International Journal of Security and Its Applications
Vol. 9, No. 5 (2015), pp. 133-140
http://dx.doi.org/10.14257/ijsia.2015.9.5.14

Research on Encryption Key Extraction From Iris Feature

Wei Hao and Yong Wei

Xinxiang Vocational and Technical College, Xinxiang 453002, China
13782570917@139.com

Abstract

The current encryption algorithm has some problems that key length is long, which is difficult to be memorized and kept so that a potential threat is caused to information security. The encryption key extracted from the biological feature is used for the encryption method, which has become a hotspot of the research. The Haar wavelet decomposition is carried out to the iris image, to extract the third-layer high frequency coefficient as the iris feature code. The random mapping function is used to generate a 128 - bit key. chi-square (χ^2) test is used to analyze the key safety extracted. The results show that the key extracted from iris feature can meet requirements of the randomness and security of the encryption algorithm.

Keywords: *iris feature, wavelet decomposition, key extraction, χ^2 test*

1. Introduction

Many researchers have studied the identification and authentication based on biological characteristics, which has significant differences from the key generation based on the biological characteristics. The former uses biological extracted features to distinguish users certified and counterfeiting, while the latter converses the biological characteristics conversion as a unique key, which cannot be obtained by the counterfeiting [1]. Chang Yaojen [2] proposes a framework of key generation based on biological features.

However, the main problems of key generation based on the biological features are: the biometric data of the extraction is affected by the nature and acquisition environment. Even though, the data is extracted and received again to the same biological characteristics, it is similar to the data of the biological features which is previously extracted. The previous and late data of the encryption key must be completely consistent. Otherwise, the encryption system cannot run. The "similarity" and "uniqueness" of the biological characteristics have become the principal contradiction in the research, and therefore, it is necessary to solve this kind of contradiction by a protocol or fuzzy method.

The earliest key generation based on the iris features is the private template scheme of Davida in 1998 [3]. A typical iris features is extracted directly as a key, binding correcting error code. The database can only store hash values and personal information of the key. In the stage of the recovery key, after the verification of personal information, correcting error codes released and biological features gathered are used to reconstruct the key. The hash values of the key reconstructed are compared to the stored hash values to decide whether to accept the key. The error-tolerant rate of this method is only 10%, while the different digits of feature codes between different samples from the same iris can reach 32% [4]. Furthermore, the experimental results of the scheme are not reported, so that this algorithm only stays in the theoretical level. The representative study is the key generation technology of the iris features based on correcting error codes proposed by Feng Hao in 2006 [5]. First, a random key of 140bit is generated, and then the key is expanded as the pseudo code of the iris features of 2048bit through the two-level encode of Reed-Solomon and Hadamard. Finally, XOR is operated between the iris features template and the pseudo code to gain the lock code and is stored in the smart card. Decoding restores the iris pseudo code through collecting the sample features of the same

ISSN: 1738-9984 IJSIA

types of iris and correcting error code to recover key and be verified by the Hash function. Feng Hao thinks that difference is noise between the iris features used by the encryption and the iris features used by the decryption. The two-level coding of Hadmard and Reed-Solomon is used to remove such noise, but fault tolerance rate of the algorithm is 27%. In addition, improved fuzzy extraction technique is used by Boyen to generate a key based on iris features [6]. The place improved is: before using hash function to generate the key extracted from iris features, the sequence positions of iris features codes are carried out fixed replacement so that security of keys from the same iris features doesn't influence each other.

Firstly, among the existing methods of the key generation based on the biometric, excluding Feng Hao's key generation method based on iris features, FRR is more than 20%, which has a larger gap from the actual application. Secondly, because of the iris with the rich texture features, the generated feature code is long, which means that the space of key generation is larger, the extracted key is longer, and the security of the encryption is higher. However, the generated key of other biological features is shorter not to meet the demand of the general encryption algorithm to the key length (at least 56bit), which becomes the unique advantage of a key generated by iris features. Finally, in the key generation method, Davida and Boyen directly use the iris features as a key, while Feng Hao uses the "binding method" actually between the key and iris features, but correcting error codes are basically used for the inhibition to overcome the inconsistency the same iris features codes before and after the acquisition. The correcting error codes and fuzzy extraction techniques are unanimous in thinking and both employ the similarity of the same biological features to hide the key, but restoring ways of keys are different.

