Secure Data Hiding in Coded Image using Arduino Speech Recognition

S N Abd -Alwahab1, M K Wali2, H H Hasan3
1 Department of computer Engineering, College of Technical Electrical Engineering, Middle Technical University, Baghdad, Iraq
2 Department of Electronic Engineering, College of Technical Electrical Engineering, Middle Technical University, Baghdad, Iraq
3 Department of Electrical Power Engineering, College of Technical Electrical Engineering, Middle Technical University, Baghdad, Iraq
1 shamsnaseer3a@gmail.com , 2 musa.wali@mtu.edu.iq , 3 husam@eetc.mtu.edu.iq

Abstract. The cover medium is used as an envelope to hide inside information. Embedded data may represent a text, image or video to be transferred in the spread media. In this paper, secret information hiding in coded image is exhibited relying upon quantization level and discrete wavelet transform (DWT). Secret data is encrypted and controlled by Arduino speech recognition board in both transmitter ad receiver side to decrypt the secret data. The color image is utilized as a cover image with the discrete wavelet transform it into another form and Absolute Moment Block Truncation Coding (AMBTC) technique is used for embedding the data. The mapping of secret data and image scrambling is depended on the Henon map, while the logistic map is used to encrypt secret data. The proposed method achieved accurate results in peak signal to noise ratio about 34.0524 at average for tested image.

Keyword: Discrete Wavelet Transform (DWT), Speech Hiding, Henon Map.

1. Introduction
Steganography is the art of hiding data that is represented by text, image or video inside another document that is called cover medium. Data hiding is used for secret communication when secured channels are not available [1]. The size of embedding data is related to the cover medium and hiding algorithm, while the robustness of retrieval hidden data is related to the ability to extract them [2]. All information hiding techniques are used in different information technologies and communications (ITC), some of them are used in communications such as military and intelligence applications and authentication data as original or fake. In addition to identification and proof of ownership for commercial purposes using watermarking algorithms, and the applications of customer tracing in which the IDs of the users are inserted in the products before selling [3].

In this paper, a modern method in steganography of encrypted text files in image is presented with details. There is a simple and efficient image compression technique called Block Truncation Coding (BTC) that is represented by two quantization levels \((a_i, b_i)\). A bitmap \((B_i)\) is applied through the compressing code of an image block \(C_i\) by using an AMBTC that is used for data hiding [4]. The processes of quantization level modification and bitmap representation are used for hiding secret information with a discrete wavelet transform (DWT). In general, the utilization of such processes provide a low cost of computational and a highly acceptable level of image quality. To increase the complexity, a specific mapping for hiding data may be used and the coded image may be scrambled. The chaotic system is used for mapping and scrambling secret information and cover medium. There
are several chaotic systems in multiple dimensions that are used for this purpose [5]. Logistic map and Henon map are types of chaotic systems that are sensitive to the initial conditions which make the changing of the whole system is depending on small changing on initial parameters [6]. The contribution of this work are produced in multiple way and satisfied security requirement in privacy and confidentiality. The Arduino with speech recognition verified and fixed destination recipient, while, the software speech recognition satisfied only the recipient is used the secret data. The DWT increased complexity and capacity, furthermore reduced distortion in cover medium. The scrambling, encryption and mapping increased also complexity of retrieval secret data. Section (2) presents chaotic map, section (3) presents materials and methods, while results and discussion are presented in section (4), and section (5) presents the conclusion.

2. Chaotic Map
Chaotic map (evolution function) demonstrates the sort of chaotic behaviour in mathematics. It can be used for continuous-time or discrete-time depending on its type. Maps may be parameterized by the parameter and the separated maps generally take this method of repeated functions. Chaotic maps are commonly used in the research of dynamic systems [7]. Two types of chaotic map are used in proposal; logistic map and Henon map. Firstly, the logistic map is a polynomial mapping, often is cited as an archetypal example of how complex, chaotic behaviour can arise from very simple nonlinear dynamical equations [8]. Mathematically, the logistic map is written as in equation (1):

\[ x(n + 1) = X(n) \times r \times (1 - x(n)) \]  

