A Secure Internet-Based Personal Identity Verification System Using Lossless Watermarking and Fingerprint Recognition

Guorong Xuan[1], Junxiang Zheng[1], Chengyun Yang[1],
Yun Q. Shi[2], Dekun Zou[2], Liu Liansheng[3], and Bai Weichao[3]

[1] Tongji University, Shanghai, China
[2] New Jersey Institute of Technology, New Jersey, USA
[3] Baojia Electronic Equipments, Co. Ltd, Shenzhen, China

Abstract. This paper proposes an internet-based personal identity verification system using lossless data hiding and fingerprint recognition technologies. At the client side, the SHA-256 hash of the original fingerprint image and sensitive personal information are encrypted and embedded into the fingerprint image using an advanced lossless data hiding scheme. At the service provider side, after the hidden data are extracted out, the fingerprint image can be recovered without any distortion due to the usage of the lossless data hiding scheme. Hence, the originality of the fingerprint image can be ensured via hash check. The extracted personal information can be used to obtain the correct fingerprint feature from the database. The fingerprint matching can finally verify the client's identity. The experimental results demonstrate that our proposed system is effective. It can find wide applications in e-banking and e-government systems to name a few.

1 Introduction

As more and more activities are being conducted through Internet, people are more willing to use this convenient way to handle personal business including financial activities. Some of them involve large amount of money. However, Internet by itself is not a safe place. Criminals prefer to rob banks by using advanced technology rather than by using gun nowadays. Most of the online crimes rely on feigning the identity of other people. Thus, personal identity verification has emerged as a crucial issue for the safety of online activities.

In this paper, we proposed a smart internet-based personal identity verification system using lossless digital image watermarking and fingerprint recognition. On the one hand, our system utilizes the fingerprint to identify person. We use one way hash SHA256 to ensure the originality of the fingerprint image. On the other hand, personal information is embedded into the fingerprint

I.J. Cox et al. (Eds.): IWDW 2004, LNCS 3304, pp. 55–65, 2005.

image losslessly and imperceptibly so that it can be transmitted through the Internet covertly. At the client side, fingerprint and encrypted user information serve as input of the identity verification. At the service provider side such as banks verify the fingerprint by using the hash value and the encrypted data to authorize the remote client. We claim that the combination of fingerprint and encryption can achieve better security than either of them alone. Firstly, secure messages are embedded into the fingerprint images. As a result, the data amount needed to transmit will not increase. Secondly, since the secure message cannot be seen directly by human eyes, it is safer than encryption alone in which the malicious attackers will easily know that there exist secret messages. Thirdly, the fingerprint is hashed and the hash value is embedded into the fingerprint image itself. As a result, the originality of the fingerprint can be guaranteed via authentication.

To the best knowledge of the authors of this paper, there is no prior art concerning embedding information into fingerprint images in the literature. Our proposed system can be applied to various applications such as e-banking, e-trading and e-government systems. In fact, it has been applied into an online pension distribution system in China.

2 Invertible Data Hiding

The invertible data hiding scheme adopted in this paper is to embed watermark into middle bit-planes of the image wavelet coefficients of the highest frequency sub-bands. We apply histogram modification before embedding to ensure that no over/under-flow will occur after data hiding. The watermark scheme we used has very high capacity and low visual artifact. In addition, the computational complexity of the proposed algorithm is low. In this section, we present a brief introduction of the basic idea of the scheme. The new improvements of the scheme [1], that has made the scheme much more efficient, and the application of the updated scheme to the verification system are described.

2.1 Selection of Wavelet Family

Although our invertible watermarking scheme can be applied to various wavelet families, after experimental comparison, we have discovered that CDF(2,2) is better than other wavelet families in terms of embedding capacity and visual quality of watermarked image in general.

