A New Secure Storing System for Biometric Templates Based Encryption and Concealment

Emad Taha Khalaf and Norrozila Sulaiman
Faculty of Computer Systems and Software Engineering, University Malaysia Pahang, Kuantan, 26300, Malaysia

ABSTRACT
The security of templates is the critical part of biometric system and one of the most crucial issues in any proposed system. In fact, features such as voice, face, fingerprints and many others can be covertly acquired or stolen by an attacker and misused. Therefore, storing biometric templates a secure way is crucial. This study proposes a novel approach that combines an improved encryption method with a new concealment technique to establish a secure data template storing system. The Hill Cipher algorithm has been improved to be more secure by using a large and random key with large data block and also extending it to include the special characters. In the other hand a new concealment method proposed by combine two of the popular methods DWT and DCT for embedding and extraction the secret data in order to compensate the drawbacks of both of them and to make the hidden information much more secure against the attacks, wavelet transform which use dyadic filters to decompose cover image into 4-levels (HH, HL, LH and LL) and discrete cosine Transforms to convert a signal of the selected coefficients (HH, HL and LH) into elementary frequency components then according to the percentages entering by a user the encrypted data will distribute. The efficiency of encryption and concealment have been checked out with a number of widely used metrics such as Peak Signal to Noise Ratio (PSNR) and Normalized Correlation Coefficient (NCC).

Key words: Biometric, template security, encryption, data hiding

INTRODUCTION
Biometric field has taken a huge interest by global industry with protect and safeguard information as an everlasting necessity, biometric is uses to identify authorized person based on specific physiological or behavioral features. Examples of behavioral characteristics are gait, signature and voice. Physical characteristics include: DNA, ear, face, fingerprint, hand geometry, iris and retina (Griaule Biometrics, 2012). Salient features are extracted using some feature transformation technique and get converted into digital form. This digital information is stored in the database which is known as Biometric Template. Later the template is used during authentication purpose, compromised biometric templates are unlike passwords and tokens they cannot be revoked and reissued this led to become biometric template security is an important issue and protecting the template is a challenging task due to intra user variability in the acquired biometric traits, based on knowledge of the biometric characteristics (Malhotra and Kant, 2013). Cryptography was created as a technique for transmitting secret information between persons in a way that prevents a third party from reading it, many encryption methods have been developed to encrypt and decrypt data to make it meaningless and unintelligible unless the decryption key is available. There are three types of cryptosystems: Symmetric key, Asymmetric key and Hash Functions. Symmetric key (private key or secret key) encryption uses one key to encrypt and decrypt. The Hill Cipher is a block cipher and symmetric key algorithm, it was introduced to the journal of mathematics by Lester Hill as a short paper and published in 1929 (Hill, 1929). Hill Cipher has several advantages such as disguising letter frequencies of the
plaintext, simplicity because of using basic matrix operations, high speed, high throughput (Overbey et al., 2005; Saeednia, 2000). Unfortunately it has some disadvantages such as not include special characters and digits, takes the smaller sizes of blocks so key length is shorter, very simple and vulnerable for exhaustive key search attack and known plain text attack, also the key matrix which entered should be invertible. The proposed system is inspired from a number of researches which are related to cryptography and steganography biometric techniques. Many of researchers have proposed techniques to keep the security of biometric data. One of the suggested techniques was by Monrose et al. (1999), proposed a behavioral biometric key generation based on keystroke biometrics. They use dynamic features (duration of keystrokes and latencies between keystrokes) to strengthen a user’s password. This scheme makes the system more secure by adding 15 bits of entropy to the password for 15 dynamic features (Uludag et al., 2004). In their later studies, Monrose et al. (2001, 2002) applied this scheme to voice data. The algorithm to generate cryptographic keys from voices was mainly based on the speaker verification and identification technologies, such as digital signal processing, feature extraction and the vector quantization technique. Consequently, their system was eventually able to generate cryptographic keys up to 60 bits from voice features. However, the False Rejection Rate (FRR) was still high (20%). Moreover, since their approach was based on the VQ technique, the system left useful information which can be used in gaining access to the system. Another suggested technique was by Jain and Uludag (2003) with two scenarios of hiding data, first one in a cover image not related to the template data, other scenario by using the fingerprint image to hiding the facial information. Wang et al. (2010) and Sonsare and Sapkal (2011) use DCT transformation method to hiding the iris code and the secret information after encrypting in random blocks of the coefficients. Another security system have been proposed by Nitalianis et al. (2011) based on DWT transformation method to hides biometric signals in video objects over open network. One of the common methods is uses skin region of the image for embedded the template data (Kharage et al., 2013), HSV color space has been used to detect skin color tone, also DWT is used to embedded in one of the high frequency sub-band by tracing the number of skin pixels in that band, then cropped a region of the image which will be used as a key at the decoder side (Amritha and Meethu, 2013).

