Journal of Information Hiding and Multimedia Signal Processing ©2018 ISSN 2073-4212
Ubiquitous International Volume **9**, Number **4**, July **2018**

Karhunen Loeve Transform-based Online Handwritten Signature Scheme via Dynamic Bit Allocation Method and Naive Bayes for Biometric Key Generation

Lei Meng[1] and Lin Teng[1*]

[1]Software College
Shenyang Normal University
Shenyang, 110034 - China
*Corresponding author: ysl352720214@163.com;

Received July, 2017; revised April, 2018

Abstract. *Biological characteristics and information security have highly complementary attributes, which is a hot topic in the field of information security research nowadays. Some traditional keys are easily suffered from dictionary attack. To mitigate this demerits and obtain reliability coding of biological key, we propose an online handwritten signature scheme based on Karhunen loeve transform via dynamic bit allocation method and naive Bayes for biometric key generation in this paper. The new scheme is partitioned into three steps. Firstly, feature selection of users is executed by karhunen loeve transform. A feature that mostly can represent the characteristics of the user is selected. Additionally, Naive Bayes in dynamic bit allocation method is adopted for determining probability quality of registered samples characteristics quantization. Thirdly, the trapezoid size replaces the definite integral to improve the computation time. Experimental results show that our new scheme greatly improves the effectiveness of handwritten signature and has powerful competitive advantage than other methods.*
Keywords: Biological characteristics, Online handwritten signature, Karhunen loeve transform, Dynamic bit allocation method, Naive Bayes

1. **Introduction.** With the development of computer network, information security has become a key technical problem in the network environment. Biological characteristics refer to automatically match the physiological or behavioral characteristics of people, it is the only way to fully identify a person. Biological characteristics and information security are highly complementary properties, which is a hot-spot in the field of information security research. The key that is generated based on user-specific biometric information is not only secure but also difficult to decode.

In order to solve the problem that the user's digital identity is not accordant with the physical identity, based on biological characteristics, the researchers use a various of methods to carry out the research on certification. Because biology has the characteristics of being portable, without memory, not easily to lost, small repetition probability, difficult to copy and so on. Biometrics-based certification has gradually become the mainstream developmental direction of cryptography, and it is expected to provide a more natural and reliable solution.

When biometrics are combined with cryptography, biological coding will be generated, and it consist of three types, known as key release [1], key binding [2] and key generation [3-5]. Both the key release and key binding adopt the combination of biometrics and external input key. The inaccuracy of combination may cause the result that external input key occupies dominating status while the safety of the system still relies on the key, instead of biometrics. The generation of the secret key by the key generation scheme is based on biometrics discretization, thus it is difficult to generate the same biological password for impostor. This paper will mainly focus on the study of the key generation. The key in computer cryptography must be determined, and the biometrics are composed of some continuous data. Therefore, the directly expression of binary must depend on the high effective discretization scheme.

The bio-discretization scheme includes two stages: qualification and coding. The static distribution method is an early discretization method, which maps the feature of each dimension into the code string with fixed length. And the secret key used for identity certification is generated by the combination of code string in dimensions. Wu et al., [6] proposed the first public key encryption with keyword search scheme with a designated server based on ID-based systems that possessed the advantage (removing certificate management) of ID-based systems. The Krishna et al., [7] put forward another static multi bit discretization scheme based on likelihood ratio aiming to the features of face and finger. It could determine the function interval of likelihood ratio based on probability mass and segment the left and right section. The original feature will be mapped into binary gray code. According to the positive and negative of the third layer wavelet coefficients of the iris, Rathgeb et al., [8] proposed a single bit code. Wu et al., [9] presented a new user authentication and key exchange protocol using bilinear pairings for mobile clientCserver environment. However, the distinguish ability in each dimension is different, so does the volume of information. Thus to some extent, using a fixed strength to code string can blur the useful information. The equal interval quantization method is improved [10-12]. And the length of the feature code is related to its own characteristic distribution. The larger the feature separability is, the longer the assigned code string is, However, once the users boundary and the global feature boundary are known for attacker, the information will be leaked. According to the facial characteristics, the reliability wight can distribute the length of code string in sequence, adjust the weight threshold until it can fulfill the length of code string. Even though the reliable features can get distribute more code string, a approximately comprehensive qualification scheme are not very ideal for the performance of certification.