Furthermore, the application of correcting error codes will increase the extra storage space, reduce the coding efficiency and decrease the speed of the code transmission. If the Reed-Solomon code is used as the correcting error code, the RS code can correct t mistakes, and the number of its check code is $r = 2t$. 2048bit iris features code of Daugman is taken as an example. Generally, the inconsistent ratio of feature codes of the same type of the iris is 25%, that is to say, the error needs correct t=512bit, and then Supervision code r=2t=1024bit. The iris feature codes with RS correcting error code gets up to 3072bit long, which means 1/3 of the storage space of iris feature codes is used for the storage of correcting error codes.

2 The Two-dimensional Discrete Wavelet Decomposition of the Image

The principle of the fast algorithm of two-dimensional discrete wavelet transform of the image is shown in Figure 1. Hi_D represents the high pass filter of the wavelet decomposition, while Lo_D is the low pass filter of the wavelet decomposition. In the figure, two-dimensional discrete wavelet transform to the image is to use the fast algorithm of one-dimensional wavelet transform for the rows and columns of the image [7]. Two-dimensional discrete wavelet transform to the image is made from high to low scales in Figure 2. The first level decomposition is unfolded in the original image, which is similar to the second level decomposition being carried out in the approximate sub-graph LL1. According to the same method, the image is decomposed hierarchically. Each level of the wavelet decomposition gets approximate sub-graph LL, and details sub graph HH, HL and LH.

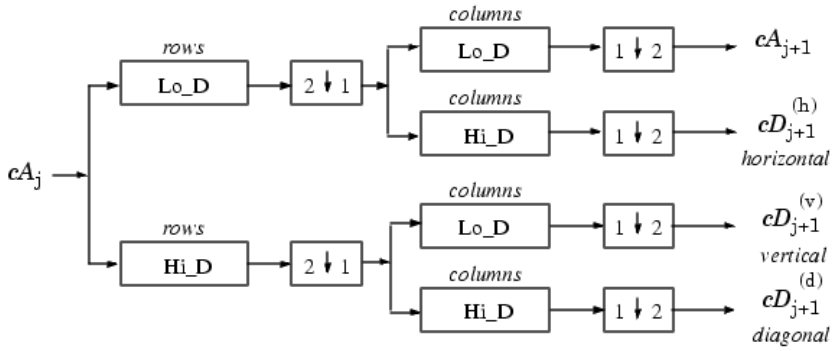


Figure 1. The Fast Algorithms of the 2D Discrete Wavelet Transform

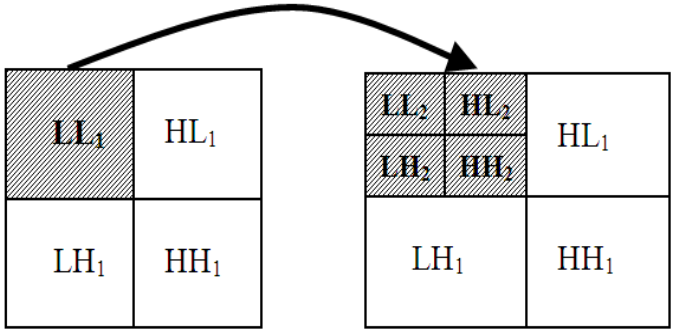


Figure 2. The Wavelet Decomposition

3 Iris Features Extraction

After the steps of preprocessing, localizationing, eyelash and eyelid detection to the iris image [8], the normalized extraction region of the iris features is obtained. In Figure 3, the white area is interfered by the eyelash and eyelid. R_2 region contains less iris texture features, which is easy to be shaped by the upper and lower eyelids. R_3 area is seriously interfered by the upper eyelid and eyelash. Based on the analysis to a large number of the original iris image, it is found that the upper eyelid covering the iris is more serious than the lower eyelid, and the upper eyelash interferes more severely to the iris region [9]. Based on the above analysis, R_4 region is selected as the extraction region for iris features.