A Classical Least Significant bit Technique (LSB) introduced earlier steganography schemes in the spatial domain which directly embed the secret data within the pixels of the cover image (Masc and Quisquater, 2005). This approach has few disadvantages like the relative easiness to implement it makes the method popular, the proper cover image required to hide a secret message inside an image (Kolputzuk and Bouman, 1992). Many previous stenographic algorithms have been used pixel domain, its provides the space (capacity), reliability and controllability in encoding/decoding while embedding the hidden message but more of the steganalysis attacks will focus on this domain as expected. Kumar and Shumuganathan (2010) and Ramani et al. (2008) studied the steganography in transform domain since the compress of digital data will be provides a reduction of storage space and transmission cost. In this study some improvements have been proposed to the Classic Hill Cipher to overcome these problems. In the other hand, encryption sometimes not enough to keep the contents of information secret, the attackers can guess that there is confidential information passing on from the source to the destination and can be broken with enough computational power. One solution to this problem is through the use of steganography. Steganography is a technique of concealment information in digital media. In contrast to cryptography, it is not to keep others from knowing the hidden information but in a way that does not allow an attacker to even detect that there is secret message. Two of the popular methods are combined DWT and DCT in the proposed security system for concealing and extraction the secret data in order to compensate the drawbacks of both of them and to make the hidden information much more secure against the attacks. The proposed protection system goal is giving more the complexity to the encryption and to make the concealment capacity as high as possible with high visible quality.

MATERIALS AND METHODS

Proposed system: The proposed system has been consist of three phases, improved encryption/decryption method, improved concealment/extracting method and finally combining the two methods. This system has been implemented in MATLAB 9.

Encryption and decryption phase: The proposed method is improvement to the Classic Hill Cipher (Hill, 1929). In this proposed method the disadvantages of Classic Hill Cipher algorithm have been addressed, the encryption comparatively became more secure and it extended to include special characters and digits by using the extended ASCII table to replace characters, special characters and digits. The key has become random and variable size based on the password and the value which became entered by a user. The Invertible-Matrix problem has been solved also, where key matrix should have an inverse (Invertible-Matrix) that is required in the decryption process which accounted as the disadvantage of the Classic Hill Cipher. Repeat entering the encryption key for N of times with testing processes to check if the key matrix invertible or non-invertible using long procedures, this may take long time which is making the method not efficient. Therefore and to solve this problem the proposed method suggest replacing summation instead of multiplication in the encryption process.
The steps of the encryption will be as follow:

**Input:** Template data/password

**Output:** Encrypted data vector

**Step 1:** Convert the secret data into a vector of digits using extended ASCII

**Step 2:** Calculate the length (L) of the entered password

**Step 3:** Generate random key (K) from the entered password: Pseudo Random Generator (PRG) function has been used to generate a random key which generate different key each time that create different key in the decryption process, to solve this problem the default seed has been reset programmatically before using it that will help to having same key in the decryption process. Then, apply Pseudo Random Generator (PRG):

\[
K = PRG(\text{L}, S)
\]  

(1)

where, S is the password entered by the user, L is length of the entered password

**Step 4:** Arrange Key vector in 2D matrix:

\[
K = \begin{bmatrix}
    k_{11} & k_{12} & \cdots & k_{1m} \\
    k_{21} & k_{22} & \cdots & k_{2m} \\
    \vdots & \vdots & \ddots & \vdots \\
    k_{m1} & k_{m2} & \cdots & k_{mm}
\end{bmatrix}
\]  

(2)

**Step 5:** Calculate the length the message vector P and if the length is not a multiple of L then add some 0s to the vector as follow:

\[
R = \text{(Remainder P/L)} \text{... If (R+0)}
\]  

(3)

Then:

\[
P(n+1|n/L)|L+1)
\]  

(4)

**Step 6:** Segment the message vector into N of blocks:

\[
P(i) = \{P(1:L), P(L+1:2L), \ldots, P([n/L]\ell L+1)\}
\]  

(5)