The collection of handwritten signature data has the characteristics with non-intrusion, and there are no problems of user's feelings exclusion or privacy infringement when collecting fingerprints, face, iris. From the perspective of sports biomechanics and physiology, Yoshimura et al., [13] suggests that handwritten signatures is a fast and skilled "ballistic motion" that reflects the individual differences in writing habits and handwriting signatures. So that it can be used for identity identification. The characteristics of handwritten signature were evaluated from characteristic deviation, characteristic soil entropy, characteristic stability and the correlation of entropy [14]. And also other researchers had represented some new approaches [15-22].

In order to realize the feature reliability encoding, this paper proposes online handwritten signature scheme based on karhunen loeve transform (KLT) via dynamic bit allocation method and Naive Bayes for biometric key generation. In the new scheme, distribution of the same characteristics of different users may be different. Different characteristics distribution provides guarantee for real dynamic bit distribution and it can generate stability, longer code string. In addition, Naive Bayes method can ensure that the attacker

cannot completely crack the feature string under the situation of characteristic boundary leakage.

The remainder of this paper is organized as follows. In Section 2, feature selection based on Karhunen Loeve and Transform feature quantization are analyzed first, then the overhead of feature coding for our new scheme is assessed, then detailed algorithm of new scheme is presented. In Section 3, extensive experiments are conducted using our method to evaluate the performance. The conclusions are provided in Section 4.

2. Dynamic bit allocation method based on Naive Bayes.

2.1. **Feature selection based on Karhunen Loeve Transform (KLT).** There are many inseparable features in the process of biometrics extraction, and the inseparable features of each user cannot be identical. This paper presents a feature selection method based on user-dependent feature selection method. KLT [23,24] is a commonly feature selection method. Its aim is to look for the main component of any statistical distribution of data collection. The corresponding basis vector satisfies orthogonality and by which defines the optimal subspace. It transforms the original data to principal component space to make the cross correlation of single data sample reduce to the lowest point.

$$x = [u_1, u_2, \cdots, u_d] = Uy. \tag{1}$$

Where x and y represent the vectors on the registration sample set respectively. It shows that the higher the score of a dimension is, the better class separability is. The former d dimensions features are selected as the subset of features to achieve the purpose of removing noise and weakly separable features.

By using the KLT that the user depends on to make a feature selection, it can obtain I n-dimension vectors. Each vector may be different, but for the corresponding feature can be selected correctly, the registration process needs to be selected as a secondary data storage. When the registered samples of online handwritten signature is enough, according to the central limit theorem, it can be assumed that the feature distribution follows the Gaussian distribution. In this paper, the registration feature interval controlling coefficient is used to determine the registration feature interval of registered sample, and the biological characteristics with different statistical parameters provide protection for dynamic allocation.

2.2. **Feature Quantization and coding.** As we all know, naive Bayes is a simple probabilistic method based on applying Bayes theorem with naive independence assumptions between the features [25,26]. For some probability models, naive Bayes can be trained very efficiently by using maximum likelihood in a supervised learning setting. Abstractly, naive Bayes is a conditional probability model. Given a vector $v = (v_1, v_2, \cdots, v_n)$ represents n features. Probabilities of each feature f_k are $p(f_k|z_1, z_2, \cdots, z_n)$. According to Bayes theorem, the conditional probability can be decomposed as:

$$p(f_k|z) = \frac{p(f_k)p(z|f_k)}{p(z)}. \tag{2}$$

Assume that each feature z_i is conditionally independent of each other feature z_j for $j \neq i$, given the feature C. This means that,

$$p(z_i|z_{i+1}, \cdots, z_n, f_k) = p(z_i|f_k). \tag{3}$$

Under the above independence assumption, the conditional distribution can be expressed as:

$$p(f_k|z_1, \cdots, z_n) = \frac{1}{Z} p(f_k) \prod_{i=1}^{n} p(z_i|f_k). \tag{4}$$

Where the evidence $Z = p(x)$ is a scaling factor dependent only on $(z_1, \cdots, z_n)$, that is, a constant if the values of the feature variables are known.

