}$, and standard deviation, $(\delta_i^k)^2$, of each aligned feature i of user k are calculated where,

$$\overline{f_i^k} = \frac{1}{n}\sum_{j=1}^{n} \hat{f_{ij}^k}\,, (\delta_i^k)^2 = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(\hat{f_{ij}^k} - \overline{f_i^k})^2} \qquad (2)$$

and n denotes the number of enrollment samples. The interval of the i-th feature of user k, I_i^k, resulting out of feature values $f_{ij}, j = 1...n$, is defined as,

$$\begin{aligned} I_i^k &= [b_L, b_R]\,, \qquad (3) \\ b_L, b_R &= \overline{f_i^k} \pm (\delta_i^k)^2 \cdot d \bmod l \end{aligned}$$

where b_L denotes the left border, b_R denotes the right border of the interval and d is a parameter to be adjusted. Having calculated the interval of features these have to be encoded to form parts of a cryptographic key. Random binary codewords $c_i \in \{0, 1\}^m$ of length m are assigned to each interval I_i, where the interval mapping is based on the idea presented in [13] where adequate Gaussian functions are used to map extracted biometric features to decimal codewords which are combined to generate a cryptographic key. Additionally, fake Gaussian functions are added in order to hide genuine intervals. Instead of Gaussian functions polygonal chains are used to encode intervals, which is easier serving the same purpose. In order to hide genuine intervals fake intervals are added, as illustrated in Fig. 2:

1. Rays are shot through (b_L, c_n) and (b_R, c_n) starting at $(\overline{f_i}, c_{n+1})$, where c_n and c_{n+1} denote binary codewords of decimal values n and $n+1$ respectively. Random points are chosen on these rays such that the y-values of these points are less than n.

2. Rays are launched at these random points such that the angle of entry is equal to the angle of reflection. Random points are chosen on these rays and the algorithm is applied recursively until enough peak points are calculated.

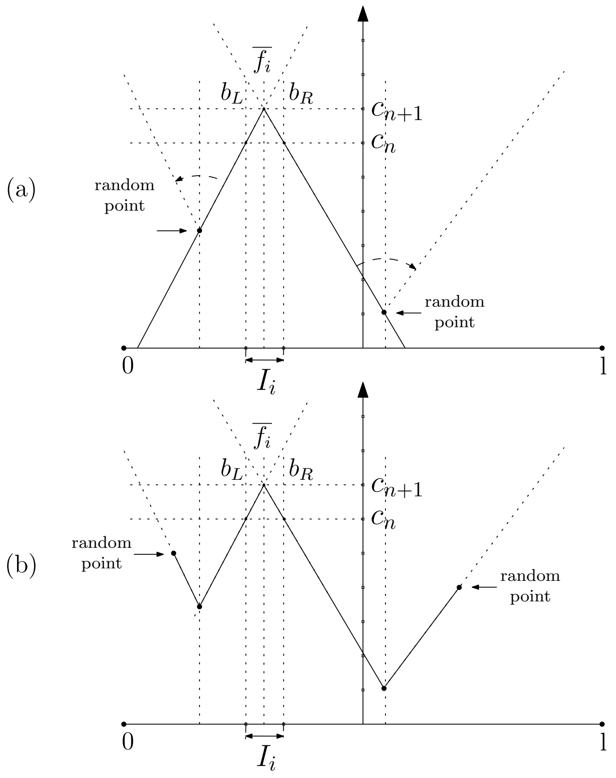

Figure 2. The proposed scheme: (a) Feature intervals are used to generate polygonal chains which are used to map features to codewords. (b) fake intervals (peaks) are added to hide information.

The enrollment procedure produces several arrays of points $\{P_{i1}, P_{i2}, ..., P_{iN}\}$, $P_{ij} \in \mathbb{R}^2$, defining polygonal chains for the i-th feature of a feature vector. Each array consists of $N/2$ peak points, defining one genuine interval as well as $N/2 - 1$ fake intervals, overlapping with predefined feature space $[0, l]$. For a feature vector of length s a set of s point arrays defines the a user's template.

Distinct features create tight interval boundaries and sharp peaks while unsteady features stretch intervals and reveal peaks which may even lie outside the entire interval $[0, l]$. Thus, codewords of distinct feature intervals are hidden better than those of unsteady features, which means distinct feature intervals bring about more fake intervals lying within $[0, l]$.