**Step 7:** Process each block matrix P(i) individually. Where:

\[i = 1, 2, \ldots, N\]

**Step 8:** Arrange each Block P(i) into 2D square matrix

\[
P(i) = \begin{bmatrix}
    P(i)_{11} & P(i)_{12} & \cdots & P(i)_{1m} \\
    P(i)_{21} & P(i)_{22} & \cdots & P(i)_{2m} \\
    \vdots & \vdots & \ddots & \vdots \\
    P(i)_{m1} & P(i)_{m2} & \cdots & P(i)_{mm}
\end{bmatrix}
\]  

(6)

**Step 9:** Perform the encryption equation:

\[
C(i) = [P(i)+K] \mod 255
\]  

(7)

**Step 10:** Convert the encrypted the square block matrix C(i) into 1D matrix

**Step 11:** Repeat steps 6 and 9 for N times, to complete all segments

**Step 12:** After encrypt all segments, recombine the segments and remove the added “0”s to get the encrypted vector:

\[
C = C \{1: \text{length}(p)\}
\]  

(8)

**Proposed algorithm:** Encryption algorithm shown in Fig. 1, where, P is Plaintext, P(i) is block of plaintext, C is ciphertext, C(i) is block of Ciphertext, K is key, L is length of encryption key, N is number of blocks and S is password.

The steps of the decryption will be as follow:

The decryption algorithm is basically follows the reverse processes of the encrypting steps to obtain the plaintext. After extracting the ciphertext bits stream from the cover image, decryption processes will begin as follows:

**Input:** Encrypted data vector/password/length of key

**Output:** Secret message

**Step 1:** First process of decryption is convert encrypted data into Vector (P) of numbers using Extended ASCII codes

**Step 2:** Generate the random key from the entered password by applying Pseudo Random Generator (PRG), as in the encryption process the default seed is first reset to get same random key

**Step 3:** Arrange Key vector in 2D matrix:

\[
K = \begin{bmatrix}
    k_{11} & k_{12} & \cdots & k_{1m} \\
    k_{21} & k_{22} & \cdots & k_{2m} \\
    \vdots & \vdots & \ddots & \vdots \\
    k_{m1} & k_{m2} & \cdots & k_{mm}
\end{bmatrix}
\]  

(9)

**Step 4:** Segment the encrypted message vector into N of blocks:

\[
C(i) = \{C(1:L), C(L+1:2L), \ldots, C([n/L]\ell L+1)\}
\]  

(10)

**Step 5:** Arrange each Block P(i) into 2D square matrix:

\[
P(i) = \begin{bmatrix}
    C(i)_{11} & C(i)_{12} & \cdots & C(i)_{1m} \\
    C(i)_{21} & C(i)_{22} & \cdots & C(i)_{2m} \\
    \vdots & \vdots & \ddots & \vdots \\
    C(i)_{m1} & C(i)_{m2} & \cdots & C(i)_{mm}
\end{bmatrix}
\]  

(11)

**Step 6:** Process each block matrix C(i) individually. Where, \[i = 1, 2, \ldots, N\]

**Step 7:** Perform the decryption equation:

\[
P(i) = [C(i)-K] \mod 255
\]  

(12)
**Fig. 1:** Encryption process

**Step 8:** Convert the encrypted the square block matrix $P(i)$ into 1D matrix

**Step 9:** Repeat steps 6 and 8 for $N$ times, to complete all segments

**Step 10:** After decrypt all segments, recombine the segment to get the original vector:

$$P = P(1: N) \quad (13)$$

**Step 11:** Replace the numbers of the vector with the extended ASCII table to get original template data

**Proposed algorithm:** Decryption algorithm shown in Fig. 2, where $P$ is Plaintext, $P(i)$ is block of plaintext, $C$ is ciphertext, $C(i)$ is block of Ciphertext, $K$ is key, $L$ is length of encryption key, $N$ is Number of blocks and $S$ = Password.

**Concealment and extracting phase:** In the proposed technique, information has been embedding through reparation technique in frequency domain based on combination of two transforms Discrete Wavelet Transform (DWT) and Discrete Cosine Transform (DCT). Frequency domain steganography is very secure and more complex way of hiding a secret inside an image comes with the use and modifications of discrete cosine transformations, it is more secure than spatial domain steganography because information can be spread out to entire image. Embedding process is started by applying 1-level 2D Haar DWT to the host image (Fig. 3), then perform the block base DCT with selecting DWT coefficient sets (HL, LH and HH) and embed data in middle frequencies in each sub-band depending in percentages entered by the user. Embed data in the sub-bands (HL, LH and HH) is better in perspective of security and quality. Besides, distribute data in these sub-bands depending on the percentages and nonspecific number of bits in each pixel, this will provide confidentiality for embedded data.