Although the above mentioned quantization based on naive Bayes can better resist that attacker obtain advantage from guessing discretization output, the time performance of registration is completely lost on the interval-based quantization. This is mainly due to the fact that the probability quality calculation needs to integrate the background probability density function, and the curve integral operation is a time consuming operation.

The curve integral [27] is the area size below the curve. In the quantization segment, the trapezoidal area below the segmentation curve of the background probability density function is used to approximate the curve integral to achieve the purpose of reducing the computational time and improve the efficiency of the operation.

The number of segments of the background probability density function determines the number of bits of the feature code. When the function is divided into S segments, at least $\lceil 1z(S) \rceil$ is needed to satisfy the coding requirement. Directly binary coding is easy to cause large within-class distance in the Hamming domain. However, Gray code is a coding method that is minimized by error coding. In the Hamming domain, the single-step feature of Gray code can reduce the quantization error caused by within-class change and reduce the false rejection rate (FRR). In this paper, we use Gray code to implement feature coding. The n-dimensional feature selected according to the KLT is quantized and the n sub-codes are sequentially concatenated into a code string, which is a biological key.

The feature quantization interval and the biological key are required to be stored as auxiliary data. When the user's identity is authenticated, the user authentication key is generated by judging in which interval of the biometric feature and matching it with the stored key template. If it is matched, they are the real users, otherwise they are forged users.

2.3. **Detailed algorithm processes.** The D-dimensional feature is extracted from the signature data, and the former n-dimensional feature is selected as the input of the bit allocation method according to the KLT that the user depends on. In the user registration phase, the specific algorithm of probability-based dynamic bit allocation method in this paper is as follows:

- Step 1. Calculate the expectation and variance of the D-dimensional features of all user registration samples in the database. And calculate the expect and variance of its d-dimensional feature according to the registered sample of user i and construct the original probability density function according to the statistical characteristics.
- Step 2. Read the expectation, variance from step 1 corresponding to d-dimensional feature of the constructed background probability density function.
- Step 3. Use the feature interval control coefficient a to determine the width of the background probability density function of the original probability density function and calculate the quantization probability quality $p(a)$ according to the width.
- Step 4. Quantify the segmentation of the background probability density function according to $p(a)$ and perform the gray code on the user's characteristic according to the quantization segmentation result, which is the d-dimension code string of the user i.

- Step 5. The n-dimensional features of user i are executed from step 2 to step 4, and the n signature strings are cascaded as the online signature biological key of user i.

3. **Experimental results and performance analysis.** In order to verify the performance of the dynamic bit allocation method based on the Naive Bayes, this paper uses the part of public MCYT Online and Offline Signature Database available in ($http : //www.gavo.t.u - tokyo.ac.jp/\ qiao/database.html$) as experimental data set. The public part contains the English signatures and Chinese signatures, a total of 40 users' signature data, each user contains 20 real signatures, 20 forged signatures. 20 real signatures are acquired within two times. Each time collects 10 data. Time period is two weeks.

As the online handwritten signature is a biological behavioral characteristics, in terms of stable, it is worse than fingerprints, iris and other biological physiological characteristics. In order to improve the stability of online handwritten signatures, the experimental scheme is designed by randomly selecting five samples as the registered samples from the real signatures. The remaining 10 real samples and 20 forged samples are the test samples in authentication phase. According to the above description, it repeats 10 times and takes the average value as the system's authentication performance. Real samples and forged samples are shown in Figure 1.

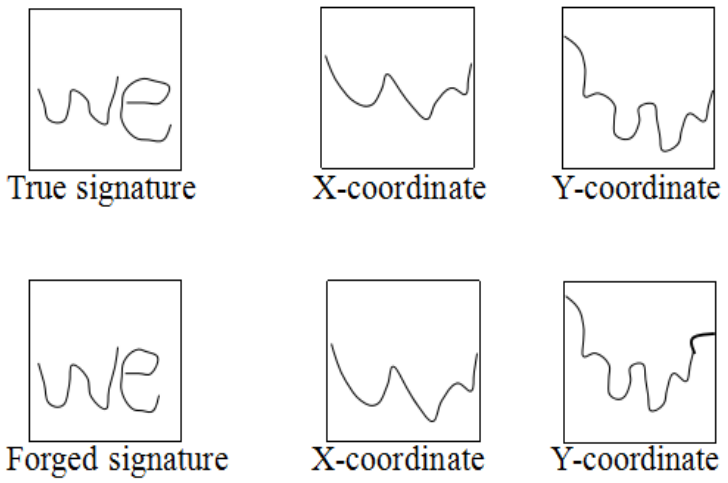

FIGURE 1. True signature and forged signature.