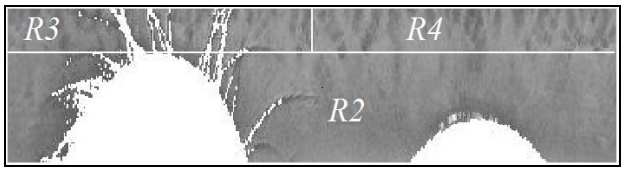
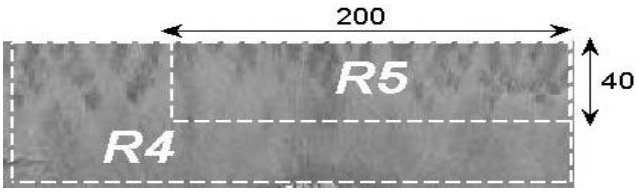


Figure 1. The Normalized Image Interfered by Eyelids and Eyelashes

To generate the fixed iris code, R_5 region with the fixed size is selected from up to down and from right to left for the features extraction from R_4 region as shown in Figure 4. The part of the iris region undisturbed is used for the features extraction and comparison. The features extraction has a certain limitation to the size of the region. According to the experimental results of the literature [10], the size of the effective R_4 of the iris is not less than 1/6 of the size of the normalized iris image. After testing the large number samples, the size of the fixed region is selected as 40×200.

(a) The Effective Area of the Iris



(b) The Features Extraction Area Selected in the Iris Effective Area

Figure 4. The Region Selection of the Features Extraction

The region R_5 of the features extraction in Figure 4(b) is carried out three-layer decomposition of two-dimensional Haar wavelet, and in Figure 5 the results of decomposition is shown.

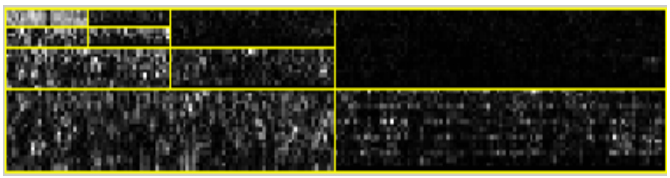


Figure 5. The Three-layer Decomposition of Two-dimensional Haar Wavelet in the Region of the Features Extraction

The component of the iris texture features mainly concentrates on the third layer of decomposition [11]. At the same time, the iris texture information is the detail information of grayscale change, mainly concentrating on the high frequency coefficients. If the high frequency coefficients of the first or second layers as features will lead to too large feature space, to not only influence the coding efficiency, but also the matching speed [12].

Extracting the high frequency sub image of the third layer, LH_3, HL_3 and HH_3 is regarded as the iris features. The sub image size of each high frequency is $(40\times200)/(2^3\times2^3)=125$. The high frequency sub images of three directions are composed of the space vector of the iris feature C={ LH_3 ,HL_3 HH_3}, and its space size is $125\times3=375$. The features extraction region is carried out the two-dimensional wavelet decomposition, and the intrinsic correlation between 2D textures of the iris image is considered. The space vector of iris feature extraction C contains 375 wavelet coefficients with suitable size to store and encode easily. Space vector C combines the high frequency coefficients in three directions, which can represent the iris texture features well.

The statistical analysis is done for all the wavelet coefficients generated by Haar wavelet decomposition and high frequency wavelet coefficients of the third layer (iris feature space vector C), as shown in Figure 8.