3.2 Authentication

At the time of authentication features of one biometric sample are extracted, which have to be mapped to intervals. The i-th feature value, f_i, of the extracted feature vector is first mapped to its aligned correspondence $\hat{f}_i$ (just like in the enrollment phase). Aligned features are then mapped onto the according polygonal chain, created during enrollment, in order to generate a codeword. The process of mapping feature values onto a polygonal chain is described in detail subsequently. The concatenation of all codewords forms the

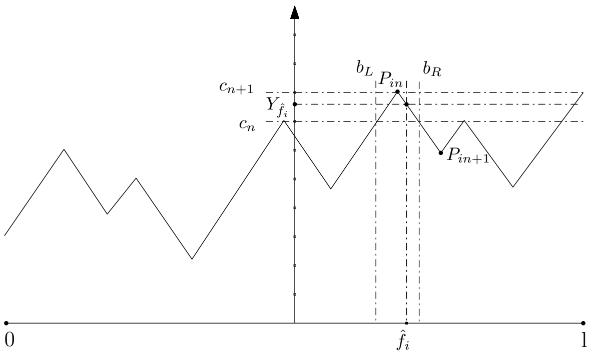

Figure 3. The y-value for a feature value f_i is calculated by interpolating two points. The lower absolute value of the result is the codeword for the feature f_i. To avoid that $\lfloor(Y_{\hat{f}_i})\rfloor = c_{n+1}$ a small δ should be subtracted from the peak.

cryptographic key.

Let $\{P_{i1}, P_{i2}, ..., P_{iN}\}$ be the point array, ordered with respect to x-values, which was stored in a user's template to form the intervals for the ith feature of an extracted feature vector. Since a perpendicular line to the x-axis intersects with a polygonal chain at most once (polygonal chains are monoton with respect to the x-axis), at most two points P_{in} and P_{in+1} exist, such that $x_{P_{in}} \leq \hat{f}_i \leq x_{P_{in+1}}$ where $x_{P_{in}}$ is the x-value of P_{in} and $x_{P_{in+1}}$ is the x-value of P_{in+1}. If such two points can be found, $\hat{f}_i$ is mapped onto the polygonal chain. The corresponding y-value of $\hat{f}_i$ is calculated by means of interpolation such that,

$$Y_{\hat{f}_i} = Y_{P_{in}} + (\hat{f}_i - X_{P_{in}}) \cdot Y_{\overrightarrow{P_{in}P_{in+1}}} \tag{4}$$

where $Y_{P_{in}}$ is the y-value of the point P_{in}, $X_{P_{in}}$ is the x-value of the point P_{in} and $\overrightarrow{P_{in}P_{in+1}}$ is the vector between the points P_{in} and P_{in+1} (see Fig 3). After the y-value of each feature value $\hat{f}_i$ of the feature vector is calculated, codewords $C_{\hat{f}_i}$ are returned for each feature value to generate the cryptographic key where,

$$C_{\hat{f}_i} = \lfloor(Y_{\hat{f}_i})\rfloor. \tag{5}$$

By now, the importance of fake intervals becomes clear, since imposters would be able to identify genuine intervals yielding to genuine cryptographic keys if only a single peak, consisting of two line segments, is stored in the template to create a codeword. In the context of biometric cryptosystems fake intervals are comparable to so-called chaff-points in the "fuzzy vault" approach [7].

4 Applied Biometric Algorithms

Unlike iris recognition algorithms based on iris codes the algorithm of Zhu *et al.* [18] fulfills the requirements of generating feature vectors consisting of real numbers. By applying a 2D wavelet transform a total number of 26 floats is extracted for one input sample. As result of the preprocessing step normalized iris textures are generated similar to Daugman's [2] approach. In the feature extraction stage wavelet transform is applied on original images, subsequently a set of sub-bands is obtained at different resolution levels. Means and standard deviations of each wavelet sub-band are extracted as texture features, where DAUB4 is used as basis for the wavelet transform.

In order to demonstrate that the presented technique is generic a second, more apparent, biometric modality is applied to the interval-mapping scheme. An online-signature recognition algorithm was constructed, extracting 44 features out of six signals, including x- and y-position of the pen, pen pressure, pen up/down signal, azimuth and altitude[2]. The feature vector contains magnitudes similar to those presented by Vielhauer *et al.* [16], such as duration of signature, number of strokes, average x- and y- position as well magnitudes measured in four intervals such as average pressure and writing speed.