The performance evaluation of handwritten signature authentication is mainly based on FAR and FRR indicators. Feature interval control coefficient a needs to make a compromise. When a is larger, the registered user feature interval is also large. It is easy to authenticate the forged signature as a real signature, that is, it obtains a larger FAR. On the other hand, when a is smaller, the registered user characteristic range is also smaller, and the authentication is easy to authenticate the real signature as a forged signature. Namely, it gets a bigger FRR.

Under the constraint of given FAR, by solving the optimization problem of minimum FRR, we can get different a with different dimensions. In order to simplify the problem, this paper uses the same a for registered. Without introducing code string fuzzy parameters, the FRR and FAR change with a as shown in Figure 2.

According to the curve, the error rate is 10%, and the error rate result given in the reference [28] is 10.90%. The best test result in the MCYT signature database is 6.90%.

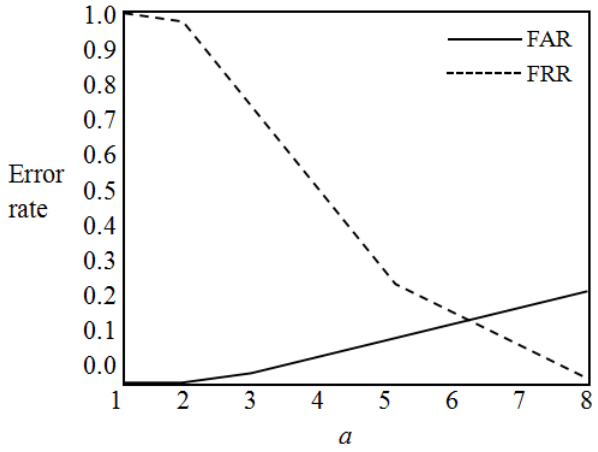

FIGURE 2. Error rate curve with a.

The comparison shows that although the method proposed in this paper cannot acquire the best certification, the test results also illustrates the validity of the method based on the online handwritten signature. FAR represents the ratio of the legal secret key that can be generated from the forger. The lower the FAR is, the more reliable the user is. Then we choose a=3, FRR is 29.63%, FAR is 2.64%. If the code string length fuzzy ratio is 99%, then FRR is 15.99 and FAR is 6.28% respectively.

Different key generation methods are compared in this paper with CADB [29], BCVD [30] , KGCI [31] and our method. Table 1 is authentication result based on different biometric key generation methods. From the table it can be seen that, the results of identity authentication based on the behavior feature based on physiological characteristics are different. The former method has poor certification performance, which also reflects the within-class distance of behavioral characteristics longer than that of biometrics. The stability of behavioral characteristics is weaker than the physiology characteristic features. However, although the acquired FRR is much higher than corresponding physiological characteristics FRR, users only need average 1.4 times to sign and overcome the error caused by the distance within the class error, which is acceptable for most users. Random forged the number critically affects the system of FAR. However, the method presented in this paper still has a lower FAR than the early behavior-based feature key generation method without random forgery. In general, our new method has good authentication performance. In order to verify the effectiveness of the improved method in terms of reducing the computational complexity, table1 records the time consumption of user signature registration when a=3. It can be seen that signature registration time of our proposed method is 2.68s.

TABLE 1. Results of key generation method based on biological characteristic.

Scheme	FRR%	FAR%	Code length	Time consumptions
CADB[29]	12.58	10.32	236	7.46
BCVD[30]	19.78	5.67	128	5.21
KGCI[31]	25.45	3.78	128	4.53
Our method	27.59	2.49	123	2.68

Guessing entropy is an important concept in the security system. It describes the expected guessing number of trusted key obtained by attacker. When the key's length is S bit, the number of exhaustive guessing expectation is about $2S$, the relationship between

S and guessing entropy is index relation. When S is larger, the system is more secure. In this paper, a longer secret key is obtained without introducing a random variable, and the average length of the key is 123 bits, which is sufficient to resist the exhaustive search attack. In addition, the quantization of equal probability quality can better resist the attacker from guessing discretization output to obtain the key information.