**Concealment procedure:** The data embedding processes is represented in Fig. 4, followed by the steps of proposed technique:
Fig. 2: Decryption process

Fig. 3(a-c): 1-level 2D Haar DWT, (a) Original image, (b) After haar image and (c) Haar DWD sub-bands

**Input:** Cover image, encrypted message, selected percentages

**Output:** Stego image

**Step 1:** Perform DWT on the host image to decompose it into four non-overlapping multi-resolution coefficient sets: LL, HL, LH and HH

**Step 2:** Perform DCT on the chosen coefficient sets (HL, LH and HH). These coefficients sets are chosen to inquire both of security and efficiency of the proposed algorithm

**Step 3:** Compute the length of each message will be embed in each sub-band based on the input percentages. Number of bits of each pixel of sub-bands which will be used to embedding data message will be agreed between sender and receiver. If the entered message is greater than the maximum capacity of the three sub-bands, the program will return an error message to the user to change the cover image

**Step 4:** Convert the secret message and the cover matrix into binary vector
**Fig. 4: Concealment process**

**Step 5:** Embedding process is done by insert the message vector bits in LSBs of the DCT coefficients of the cover vector, numbers of bits which will be used for embedding are equal for each pixel of cover vector. The process of embedding in sub-bands matrix will be in this sequence (HL, LH then HH)

**Step 6:** Convert the resultant modified cover vector for each sub-band from binary to decimal vector. Then, Apply inverse DCT (IDCT) to each sub-band after its mid-band coefficients have been modified to embed the message bits as described in the previous step.

**Step 7:** Apply the inverse DWT (IDWT) on the DWT transformed image, including the modified sub-band, to produce the stego image.

**Fig. 5: Extracting process**

**Step 2:** Perform DCT on the chosen coefficient sets (HL, LH and HH)

**Step 3:** Extract the total length of embedded message from the first twenty bits of the coefficient sets then compute the length of each embedded message based on input percentages.

**Step 4:** Extract the embedded data vector bits from the LSBs of the coefficient sets (HL, LH and HH)

**Step 5:** Reconstruct the scrambled cipher message using the extracted data bits.

**Combining encryption and concealment methods phase:**

The proposed system consists of combining the two aforementioned methods which were enhanced to be more efficient and more secure, where encrypting data using the enhanced Hill Cipher and then concealing the encrypted data using proposed concealment technique will produce an image can transmitted without revealing that secret template information is being exchanged. Moreover, even if the attacker were to defeat the concealment technique and detect the information from the stego-object, there is need for a key to decipher the encrypted information, Fig. 6.
Fig. 6: Proposed secure system

RESULTS

Analysis of encryption: In the proposed method a block cipher have been improved by modifying the Hill Cipher. Several images have been used standard and nonstandard images like Lena, Baboon, Einstein etc. and investigate text messages in different sizes.

Encryption results
Encryption efficiency analysis: The performances of the experimental results have been evaluated by Normalized Cross-correlation (NC) measurement (Wang and Cheng, 2005) as follows:

\[ NC(M, M') = \frac{\sum_{k=1}^{N} M(k) M'(k)}{\sqrt{\sum_{k=1}^{N} M(k)^2} \sqrt{\sum_{k=1}^{N} M'(k)^2}} \]  

(14)

where, M is original secret message, M' is extracted secret message.

The “normalized cross correlation” is a measure of the dissimilarity between the original and encrypted message, the range of NC metric values is between 0 (dissimilar) and 1 (similar). Table 1 shows the NC of the encrypted images and compare the result with the original Hill Cipher algorithm, decrease in NC value indicates difference between the original and the encrypted images.

Figure 7 sample of image, the resulting images showing the change in images after encryption procedure for both original and improved Hill Cipher algorithm. The image sample shows that the proposed algorithm is more effective in encryption quality than the original Hill Cipher.

Sample of text message before and after encryption are shown in Fig. 8 which performed changing the text after encryption.