In Figure 6, it can be found that the distribution of the high frequency wavelet coefficients of the third layer is similar to the normal distribution. The peak is reached when the coefficient value is about 0. According to the statistics in Figure 6, all the wavelet coefficients are negative and the probability is 0.480, while the wavelet coefficient is positive and the probability is 0.520. According to the statistical data in Figure 7, the wavelet coefficient of the third layer is negative and the probability is 0.553, while it is positive and the probability is 0.447. The wavelet coefficient is positive or

negative and the probability is 1:1, which provides good conditions for the feature encoding.

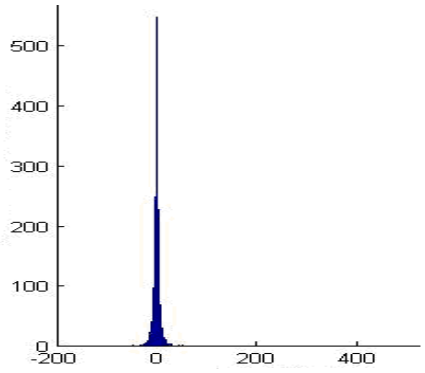


Figure 6. The Distribution of All Wavelet Coefficients

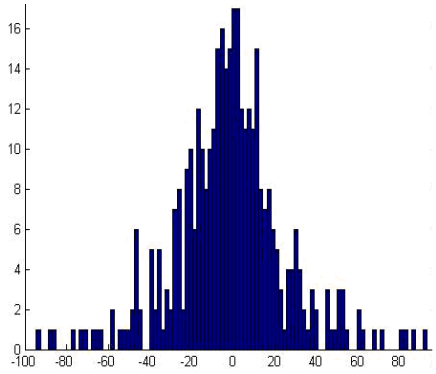


Figure 7. The Distribution of High Frequency Wavelet Coefficients of the Third Layer

Binary encoding to the feature vector C is a very important step. In practice, it is always easy for the Boolean vector to compare and operate by computer. Based on the above analysis to the wavelet coefficients distribution, the wavelet coefficients also express the similar degree of wavelet and signal. The similar degree of the description of the positive wavelet coefficients and the negative wavelet coefficients has the great differences. To improve the encoding efficiency of the iris features and reduce the matching time, 0 is taken as the threshold value and make the feature vector C into binary code.

4. The Encryption Key Extraction based on the Iris Features

The iris feature codes are set as a sequence of $K=\{k_1,\ldots,k_i, \ldots, k_m\}$, the key is selected from the sequence K through the random function. A binary sequence extracted from the iris feature codes is set as $P_1=\{p_1,\ldots, p_j,\ldots,p_n\}$. The mapping function $f:P_1\rightarrow K$ is used for one to one bit mapping. There are a variety of mapping methods. The simplest method is to circulate according to the natural arrangement order of key P one by one for cycle of mapping, namely $p_1=k_1,\ldots,p_i=k_i,\ldots,p_n=k_n$. After k_n is selected and the next round of selection starts from k_{n+1}. In order to improve the security of encryption and enhance the difficulty of deciphering, the key sequence P_1 can be mapped to K according to a certain random function f.

$p_j=k_i$, is set and the function f actually performs the subscript of the key sequences, that is $i=f(j)$. $f(j)$ is defined as:

$$f(j)=[(m-j+z_r) \bmod m]+1,\ z_r \in Z \quad (1)$$

z_r is a pseudo-random integer, and the linear congruential method is used to generate the pseudo-random integer. The large integer is selected as the seed, and$0<z_r<10^7$. The length of extracted iris feature codes is 375bit, that is m=375. The length of key is 128 bit, that is n=128. By formula (1) mapping, 128 bit key P_1 is obtained.A total of 10^5 key extraction experiments have been done and the average cost time is 0.11 seconds.

5. Safety Analysis of the Encryption Key Extracted from Iris Features

The performance of the key generation method actually is determined by the good or bad key generated. The important characteristics of the safety key are random or unpredictable. The index j of P_1 is randomly selected (such as j=10). According to formula (1) the 10^7-time mappings have been done, and k_i of p_j mapping in K is observed, namely the j mapping distribution in i. In Figure 8, the stick figure shows the distribution. It can be seen from the figure that the value of j in i is basically a uniform probability distribution, which describes randomness of $P_1 \rightarrow K$ bit mapping from the qualitative perspective.