2.2 Selection of Embedding Sub-band and Bit-Plane

Manipulating on sub-band HL, LH and HH of the highest resolution level will cause the least visual artifact because they contain only the highest frequency components. In these sub-bands, we embed data into a particular bit-plane. For a coefficient, if a flip occurs in the n^{th} bit-plane, its value will change by amount of 2^{n-1}. For example, a flip on 4^{th} bit-plane will either increase or decrease the coefficient value by 8. The higher the bit-plane is, the larger the change is and the lower the marked image quality will be. However, usually, the higher the bit-plane is, the shorter bit stream it can be compressed into. As a result, higher embedding capacity can be achieved. Hence, there is a trade-off between image quality and embedding capacity. Instead of only embedding data in one bit-plane as in [1], we propose here to use multiple bit-planes for data hiding if the capacity provided by one bit-plane is not enough. We define the bit-plane(s) used to embed data as *Embedding Bit-plane(s)* , and the remaining as *Untouched Bit-planes.*

2.3 Histogram Modification

The purpose of histogram modification is to compress the histogram range of the original fingerprint image so that, after data hiding, the pixel value of the marked image will not exceed the permitted value range which is (0,255) usually. If over/under-flow occurs, the lossless property of the watermarking scheme will be lost and the original fingerprint image cannot be recovered.

We select G grey levels, G/2 on both borders, which need to be merged as illustrated in Fig. 1 After modification, the histogram range is reduced to [G/2, (255-G/2)]. In order to recover the original fingerprint losslessly, the information about how to recover the original histogram will be embedded into the fingerprint as well as overhead. The parameter G is related to which wavelet family we choose and which sub-band we choose as the embedding carrier. An analysis is provided as follows:

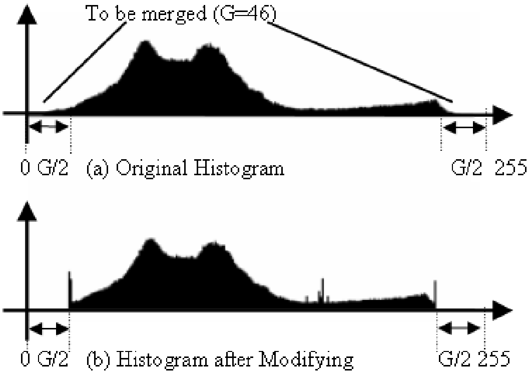

Fig. 1. Histogram modification

If CDF(2,2) is selected and the data are to be embedded into the fourth bit-plane of HH, HL, LH sub-bands of the highest resolution level, according to integer lifting scheme formula,

$$d_i \Leftarrow d_i - \left\lfloor \frac{1}{2}\left(s_i + s_{i+1}\right) + \frac{1}{2} \right\rfloor \tag{1}$$

$$s_i \Leftarrow s_i - \left\lfloor \frac{1}{2}\left(d_{i-1} + d_i\right) + \frac{1}{2} \right\rfloor \tag{2}$$

We can estimate pixel value change. If a flip on the 4^{th} bit-plane occurs, the coefficients will change by amount of 8. We can find out that the pixel value will change by amount of M=42 in the worst case. Thus, the histogram adjusting value G should be 84.

The estimation of M considers only the worst case in all steps. In reality, the probability that this result occurs is very small. According to our experiments, when moderate watermarking capacity is required, it is enough for G to be 30 to avoid over/under-flow for most natural images [1].

Although the histogram modification is to prevent over/under-flow, it will degrade the visual quality of the marked image. Two problems are brought out by big G. One is that the number of pixels needed to be changed increases. As a result, visual quality degrades. The other problem is the information for recovering original histogram increases. Thus, the effective capacity will decrease. In conclusion, the value G should be as small as possible.

Obviously, we need to embed the data representing the bookkeeping information in histogram modification in order to late recover the original image. In [1] a coordinate-based bookkeeping is used, in which the coordinates of pixels moved towards the middle and their gray levels are recorded. This results in a large amount of bookkeeping data when either the image size is large or the number of pixels assuming two end levels is large. Here, we propose to use the merge-and-shift based bookkeeping method. That is, to empty G/2 gray levels at one side of the histogram, we choose G/2 suitable gray level pairs (two neighboring gray levels form a pair) from examining the histogram. We merge a pair of levels and leave an empty level. The rest of gray levels are shifted from the end towards the middle of the gray levels. For the other side of histogram, we do the similar modification. This new method can save bookkeeping data effectively.