4. **Conclusions.** In this paper, a new online handwritten signature scheme based on Karhunen Loeve Transform (KLT) via dynamic bit allocation method and Naive Bayes for biometric key generation is proposed. K-L criterion is used to select the feature. Based on the likelihood ratio quantization method of the dynamic bit allocation method, the feature quality of the registered sample feature is determined by the feature interval control coefficient. Time performance of the registration phase is greatly improved with our method. In the future, we will study more advanced Biological key generation methods to improve this paper's method.

Acknowledgment. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] L. Medrihan , E. Ferrea , B. Greco, et al. Asynchronous GABA Release Is a Key Determinant of Tonic Inhibition and Controls Neuronal Excitability: A Study in the Synapsin, II-/- Mouse[J]. 2015, 25(10):3356-3368.

[2] Y. J. Lee, K. Bae , S. J. Lee , et al. Biometric Key Binding: Fuzzy Vault Based on Iris Images, *Journal of Advances in Biometrics*, vol. 4642, pp. 800-808, 2007.

[3] J. Delvaux , D. Gu , D. Schellekens , et al. Helper Data Algorithms for PUF-Based Key Generation: Overview and Analysis, *Journal of IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 34m, no. 6, pp. 889-902, 2015.

[4] K. Zeng, Physical layer key generation in wireless networks: challenges and opportunities, *Journal of Communications Magazine IEEE*, vol. 53, no. 6, pp. 33-39, 2015.

[5] R. Maes , V. V. D. Leest , E. V. D. Sluis, et al. Secure key generation from biased PUFs: extended version, *Journal of Journal of Cryptographic Engineering*, vol. 6, no. 2, pp. 121-137, 2016.

[6] T.Y. Wu, T.T. Tsai, Y.M. Tseng, ?Efficient searchable ID-based encryption with a designated server?, *Annals of telecommunications*, vol. 69, no (7-8), pp. 391-402, 2014.

[7] M. He, Z. C. Wu, F. Li, Key Generation Method Based on On-line Handwritten Signature, *Journal of Computer Engineering*, vol. 42, no. 10, pp. 164-168, 2016.

[8] Rathgeb C, Uhl A. Context-based biometric key generation for Iris, *Journal of Computer Vision Iet*, vol. 5, no. 6, pp. 389-397, 2011.

[9] T.Y. Wu, Y.M. Tseng, An efficient user authentication and key exchange protocol for mobile client-server environments, *Computer Networks*, vol. 54, no. 9, pp. 1520- 1530, June 17 2010.

[10] S. Kadono, Moving picture coding method for quantizing a plurality of pictures using a quantization step and a small quantization step having a size smaller than a size of the quantization step:, US8948255[P]. 2015.

[11] C. Wu,H. Li ,H. K. Lam , et al. Fault detection for nonlinear networked systems based on quantization and dropout compensation: An interval type-2 fuzzy-model method, *Journal of Neurocomputing*, vol. 191, pp. 409-420, 2016.

[12] T. Szirnyi, Subpixel pattern recognition by image histograms, *Journal of Pattern Recognition*, vol. 27, no. 8, pp. 1079-1092, 2015.

[13] A. Yoshimura , A. Matsugi , Y. Esaki , et al. Blind humans rely on muscle sense more than normally sighted humans for guiding goal-directed movement, *Journal of Neuroscience Letters*, vol. 471, no. 3, pp. 171-174, 2010.

[14] C. Vielhauer , R. Steinmetz, Handwriting: Feature Correlation Analysis for Biometric Hashes, *Eurasip Journal on Advances in Signal Processing*, vol. 4, pp. 1-17, 2004.

[15] C. T. Li, T. Y. Wu, C. L. Chen, C. C, Lee, C. M. Chen, An Efficient User Authentication and User Anonymity Scheme with Provably Security for IoT-based Medical Care System, *Journal of Sensors*, vol. 17, 1482; 18 pages, doi:10.3390/s17071482, June 2017.