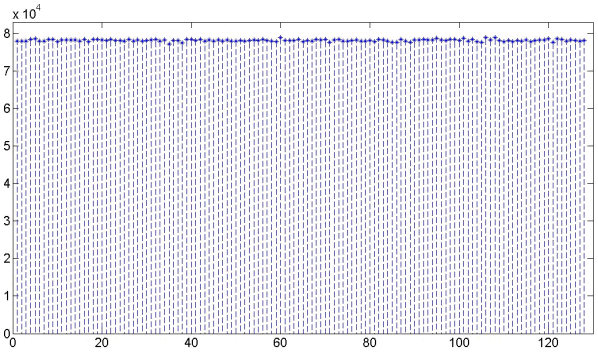


Figure 8. *j* Mapping Distribution in *i*

From the statistical perspective, χ^2 test is used to test the randomness of key selected from iris feature codes, that is to check j mapping distribution in i is homogeneous.

χ^2 test is the deviation degree between the actual observation value of the statistical sample and the theoretical value, which determines the size of χ^2 value. The greater the deviation is, the greater the χ^2 value is, which is not more suitable; the smaller the deviation is, the smaller the χ^2 value is, which is more suitable. In the χ^2 test, a hypothesis to be tested is put forward, which is called as the original hypothesis or the null hypothesis, and denoted as H_0. The hypothesis opposite to the original hypothesis is called the alternative hypothesis, and denoted as H_1 [13]. The main steps are as follows:

(1) Determining the basic hypothesis

The original hypothesis H_0: the key selected from iris feature codes is random, that is to say, the index j mapping in i is in uniform distribution.

The alternative hypothesis H_1: the key selected from iris feature codes are non-random.

(2) Calculating the test statistics.

When H_0 is true, the difference of the actual observed number O_i and theoretical observed number E_i should be relatively close to 0. Therefore, when H_0 is true, the test statistic of formula (2) shown obeys the χ^2 distribution of freedom degree as m-1.

$$\chi^2(m)=\sum_{i=1}^{m}\frac{(O_i-E_i)^2}{E_i},1\le i\le m \quad (2)$$

For any j($1<j<128$) to i($1<i<375$) for n mappings, if the index j mapping in i is uniformly distributed, and then the probability of any i value shown should be $1/m$, and m is the digit of binary sequences of iris feature codes and m=375. Theoretical frequency $E_i=n\times(1/m)=n/375$. The i value frequency shown actually is calculated, which is the actual observed number O_i.

(3) Accepting or rejecting the hypothesis.

According to the set level a, $\chi_\alpha^2(m)$ of a quantile is found, and $\chi^2(m)$ and $\chi_\alpha^2(m)$ are compared, if $\chi^2(m)>\chi_\alpha^2(m)$, H_0 is refused; otherwise H_0 is accepted.

all j is carried out for i mapping experiments, and the time of mapping n=10^7. χ^2 test is performed for mapping j to i and a total of 128 test statistics are obtained, and degree of freedom m-1=374. The choice level of significance a =0.05 and based on the tables, $\chi_{0.05}^2(374)=419.8$. In Figure 9, the χ^2 statistics distribution of different j is shown.

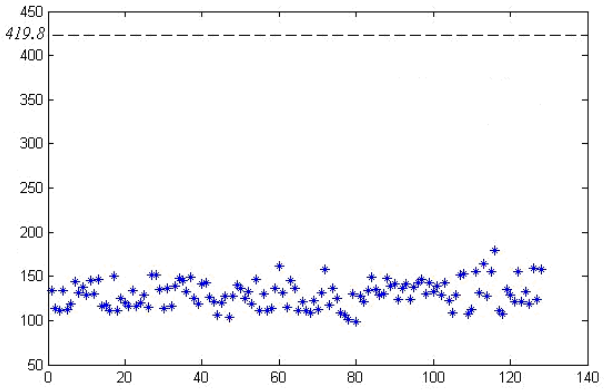


Figure 9. Distribution of χ^2 Statistics Value

From the Figure 10, the test statistic data of any mapping of j to i $\chi^2(374)<\chi_{0.05}^2(374)$ is obtained. Therefore, the H_0 hypothesis is accepted, that is to say, the key selected from the iris feature codes is a random sequence.