2.4 Data Embedding and Extraction

As illustrated in Fig. 2, the original fingerprint image is subject to histogram modification first. After forward wavelet transform applied to the modified image, the embedding bit-plane(s) is(are) compressed with JBIG algorithm. The left-over part is used to embed data. Three kinds of information are needed to be embedded. They are the hash value of the fingerprint image, the information about recovering the original histogram and the user's personal information. All of them are combined together and encrypted with an encryption key. The encrypted bit-stream is embedded into the left-over part of the embedding bit-plane(s). Then, the embedding bit-plane(s) and the untouched bit-planes are

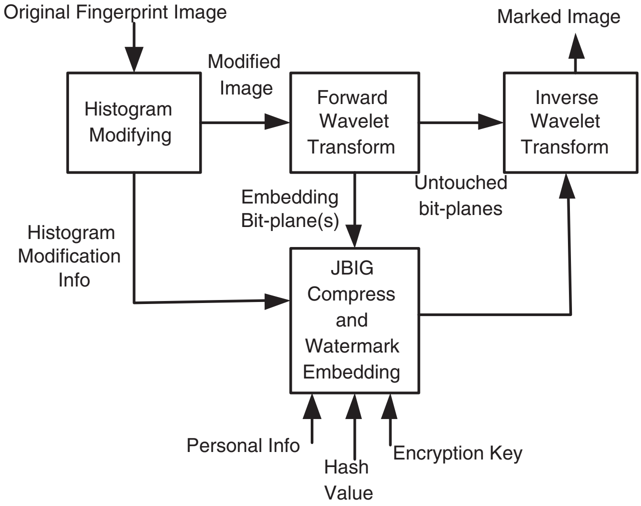
Original Fingerprint Image
Histogram Modifying
Modified Image
Forward Wavelet Transform
Inverse Wavelet Transform
Marked Image
Histogram Modification Info
Embedding Bit-plane(s)
Untouched bit-planes
JBIG Compress and Watermark Embedding
Personal Info
Hash Value
Encryption Key

Fig. 2. Watermark embedding

combined together for inverse wavelet transform. Finally, we have the marked fingerprint image which is ready for transmission. Watermark extraction procedure is the reverse of the embedding part which is illustrated by Fig. 3. After the embedded data is extracted, it is decrypted with a decryption key. Two encryption/decryption schemes can be applied in our system. One is symmetric scheme in which both keys are the same and the key is needed to be transmitted. The other is non-symmetric scheme which is known as public/private key pair. The recovered fingerprint image will be subject to hash verification to determine its originality.

2.5 Performance Analysis

We have applied our algorithm to two fingerprint libraries. Library A is FVC2000 fingerprint database[3]. Image size is 300x300. Library B is generated in our work using Verdicom's fingerprint mouse[4]. The image size is 256x300. Our experiments show that for Library A, the embedding capacity can be as large as 0.6bpp. For Library B, the capacity is 0.11bpp. Both are enough for our system requirement. In terms of image quality, for A, if watermark is embedded at the rate of 0.4bpp, the PSNR of the marked image vs. the original image is over 30dB. For B, if the embedding data rate is 0.07bpp, the PSNR is over 24dB. All of marked images have no visual artifact. It is noted that the result of Library A is better than that of Library B.