Running time analysis: For comparing running time with the original Hill Cipher been programmed these algorithms by the programming language MATLAB 9 on PC computer with Windows 7 Ultimate, processor: Genuine intel(R) CPU 585 at 2.16 GHz and Memory (RAM): 1.00 GB. The proposed system has been implemented on the text data of various sizes time calculated at the time encryption and decryption. Figure 9 shows the comparison result for encryption and decryption running time between the original and improved Hill Cipher, the results show running time less than the original method especially when the secret message increase.

Concealment results: The Peak Signal to Noise Ratio (PSNR) is introduced (Lu and Liao, 2001; Peng et al., 2004) to evaluate the performance of the proposed scheme and image quality which is defined as:

\[ PSNR = 10 \log_{10} \left( \frac{255^2}{MSE} \right) \text{dB} \]  

(15)

\[ MSE = \sum_{i=1}^{w} \sum_{j=1}^{h} (a_{ij} - b_{ij})^2 / w \times h \]  

(16)

where, w \times h is the image size, a_{ij} and b_{ij} are the corresponding pixel values of two images.

The PSNR is often expressed on a logarithmic scale in decibels (dB). A larger PSNR value means stego-image preserves the original image quality better. In general, the distortion of the stego image that caused by the embedding can be obvious when the PSNR values falling below 30 dB while the stego image considered high quality when PSNR value is 40 dB and above (Cheddad et al., 2008).

Table 1: Results comparison between original and improved Hill algorithm.

<table>
<thead>
<tr>
<th>Images</th>
<th>Original hill</th>
<th>Improved hill</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lena</td>
<td>0.3272</td>
<td>0.0138</td>
</tr>
<tr>
<td>Flower</td>
<td>0.1832</td>
<td>0.0083</td>
</tr>
<tr>
<td>Baboon</td>
<td>0.3198</td>
<td>0.0091</td>
</tr>
<tr>
<td>Einstein</td>
<td>0.2199</td>
<td>0.0076</td>
</tr>
</tbody>
</table>
The proposed method has been compared to other methods. One of the common methods uses the region of the image for embedded the template data (Kharje et al., 2013; Amritha and Meethu, 2013). Other slandered method is the adaptive LSB method (ADLSB), it is enhance to the classic LSB. As well, proposed method has been compared with a method uses the same proposed technique but with using DWT transform only. The comparison is shown in Table 2.

Table 2: Results of the comparison with the standard LSB and the other methods

<table>
<thead>
<tr>
<th>Images</th>
<th>Size (bit)</th>
<th>PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Adaptive LSB</td>
</tr>
<tr>
<td>Image1</td>
<td>776420</td>
<td>38.13</td>
</tr>
<tr>
<td>Image2</td>
<td>747850</td>
<td>38.29</td>
</tr>
<tr>
<td>Image3</td>
<td>780250</td>
<td>38.11</td>
</tr>
<tr>
<td>Image4</td>
<td>770558</td>
<td>38.16</td>
</tr>
</tbody>
</table>

**Fig. 7(a-c):** Comparing between original and improved Hill Cipher encryption effects, (a) Lena, (b) Original hill and (c) Improved hill

**Fig. 8(a-b):** Sample of text message before and after encryption
Fig. 9(a-b): Running time comparison between original and improved Hill Cipher algorithm, (a) Encryption running times and (b) Decryption running times

Table 3: Capacity and quality of images using different percentages

<table>
<thead>
<tr>
<th>Percentages (HL, LH, HH)</th>
<th>PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lena</td>
</tr>
<tr>
<td>10, 45, 45</td>
<td>52.46</td>
</tr>
<tr>
<td>40, 20, 40</td>
<td>59.40</td>
</tr>
<tr>
<td>35, 30, 35</td>
<td>57.38</td>
</tr>
<tr>
<td>45, 45, 10</td>
<td>51.94</td>
</tr>
<tr>
<td>25, 25, 50</td>
<td>56.38</td>
</tr>
<tr>
<td>20, 20, 60</td>
<td>57.66</td>
</tr>
</tbody>
</table>

Table 3 explains the effect of changing the percentages to stego image quality using different images (size of images are 256×256 with 15% data capacity).

CONCLUSION

The results showed the efficiency of the proposed security system, encryption is become very secure cryptography technique through enhanced the Hill Cipher algorithm by generate the key matrix randomly and dynamically using the entered password, also by extended it to include special characters and digits and the concealment method which combine two transform methods DWT and DCT with dynamic data distribution are highly secured. Through combine the two above techniques, there is less chance for the original message to be detected by the attacker.

REFERENCES