The performance of both algorithms is measured applying one metric, namely the weighted Euclidean distance. A user is identified as person k if the following weighted Euclidean distance is below a predefined threshold:

$$WED(k) = \sum_{i=1}^{N} \frac{(f_i - \overline{f_i^k})^2}{(\delta_i^k)^2} \tag{6}$$

where f_i denotes the ith feature of an unknown person and N is the total number of features extracted from one input sample.

Since the algorithm of Zhu *et al.* [18] does not handle presence of eyelids and eyelashes only half of the iris image, from the right side [45^o to 315^o] and the left side [135^o to 225^o] are used to get rid of most eyelids and eyelashes. The iris recognition algorithm was tested on the CASIA-IrisV3-Interval database [1], where all persons for which at least 10 images were available were tested. For a total number of 41 persons the ROC and EER are plotted in Fig. 4. The FRR and FAR are given in in Tab. 1. At a FAR of 0.0% the iris recognition algorithm reveals a FRR of 35.21%. According to iris recognition this performance is obviously unsatisfying, however, the system fulfills the precondition of extracting a sufficient number of real-valued features.

For online signature recognition user-specific thresholds make sense [10], since psychological features underlie user-specific variations. Thus thresholds for each user are set up by observing the intra-class distributions. The online signature recognition algorithm was tested on the SVC04 database [17]. The database comprises a total number of 40 persons, where for each person 20 genuine signatures as well as 20 skilled forgeries are available. The ROC and EER are plotted in Fig. 5. As shown in Tab. 1 as experimental results for a FAR of 0.0% a FRR of 18.17% is achieved. By using multiple signature acquisition during authentication this algorithm would provide a practical performance.

[2] Each of these signals is measured in equidistant timestamps, typical for online signature sensors.

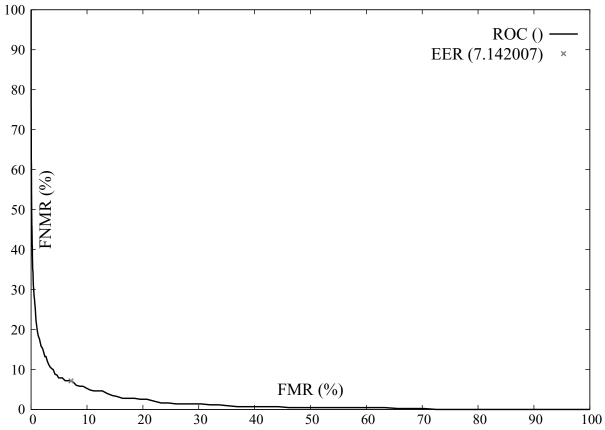

Figure 4. ROC and EER for iris recognition.

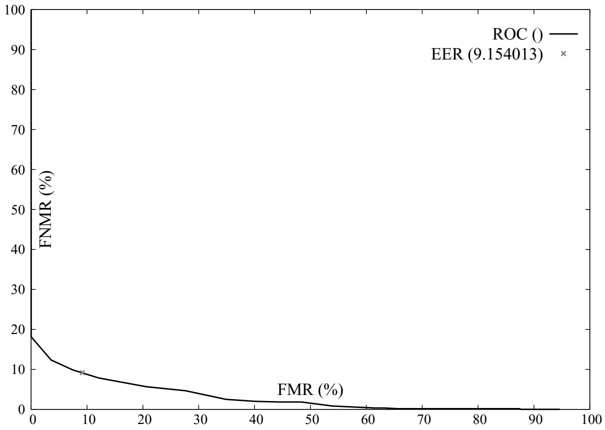

Figure 5. ROC and EER for sig. recognition.

5 Experimental Results

The proposed scheme is adapted to both recognition algorithms which means an adequate size for the extracted codewords is chosen in order to generate sufficiently long cryptographic keys. Tests are carried out using the same databases [1, 17]. Several approaches have been published applying Interval-mapping schemes to online-signatures [4, 16, 13], however, iris recognition has not yet been investigated to be applied in such schemes.

Performance is measured by calculating the Hamming-Distance between generated keys during authentication and correct cryptographic keys, used to encoded the intervals during enrollment. This means a successful authentication requires a hundred percent correct key, according to this, a zero Hamming Distance. The proposed system may produce keys which do not differ much from the correct key (for genuine as well as non-genuine users). However, such keys could be deemed by the system applying hash functions to these and test the results against hash values of the correct keys, stored during enrollment.