where \( x(n) \) is a number between zero and one that represents the ratio of the existing population to the maximum possible population and \( R \) (sometimes also denoted by \( y \)) is a control parameter and the interest value for it is within the interval (0,4). The relative simplicity of the logistic map makes it a widely used point of entry into a consideration of the concept of chaos. A rough description of chaos is that the chaotic systems exhibit a great sensitivity to initial conditions. A common source of such sensitivity to initial conditions is that the map represents a repeated folding and stretching of the space on which it is defined. Secondly, Henon Map was presented by Michel Henon as a rearranged model of the Poincare area of the Lorenz display. For the established map, an underlying purpose of the plane will approach a lot of focuses known as the Henon peculiar attractor. The Henon Attractor was a fractal, smooth in one heading and a Cantor set in other. Numerical assessments produce a connection measurement of 1.25 ± 0.02 [9], and a Hausdorff measurement of 1.261 ± 0.003 for the attractor of the traditional map. Chaotic values can be generated from applying equations (2) and (3) which are iterated several times to generate the required elements.

\[
\begin{align*}
\text{(2)} & \quad x(n + 1) = 1 - a \times x(n)^2 + y(n) \\
\text{(3)} & \quad y(n + 1) = b \times x(n)
\end{align*}
\]

The values for the constants ‘\( a = 1.76 \)’, ‘\( b = 0.1 \)’, \( x \) and \( y \) are the iterative sequence of output values to get a random sequence. The Henon Map is used for the key generation with an initial value that generate real numbers in period of (0, 1).

3. Materials and methods
The proposed method deals with hiding secret information in the cover medium (image) in the first stage (hiding stage), and extracting this secret information in the second stage (extraction stage). In the hiding method, the AMBTC method has been utilized, also, Arduino speech recognition has been used for matching the sender name, and if it is failed, no message will be sent. Three keys sequences are used for mapping, scrambling, and encryption, first and second keys are generated by Henon map, while the third key is generated by a logistic map. The embedding algorithm used the Henon map as a mapping key (not sequential). The extraction method depends on speech recognition for receiver authentication by using specific initial values for key generation (for each user) same as that used in the transmitter side.
3.1 Hiding Stage

The coding technology uses the AMBTC method which depends on pixel smoothing in each block. In this work, the utilized cover media is an image. DWT is used to transform the cover image by separating high-value rounding coefficients used to hide information from details of other low-value coefficients. The proposed method in this paper consists of two stages, the first one is the forward stage, and the second one is the backward stage. The first stage is the hiding method that needs to insert the secret text into cover media. The cover image is divided into three channels; red, green, and blue (RGB) and go through series of stages to produce stego-image as shown in figure 1, and the pseudo code of hiding algorithm is shown in algorithm 1.

![Figure 1. Block diagram of the proposed hiding method.](image-url)
3.1.1. Embedding Method

The blocks that represent by \((m \times m)\) are used for the embedding process, and for each block denoted by \(C_i\), the average pixel value in \(C_i\) denoted by \(\bar{C}_i\) is computed. The bitmap \(B_i\) can be constructed and the lower quantization level \(a_i\) is gotten by rounding the mean value of the pixel in \(C_i\) with values less than the value of \(\bar{C}_i\), similarly with \(b_i\). To decode the AMBTC trio \((a_i, b_i, B_i)\), zero’s values in \(B_i\) are decoded by \(a_i\), and one’s values in \(B_i\) are decoded by \(b_i\), consequently will decode the image \(\tilde{C}_i\). The process of creating secret text is accomplished by converting input text to ASCII code at first. After that, this code is converted to binary form for embedding it in cover band.

The Quantization level Modification (QM) technique embeds data by replacing the bitmap \(B_i\) with secret data \(S_i\) and modifies the corresponding quantization levels \((a_i, b_i, B_i)\) to \((\hat{a}_i, \hat{b}_i, \hat{B}_i)\) such that the distortion between the original and stego AMBTC blocks is minimal.

Scan each trio in \({a_i, b_i, B_i}\) \(i=1\), the QM technique is employed to embed \(m \times m\) secret bits, and to calculate the best quantization levels \((\hat{a}_i, \hat{b}_i)\) that minimize the distortion, let \(B_i = \{\beta_{ij}\}_{j=1}^{m \times m}\) be the bitmap of \(i-th\) block, and \(S_i = \{s_{ij}\}_{j=1}^{m \times m}\) be the secret data to be embedded. Since the embedding is performed by replacing \(B_i\) with \(S_i\), the distortion occurs at \(S_{ij} \neq \beta_{ij}\). The \(\{\alpha_i^c, \delta_i, B_i^c\}_{i=1}^{N}\) refers to the embedded trio. The repeating is kept until all the secret data are embedded.