Fig. 4 and Fig. 5 are the testing results of a fingerprint, named A1, in Library A. The histogram is low on borders. The pixel grey values range from 101 to 250. Hence the histogram modification will not change the image much. In the test, the amount of 0.11bpp data is losslessly embedded in the 4th bit-plane of HH subband of the highest resolution. G is set to be 12 so that the modified histogram

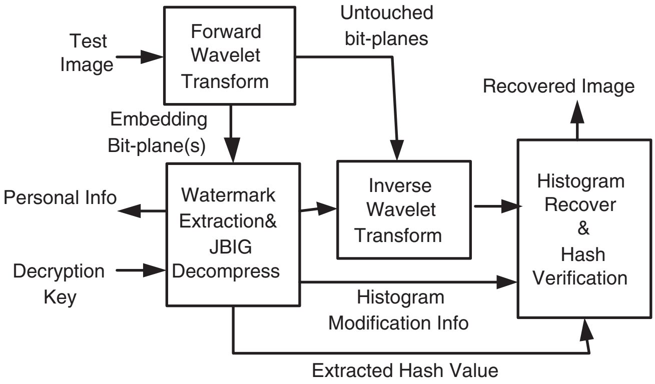
Test Image
Forward Wavelet Transform
Untouched bit-planes
Recovered Image
Embedding Bit-plane(s)
Personal Info
Watermark Extraction& JBIG Decompress
Inverse Wavelet Transform
Histogram Recover & Hash Verification
Decryption Key
Histogram Modification Info
Extracted Hash Value

Fig. 3. Watermark extraction

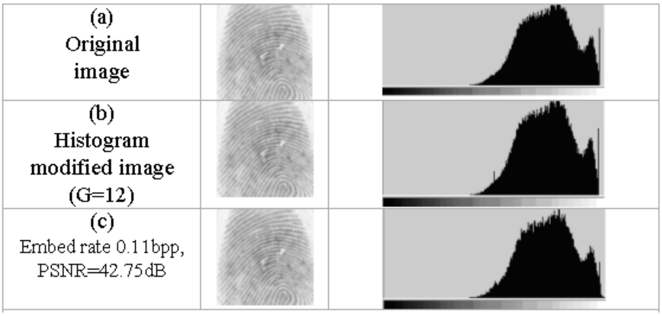

Fig. 4. Fingerprint A1 and its histogram before and after modification and embedding

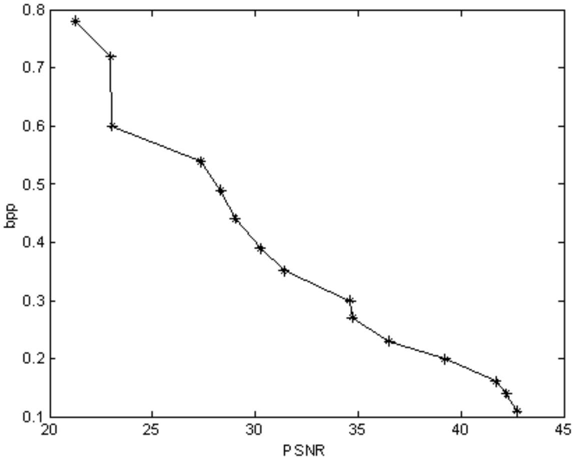

Fig. 5. Relationship between PSNR and payload for A1

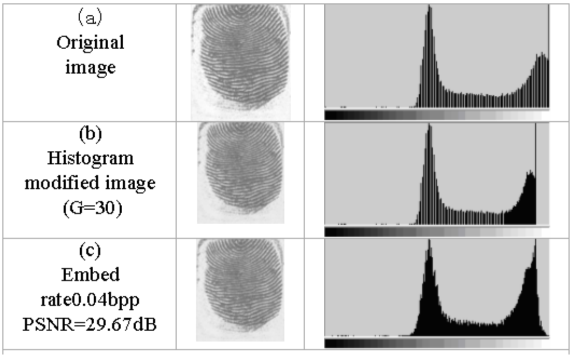

Fig. 6. Fingerprint B1 and its histogram before and after modification and embedding