5.1 Iris Key Generation

Since the implementation of the algorithm of Zhu *et al.* [18] generates a feature vector of length $N = 26$, each one of these features is encoded using 5-bit codewords in order to produce a cryptographic key of length 130-bit. A total number of five input samples is used during enrollment where genuine as well as fake intervals are constructed. The intra- and inter-class distributions are plotted in Fig. 6. The system reveals a key generation rate of almost 63.5%. This means for a 63.5% of genuine users correct keys are generated, corresponding to a FRR of 36.5% and a negligible FAR of 0.07% shown in Tab. 1. Thus the performance of the proposed system is only slightly worse than that of the original system (∼1.3%). Since the iris has not yet been investigated with respect to key generation using interval-mapping schemes, these results appear encouraging.

Table 1. Experimental results of recognition algorithms and key generation schemes.

	FRR (%)	FAR (%)
Algorithm of Zhu *et al.*	35.21	0.0
Sig. Rec. Algorithm	18.17	0.0
Iris Key Generation	36.5	0.07
Sig. Key Generation	24.84	0.02

5.2 Online Signature Key Generation

The online signature recognition algorithm extracts a total number of $N = 44$ float values for each user. Thus, each feature is encoded by 3-bit codewords resulting in an 132-bit cryptographic key, long enough to be used in general cryptographic systems. Again, five input samples are used in the enrollment procedure to set up intervals. The intra- and inter-class distributions are plotted in Fig. 7, the FRR and the FAR are summarized in Tab. 1. Here the FRR increases more drastically (∼6.7%) revealing a key generation rate of 75.16%.

Considering the performance of the applied biometric recognition algorithms both interval-mapping schemes show satisfying results. In order to avoid returning cryptographic keys, which may comprise bit errors, to genuine users, keys could be tested against a previously stored hash value of these. Compared to existing approaches [4, 16] the online signature interval-mapping scheme shows inferior performance at first glance. However, in our approach cryptographic keys of 132-bit are generated. In contrast to biometric key generation systems based on online signatures, which restrict to extract short hashes [4, 16], keys of 132-bit are suitable to be used in any standard cryptographic algorithm. As mentioned earlier, iris biometrics have not been investigated according to biometric cryptosystems based on interval-mapping schemes.

5.3 Security Analysis

The security of the whole systems depends on distinct features forming tight boundaries and generating sharp peaks in polygonal chains. Since tight intervals imply

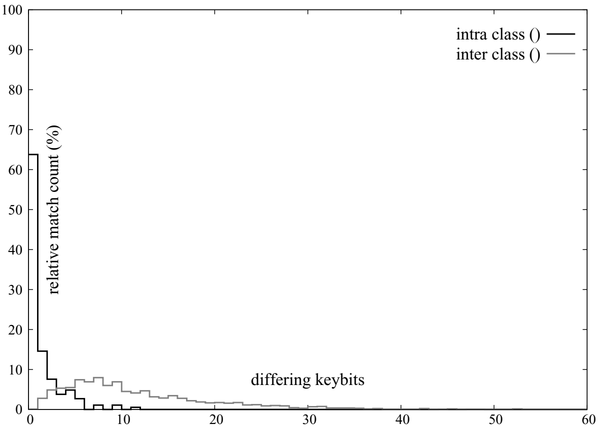

Figure 6. Iris key generation.

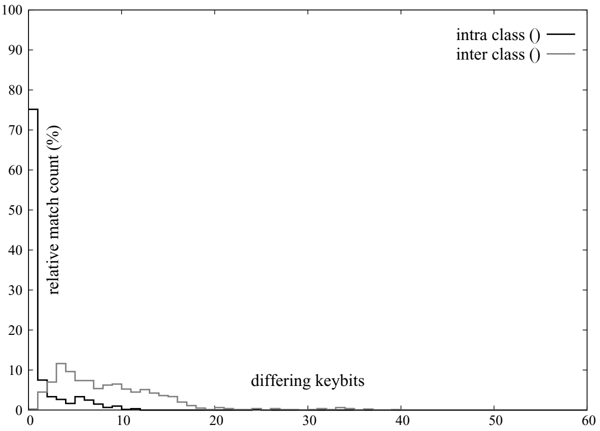

Figure 7. Signature key generation.

the generation of several fake points forming fake intervals, simple polygonal chains do not reveal any information about genuine intervals out of which correct keys are constructed. In contrast, unsteady features stretch intervals such that resulting polygonal chains hide only little information, with respect to the codeword which is used to encode these. Imposters may take advantage of unsteady features in order to construct parts of the correct key straight forward.