In this proposed method, each band of the color image is used as a grayscale image and could be used for hiding information. The image is transformed into one level by Haar DWT as shown in figure 2, to produce four bands; High-High (HH), Low-High (LH), High-Low (HL), and Low-Low (LL), bands. The Low-Low band includes the significant value of image information. Therefore, embedding may be used in any of the other bands. All text hiding in the cover medium is encrypted using the logistic map while all coddled image parameters are scrambled using Henon map.

**Algorithm 1: Hiding Algorithm**

**Input:** Speech, SecretData, CoverImage, InitSe(initial scramble), InitM(initial mapping)

**Output:** Stego image

**Steps:**
1. Im ← ReadInputImage
2. InputSP ← Match(InputSpeech)
3. GetInitialParameter(Sp)
4. Key1 ← KeyGenerationLogisticMap(InpL)
5. CipherText ← Encryption(SecretData, Key1)
6. Key2 ← KeyGenerationHenon(InitM)
   For each block in cover image do
7. PBs ← PartitionBlocks(CoverImage)
8. For i ← 1 to No of Blocks
   Current (i) ← PBs(key2(i))
   \([a(i), b(i), B(i)] ← EmbeddingMethod(current(i))\)
   End for
9. Key3 ← KeyGenerationHenon(InitSe)
10. For i ← 1 to length(PBs)
    Coddedai(i) ← ai(key3(i))
    Coddedbi(i) ← bi(key3(i))
    CoddedBi(i) ← Bi(key3(i))
    End for
11. Return Coddedai, Coddedbi and CoddedBi
The wavelet transform is an influential technique in the examination of transient occurrences for the reason that it has the capability to extract time and frequency data from the transient signal. DWT is used to develop the statistics set via [10]:

\[
\psi(c,d) = \frac{1}{\sqrt{c}} \int_{-\infty}^{\infty} f(t) \psi \left( \frac{t-d}{c} \right) dt
\]

where \(c\) and \(d\) are the constants of dilation (scale) and translation and these constants are continuous in nature.

### 3.1.2. Arduino Speech Recognition

Voice Recognition Module is a compact and easy-control speaking recognition board. This product is a speaker-dependent voice recognition module. It supports up to 80 voice commands in all. Max 6 voice commands could work at the same time. Any sound could be trained as a command. Users need to train the module first before letting it recognizing any voice command. This board has 2 controlling ways: Serial Port (full function), General Input Pins (part of function). General Output Pins on the board could generate several kinds of waves while corresponding voice command was recognized.

### 3.2. Extraction Stage

The extraction method using AMBTC retrieves the encrypted data that will be decrypted by using key generation after matching the features of receiver speech. The total steps of the extraction method are explained in figure 3, and the total steps of the extraction pseudo code are shown in algorithm 2.

**Figure 2. Image Discrete Wavelet Transforms**

**Figure 3. Block diagram of The Proposed Extraction Method**
3.1.1. Embedding Method

For extracting hidden data, the same procedure for embedding is applied in reverse order. Henon map is used in two ways; firstly, for descrambling, and secondly, for mapping hidden data. The method is summarized by generating a real number and multiplying it by power ten decimal number. As well as, the result will round and module to number N. The process of descrambling is done by performing the operation of exclusive-or between the received data and the generated key from the Henon map. The other process is used for mapping hidden data which is done by sorting the generated numbers and using the index of numbers as a map. The extracted data is dependent on the bitmap of each block in AMBTC.

3.2.2 Speech Recognition

Speech samples are pre-processed by framing and extracting Pitch and Mel Frequency Cepstral Coefficient’s (MFCC) features for each one [11] and [12]. The extracted features of voiced frames are kept while the unvoiced frame will be dropped. The remaining features are given to K-Nearest Neighbor (KNN) classifier to classify the sample by specifying the number of neighbours and the type of distance used in classification. Among the various methods of supervised statistical pattern recognition, the Nearest Neighbour rule achieves consistently high performance, without a priori assumptions about the distributions from which the training examples are drawn. A new sample is classified by calculating the distance to the nearest training case. The k-NN classifier extends this idea by taking the k nearest points and assigning the class of the majority [13].