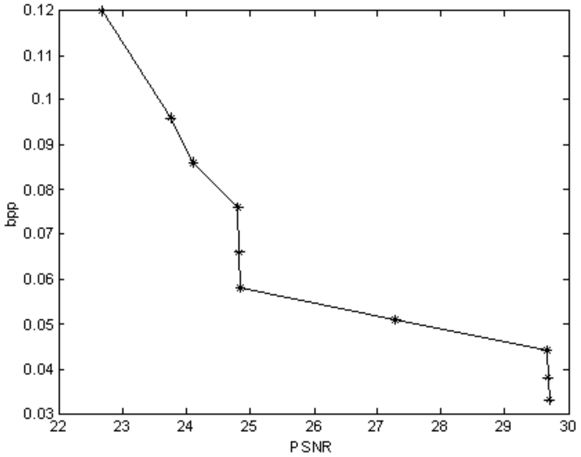

Fig. 7. Relationship between PSNR and payload for B1

ranges from 101 to 249. Only 1 grey level, 250, is needed to be merged. PSNR of marked image is 42.75dB. Library B is generated by Veridicom's fingerprint mouse[4]. All fingerprints in it are different from those in Library A in some aspects such as histogram structure. Although the result of Library B is not as good as that of Library A, both watermarking capacity and visual quality meet our system's need.

Fig. 6 and Fig. 7 are the testing results of a finger B1, in Library B. The histogram ranges from 0 to 255 and is high on both borders but low on center. 0.04bpp data is embedded in the 5^{th} bit-plane of the HH sub-band of the highest resolution. G is set to be 30, so that after modification, the histogram range will be compressed to [15, 240]. 30 grey levels are to be merged. The PSNR of the marked B1 is 29.67dB.

3 Fingerprint Recognition

We use Veridicom's fingerprint software and fingerprint mouse to capture the fingerprint image[2][4]. Fig. 8 is the mouse we used.

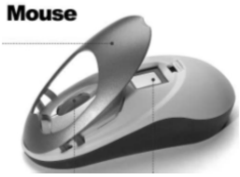

Fig. 8. Veridicom's fingerprint mouse

Because the watermarking scheme used in this paper is invertible, watermarking itself will not affect the fingerprint recognition process at all. Thus, the performance of recognition will be influenced by the fingerprint software and hardware. To test the performance, we apply the recognition to FVC2000 fingerprint database[3]. In this database, we select 8 categories of fingerprints and for each category, we choose 10 fingerprints. The total number is 8x10=80 fingerprints. Fig. 9 contains eight fingerprints in one category.

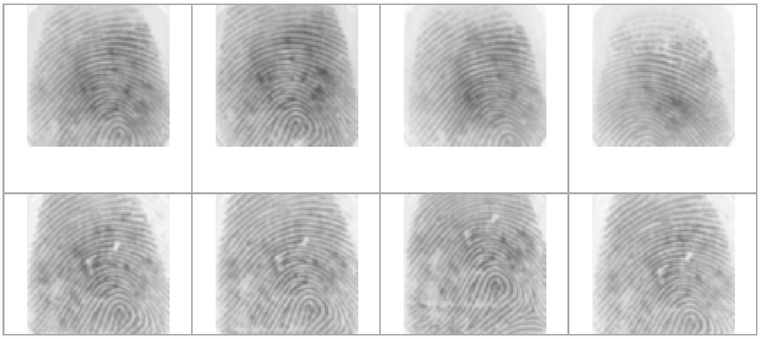

Fig. 9. Eight fingerprints in one category

Then, we select 8 fingerprints in each category to train the recognition software. Finally, we match all the 80 fingerprints to the training results. Table 1 is the matching results. If two fingerprints have the number of matched features great than the *Threshold Value*, we consider them belonging to the same person. The *Correct Matching Rate* is the rate by which the system can successfully identify a person. The *Error Rate* is the rate by which the system wrongly identify person Y in the database as person X. It is noted that the lower the Threshold Value is, the higher the matching rate will be. In addition, the error rate is zero

Table 1. Matching Results

Threshold Value	*Correct Matching Rate*	*Error Rate*
15	72.5%	0%
13	77.5%	0%
11	82.5%	0%
10	85%	0%
8	87.5%	0%
5	90%	0%
3	90%	0%
1	90%	0%

for all the threshold values listed. In order to increase the matching rate, the user can input their fingerprint by three times. Assuming they are independent event, the final matching rate will be 97.2% if we set the threshold to be 5 for all three matches.