6 Conclusion

In this work a generic approach to biometric key generation is presented, which is adapted to an iris recognition system as well as an on-line signature recognition system. Common iris recognition algorithms are not adequate to be used in interval-mapping schemes. Thus, the investigation of adopting an interval-mapping scheme to iris biometrics reveals valuable results. Altough the false alarm of the proposed systems is at a reasonable level both systems fulfil the requirement nearly maintaining the performance of applied biometric recognition algorithms.

Experimental results demonstrate that the proposed approach can be adopted to psychological as well as physiological biometric characteristics. Furthermore, both key generation systems produce cryptographic keys which are sufficiently long to be applied in general cryptosystems.

References

[1] The Center of Biometrics and Security Research, CASIA Iris Image Database, http://www.sinobiometrics.com.

[2] J. Daugman, ”How Iris Recognition Works”, *IEEE Trans. CSVT*, 14(1), 2004, pp. 21–30.

[3] G. Davida, Y. Frankel, and B. Matt, ”On enabling secure applications through off-line biometric identification”, *Proc. of IEEE, Symp. on Security and Privacy*, 1998, pp. 148–157.

[4] H. Feng and C. C. Wah, ”Private key generation from on-line handwritten signatures”, *Information Management and Computer Security*, 10(18), 2002, pp. 159–164.

[5] A. Goh and D. C. L. Ngo, ”Computation of cryptographic keys from face biometrics”, In *Communications and Multimedia Security (LNCS: 2828)*, pp. 1–13, 2003.

[6] A. K. Jain, A. Ross, and S. Prabhakar, ”An introduction to biometric recognition”, *IEEE Trans. on Circuits and Systems for Video Technology*, 14, 2004, pp. 4–20.

[7] A. Juels and M. Sudan, ”A fuzzy vault scheme”, *Designs, Codes and Cryptography*, 38, 2006, pp. 237–257.

[8] A. Juels and M. Wattenberg, ”A fuzzy commitment scheme”, *Sixth ACM Conference on Computer and Communications Security*, 1999, pp. 28–36.

[9] F. Monrose, M. K. Reiter, and S. Wetzel, ”Password hardening based on keystroke dynamics”, *Proceedings of sixth ACM Conference on Computer and Communications Security, CCCS*, 1999, pp. 73–82.

[10] V. Nalwa, ”Automatic on-line signature verification”, *Proceedings of the IEEE*, 85, 1997, pp. 215–239.

[11] N. K. Ratha, J. H. Connell, and R. M. Bolle, ”Enhancing security and privacy in biometrics-based authentication systems”, *IBM Systems Journal*, 40, 2001, pp. 614–634.

[12] C. Soutar, D. Roberge, A. Stoianov, R. Gilroy, and B. V. Kumar, ”Biometric Encryption using image processing”, *Proc. SPIE, Optical Security and Counterfeit Deterrence Techniques II*, 3314, 1998, pp. 178–188.

[13] Y. Sutcu, H. T. Sencar, and N. Memon, ”A secure biometric authentication scheme based on robust hashing”, *MMSec ’05: Proceedings of the 7th Workshop on Multimedia and Security*, 2005, pp. 111–116.

[14] W. Trappe and L. C. Washington, *Introduction to Cryptography with Coding Theory: 2nd Edition*, Pearson Prentice Hall, 2006.

[15] U. Uludag, S. Pankanti, S. Prabhakar, and A. K. Jain, ”Biometric cryptosystems: issues and challenges”, *Proceedings of the IEEE*, 92(6), 2004, pp. 948–960.

[16] C. Vielhauer, R. Steinmetz, and A. Mayerhöfer, ”Biometric hash based on statistical features of online signatures”, In *ICPR ’02: Proceedings of the 16 th International Conference on Pattern Recognition (ICPR’02) Volume 1*, pp. 10123, Washington, DC, USA, 2002. IEEE Computer Society.

[17] D.-Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, and G. Rigoll, ”SVC2004: First International Signature Verification Competition. Proceedings of the International Conference on Biometric Authentication (ICBA)”, *LNCS*, 3072, 2004, pp. 16–22.

[18] Y. Zhu, T. Tan, and Y. Wang, ”Biometric personal identification based on iris patterns”, *15th International Conference on Pattern Recognition (ICPR’00) - Volume 2*, 2000, pp. 801–804.