3.2.3. Key Generation

The same method of key generation that is used in hiding stage will be a method of key generation in extraction stage by Henon Map. Each receiver has its initial values for Henon Map key generation.

Algorithm 2: Extraction Algorithm

**Input:** Codedai, Codedbi, CodedBi, StegoImage

**Output:** SecretData, StegoImage

**Steps:**

1. Key3 ← KeyGenerationHenon(InitS)
2. For i ← 1 to length(Codedai)
   - ai(Key3(i)) ← Codedai
   - bi(Key3(i)) ← Codedbi
   - Bi(Key3(i)) ← CodedBi
   End for
3. Key2 ← KeyGenerationHenon(InitM)
4. For i ← 1 to length(SecretData)
   - Sec(i) ← Bi(key2(i))
   - CurrentBlock(i) ← Contrast(ai(key2(i)), bi(key2(i)), bi(key2(i)))
   End for
5. Image ← Concatenate(CurrentBlocks)
6. Sp ← InputSpeech(RecordSpeech)
7. SpFeatures ← ExtractMFCC(Sp)
8. InitSDatabase ← MatchFeatures(SpFeatures)
9. Key1 ← KeyGenerationLogisticMap(InpL)
10. Text ← Xor(Sec, Key1)
11. end
3.2.4. Secret Decryption

The stream of bytes that is represented as a key generated in the previous step is used to decrypt the extracted data by doing Exclusive-OR between the extracted data and the steam of key generated using logistic map. The speech recognition is used for selecting the initial parameters that are used for key generation and only the authorized person can extract secret data.

4. Results and discussion

Experiments were applied on several images (in size 512×512) for testing the embedding method. The tested images are represented by Women, car, Lena, Peppers, and House. In portioning blocks, the size of the block is specified in 4×4 pixels that are used in AMBTC. Visually, the histogram of the original images and Stego-images be almost the same as shown in figure 4.

| Original Woman | Before Hiding | Stego-Woman | After Hiding |
|----------------|--------------|-------------|-------------|
| ![Woman](Original Woman) | ![Before Hiding](Before Hiding) | ![Stego-Woman](Stego-Woman) | ![After Hiding](After Hiding) |

| Original Car  | Before Hiding | Stego-Car  | After Hiding |
|---------------|--------------|------------|-------------|
| ![Car](Original Car) | ![Before Hiding](Before Hiding) | ![Stego-Car](Stego-Car) | ![After Hiding](After Hiding) |

| Original Lenna| Before Hiding | Stego-Lenna| After Hiding |
|---------------|--------------|------------|-------------|
| ![Lenna](Original Lenna) | ![Before Hiding](Before Hiding) | ![Stego-Lenna](Stego-Lenna) | ![After Hiding](After Hiding) |

| Original Peppers | Before Hiding | Stego-Peppers | After Hiding |
|------------------|--------------|--------------|-------------|
| ![Peppers](Original Peppers) | ![Before Hiding](Before Hiding) | ![Stego-Peppers](Stego-Peppers) | ![After Hiding](After Hiding) |
After embedding data in the cover medium (image), the quality of stego-images is tested with respect to original images. Several objectives measurement with full-reference is applied to find the effect of embedding data on them. The first test is Mean Squared Error (MSE), the second test is Peak Signal to Noise Ratio (PSNR) and the third test is structural similarity index (SSIM) [14] [15], and these three tests are explained in Table 1, respectively. All these tests are used with different sizes of hidden data such as; 100, 200, 300, 400, and 500 characters respectively. The secret data and the bitmap of each block were in binary form, so the degree of distortion is depending on the differences between them. The degree of similarity between embedding binary form of data and bitmap of each block led to less distortion and high quality.

The total results of three quality measurements (MSE, PSNR, and SSIM) for the tested images are shown in figures 5, 6, and 7 respectively. The best value of MSE is (11.5248) in hiding (300) characters inside (first) cover image (woman image), the best value of PSNR is (37.5145) in hiding (300) characters inside (first) cover image (women image), and the best value of SSIM is (0.9909) in hiding (300) characters inside (fourth) cover image (peppers image).
Figure 5. Mean square error of the encrypted text hiding method.