[16] C. M. Chen, L. Xu, K. H. Wang, S. Liu, T. Y. Wu, Cryptanalysis and improvements on three-party-authenticated key agreement protocols based on chaotic maps, *Journal of Internet Technology*, 2017 accepted. DOI: 10.6138/JIT.2018.19.5.20160710.

[17] C.M. Chen, W. Fang, K.H. Wang, T.Y. Wu, Comments on: An improved secure and efficient password and chaos-based two-party key agreement protocol, *Nonlinear Dynamics*, vol. 87, no. 3, pp. 2073-2075, 2017 Feb. Doi: 10.1007/s11071-016-3171-9.

[18] C.M. Chen, L. Xu, T.Y. Wu, C.R. Li, On the Security of a Chaotic Maps-based Three-party Authenticated Key Agreement Protocol, *Journal of Network Intelligence*, vol. 1, no. 2, pp. 61-66, May 2016.

[19] T. Y. Wu, Y.M. Tseng, T. T. Tsai, A revocable ID-based authenticated group key exchange protocol with resistant to malicious participants?, *Computer Networks*, vol. 56, no. 12, pp. 2994-3006, 2012.

[20] T.Y. Wu, Y.M. Tseng, An ID-based mutual authentication and key exchange protocol for low-power mobile devices, *The Computer Journal*, vol. 53, no. 7, pp. 1062-1070, SEP 2010.

[21] H. Li , S. Yin , J. Liu , et al. Novel Gaussian approximate filter method for stochastic non-linear system, *Journal of International journal of innovative computing information and control*, vol. 13, no. 1, pp. 201-218, 2017.

[22] T. Y. Wu, Y. M. Tseng , S. S. Huang , et al. Non-repudiable Provable Data Possession Scheme with Designated Verifier in Cloud Storage Systems, *Journal of IEEE Access*, PP(99):1-1, 2017.

[23] M. Mastriani, J. Gambini, Fast Cosine Transform to increase speed-up and efficiency of Karhunen-Loeve Transform for lossy image compression, *Journal of International & Mathematical Sciences*, 2016.

[24] A. Sharma , A. K. Singh ,P. Kumar, Combining Haar Wavelet and Karhunen-Loeve Transform for Robust and Imperceptible Data Hiding Using Digital Images, *Journal of Journal of Intelligent Systems*, 2017.

[25] S. Yin, J. Liu , L. Teng A New Semi-supervised Clustering Algorithm Based on Variational Bayesian and Its Application, *Journal of TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 14, no. 3, pp. 1150-1156, 2016.

[26] H. Li, S. Yin, T. Liu , et al. An Improved Quadrature Kalman Filter Based on Gauss-Hermite, *Journal of ICIC express letters. Part B, Applications: an international journal of research and surveys*, vol. 6, no. 11, pp. 3049-3054, 2015.

[27] S. L. Yin , T. H. Liu, H. Li Application of Kalman filtering in indoor location based on simulated annealing algorithm, *Journal of Journal of Shenyang Normal University (Natural Science Edition) (Chinese)*, vol. 33, no. 1, pp. 86-90, 2015.

[28] Y. M. Qiu, H. C. Hu, J. B. Zheng , et al. On-line handwriting signature verification based on curve similarity, *Journal of XI Tong Gong Cheng Yu Dian Zi Ji Shu/systems Engineering & Electronics*, vol. 36, no. 5, pp. 1016-1020, 2014.

[29] J. Mohana, V. T. Bai, Implementation of Efficient Cryptographic Algorithm Based on Dynamic Biometric Key Generation Technique, *Journal of Sensor Letters*, vol. 14, no. 10, pp. 1044-1048, 2016.

[30] Martini , Wahyuni, C. Umbul, Subekti, Sri, et al. Competence Aedes as Vectors Based on Biological Characteristics and Vulnerability of Dengue Virus in Semarang City-Indonesia, *Journal of Advanced Science Letters*, 2017.

[31] A. Aparna, S. Ajish Cryptographic Key Generation based on Contextual Information: A Review, *Journal of International Journal of Computer Applications*, 2016, 134.