4 Personal Identity Verification

Our proposed system can be used to verify a person's identity through Internet.

4.1 System Architecture

Fig. 10 is the proposed system architecture. Central Database stores the features of the registered users' fingerprints and their other personal information. The Server receives marked fingerprints transmitted from registered users and extracts personal information. If the identity is verified, the user will be authorized for further transaction.

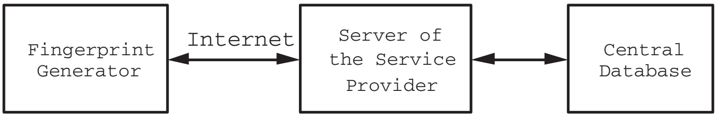

Fig. 10. System architecture

4.2 Software Module

Identity Registration. Before identity verification, user must register their fingerprint and other personal information in the central database. Specifically, fingerprint capture device will capture three fingerprint images of the same finger and extract the features of them. Finally, the features and user's personal information are stored in the central database.

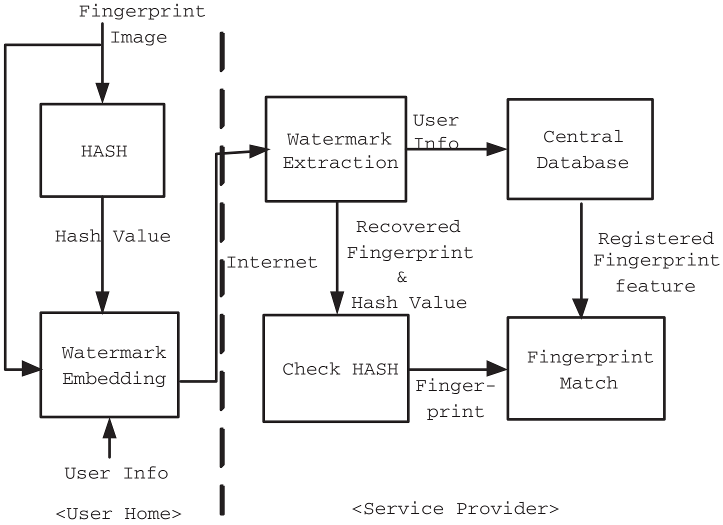

Fig. 11. Identity verification

Identity Verification. Identity verification procedure is depicted in Fig. 11. The fingerprint is compared with the registered fingerprint feature. If a match achieves, the user is authorized. It can be seen that two level of security is provided here. One is the hash check, the other is fingerprint match. The former authenticates the originality of the fingerprint, while the latter recognizes the registered person's identity. Failure of either one will cause the failure of the authentication.

5 Summary

In our proposed system, messages are embedded into the fingerprint image itself. No visible artifact can be detected. It is more secure than encryption protection method alone. The invertible feature of the adopted watermarking scheme combines with the feature of SHA, the originality of the fingerprint image is guaranteed.

Acknowledgement

This project is supported by China NSFC: The Research of Theory and Key Technology of Lossless Data Hiding (90304017), and NJCST via NJWINS.

References

1. G. Xuan, J. Chen, J. Zhu, Y. Q. Shi, Z. Ni, W. Su : Distortionless data hiding based on integer wavelet transform. Proceedings of IEEE Workshop on Multimedia Signal Processing (MMSP'02). US Virgin Islands, Dec. 2002; IEE Electronics Letters, vol. 38, no. 25, pp. 1646-1648, Dec.2002.
2. Veridicom Authentication Software Development Kit: Users Manual (For Evaluation) Version 3.0.0
3. http://bias.csr.unibo.it/fvc2000/
4. http://www.veridicom.com