Figure 6. Peak signal to noise ratio of the encrypted text hiding method.

Figure 7. The Structural similarity of the encrypted text hiding method.
5. Conclusion
Arduino with speech recognition is used to control the submission of secret information to authorized persons by retrieving the same key generated in the sending side as multiples key stream based on word speech recognition. Secret data hiding is based on the AMBTC technique to embedded the data under the transformed cover medium by DWT. Henon map and the logistic map were used for complex the scrambling stego-image and the encrypting text. Quantization level modification works on increasing the size of embedding information by insert bits equal to bitmap array for each block. Thus, the proposed method may be utilized in authentication applications. The efficiency of the proposed embedding algorithm has been obtained by providing capacity requirements with accurate PSNR (34.0524). Furthermore, DWT is used as another embedded domain (not spatial domain) with minimum distortion. The combination of wavelet and AMBTC methods is used for information hiding. Recognition the sender's voice by Arduino board adds a physical security dimension to the work. The confidentiality requirement is satisfied by extraction secret data to be used only by the authorised person.

6. References

[1] E. Cole, 2002, Hiding in Plain Sight: Steganography and the Art of Covert Communication, Wiley
[2] M. Nosrati, R. Karimi and M. Hariri, 2011, "An introduction to steganography methods," World Applied Programming, vol. 3, pp. 191-195
[3] M. K. Mihqak, 2002, "Information Hiding Codes and their Applications to Images and Audio," in the Graduate College of the University of Illinois
[4] W. Hong, T. S. Chen, Z. Yin, B. Luo and Y. Ma, 2017, "Data hiding in AMBTC images using quantization level modification and perturbation technique," Multimedia tools and applications, vol. 76, pp. 3761-3782
[5] O. A. S. Alkishriwo, 2018, "An Image Encryption Algorithm Based on Chaotic Maps and Discrete Linear Chirp Transform," ALMADAR JOURNAL FOR COMMUNICATIONS, INFORMATION TECHNOLOGY, AND APPLICATIONS
[6] X. Zhang and Y. Cao, 2014, "A Novel Chaotic Map and an Improved Chaos-Based Image Encryption Scheme," The Scientific World Journal, vol. 2014
[7] N. A. Abbas and Z. H. Razaq, 2019, "Review of DCT and Chaotic Maps in Speech Scrambling," Journal of Theoretical and Applied Information Technology, vol. 97
[8] D. Wang, C. C. Chang, Y. Liu, G. Song and Y. Liu, 2015, "Digital Image Scrambling Algorithm Based on Chaotic Sequence and Decomposition and Recombination of Pixel Value," International Journal of Network Securit, vol. 17, pp. 322-327
[9] M. Prasad and K. L. Sudha, 2011, "Chaos Image Encryption using Pixel shuffling," CCSEA, pp. 169-179
[10] K. H. Talukder and K. Harada, 2007, "Haar Wavelet Based Approach for Image Compression and Quality Assessment of Compressed Image," International Journal of Applied Mathematics
[11] M. Tamazin, A. Gouda and M. Khedr, 2019, "Enhanced Automatic Speech Recognition System Based on Enhancing Power-Normalized Cepstral Coefficients," Applied Sciences, vol. 9, p. 2166
[12] S. G. Koolagudi, D. Rastogi and K. S. Rao, 2012, "Identification of Language using Mel Frequency Cepstral Coefficients (MFCC)," Procedia Engineering, vol. 38, pp. 3391-3398
[13] H. Gweon, M. Schonlau and S. H. Steiner, 2019, "The k conditional nearest neighbor algorithm for classification and class probability estimation," Peer J Computer Science, vol. 5, p. e194
[14] K. Mallikarjuna, K. S. Prasad and M. V. Subramanyam, 2016, "Image Compression and Reconstruction using Discrete Rajan Transform Based Spectral Sparsing," International Journal Image, Graphics and Signal Processing, vol. 8, pp. 59-67
[15] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, 2004, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE TRANSACTIONS ON IMAGE PROCESSING, vol. 13, pp. 600-612