6. Conclusion

The key extraction algorithm based on iris features is proposed. The third layer high-frequency coefficient of wavelet decomposition is extracted as the iris feature codes based on preprocessing, localization and interference detection. The key generation conditions of extracting from iris feature codes are analyzed and random mapping function is used to generate a 128-bit key. The χ^2 test is used to analyze safety of the extraction key. The analysis results show that the extraction key from iris features meets the needs of the randomness and the security of the encryption algorithm.

References

[1] H. Chang, G. X. Zhou and Z. S. Wu, "A study on key generation based on biological characteristics", Application Research of computers, vol. 24, no. 7, **(2007)**, pp. 133-134.
[2] Y. J. Chang, W. D. Zhang and T. H. Chen, "Biometrics-based cryptographic key generation", Proceedings of IEEE International Conference on Multimedia and Expo, **(2004)**, Taiwan.
[3] G. Davida, Y. Frankel and B. Matt, "On enabling secure applications through offline biometric identification", Proceedings of IEEE Symposium on Security and Privacy, **(1998)**, Oakland, California.

[4] K. Bowyer, K. Hollingsworth and P. Flynn, "Image understanding for iris biometrics: a survey", Computer Vision and Image Understanding, vol. 11, no. 2, **(2008)**, pp. 281–307.
[5] F. Hao, R. Anderson and J. Daugman, "Combining Cryptography with Biometrics Effectively", IEEE Transactions on Computers, vol. 55, no. 9, **(2006)**, pp. 1081–1088.
[6] X. Boyen, "Reusable Cryptographic Fuzzy Extractors", proceedings of the 11th ACM Conference on Computer and Communications Security, **(2004)**, Washington DC, USA.
[7] J. Jang and K. R. Park, "A Study on Multi-unit Iris Recognition", Proceedings of 8th Control, Automation, Robotics and Vision Conference, **(2004)**, Noordwijkerhout, Netherlands.
[8] Q. C. Tian, "The principle and algorithm of iris recognition", National Defence Industry Press, Beijing, **(2011)**.
[9] H. Sung, J. Lim and J. Park, "Iris Recognition Using Collarette Boundary Localization", Proceedings of the 17th International Conference on Pattern Recognition, **(2012)**, Washington DC, USA.
[10] W. Boles and B. Boashash, "A Human Identification Technique Using Images of the Iris and Wavelet Transform", IEEE Transactions on Signal Processing, vol. 46, no. 4, **(1998)**, pp. 1185-1188.
[11] S. Noh, K. Bae and Y. Park, "A novel method to extract features for Iris recognition system", Proceedings of the 4th international conference on Audio- and video-based biometric person authentication, **(2003)**, Heidelberg, Berlin.
[12] J. Zhou, T. Luo, M. Li, *et al.*, "Using 2D Haar Wavelet Transform for Iris Feature Extraction", Asia-Pacific Conference on Information Theory, **(2010)** November 1-3, Xian, China.
[13] Y. R. Xiao, "The calculating method of the probability and statistics", Nankai University press, Tianjin, **(1994)**.

Authors

Wei Hao, graduated from Henan University of Finance and Economics in 2003, bachelor's degree. He is a lecturer at Xinxiang Vocational and Technical College.Research direction: computer application.He has published 5 professional papers in journals and conferences.

Yong Wei, received her Master Degree from Wuhan University of Technology in 2009. He is now a lecturer of Henan Mechanic and Electrical Engineering College in China. Her research interests include software engineering, data mining, Computer Graphics. He has published more than 7 papers in journals and conferences.

