EE 678 Application Assignment Final Report (Group 22) Audio-to-Image Wavelet transform based Audio Steganography

Akshay Sarode (110260010)
Ankur Mallick (110110013)
Krishnakant Saboo (11D070009)

April 5, 2015

1 Introduction

Online sharing and redistribution of digital media is very easy given the current technological advances. Hence there is an increased need to protect the copyright ownership of digital media. Illegal sharing is done by creating copies of digital media without the required copyright permissions and their redistribution. A way to secure the data and/or maintain information about its ownership is through encryption. Two broad techniques employed for this are Cryptography and Steganography. Cryptography aims to scramble the information sent and make it unreadable while steganography is used to conceal the information so that no one can sense its existence. Hence, steganography becomes one of the potential solutions to the problem of protecting digital media, for it allows content creators to embed data, such as author or copyright information, into a host signal. A number of algorithms for steganography in images, audio, videos have been proposed. In this assignment, we study and implement a scheme which hides information in an image representation of the audio signal, i.e., uses image steganography to hide and encrypt data inside the audio signal.[1, 2]

1.1 Image Steganography

We know that Cryptography is the science of hiding the meaning of information. A Cryptographic scheme is considered failed when an unauthorized person successfully decrypts the encrypted information [1].

Steganography, on the other hand, is the art and science of hiding secret data in plain sight without being noticed within an innocent cover data so that it can be securely transmitted over a network [3]. A steganographic scheme hides the secret data into innocent looking cover data. Different digital file formats can be used as cover data.

Nowadays, among all digital file formats, images are the easiest to find. Moreover, they have a higher degree of distortion tolerance as compared to other file formats. They also have a high hiding capacity due to the redundancy of digital information representation of an image data. As a result, image steganography has been explored more extensively as compared to audio steganography. Due to the relatively higher success in image steganography, we convert the audio data to an image and apply image steganographic techniques.

2 Steganographic Algorithm

2.1 Embedding procedure

We call the original audio signal in which we wish to embed the secret data as a_h. The embedding procedure of the audio steganographic algorithm is as follows : -

1. Conversion of the host audio signal a_h into the host image i_h using Discrete Wavelet Transform.

2. Let the secret data be d. This secret data is duplicated n times to get d^n. Embedding the secret data d^n into the host image i_h using the steganographic scheme in [2] to get the stego image i_s.

3. Conversion of the stego image i_s back into an audio file using the inverse audio-to-image transform to get the stego audio signal a_s.

It can be graphically explained as below : -

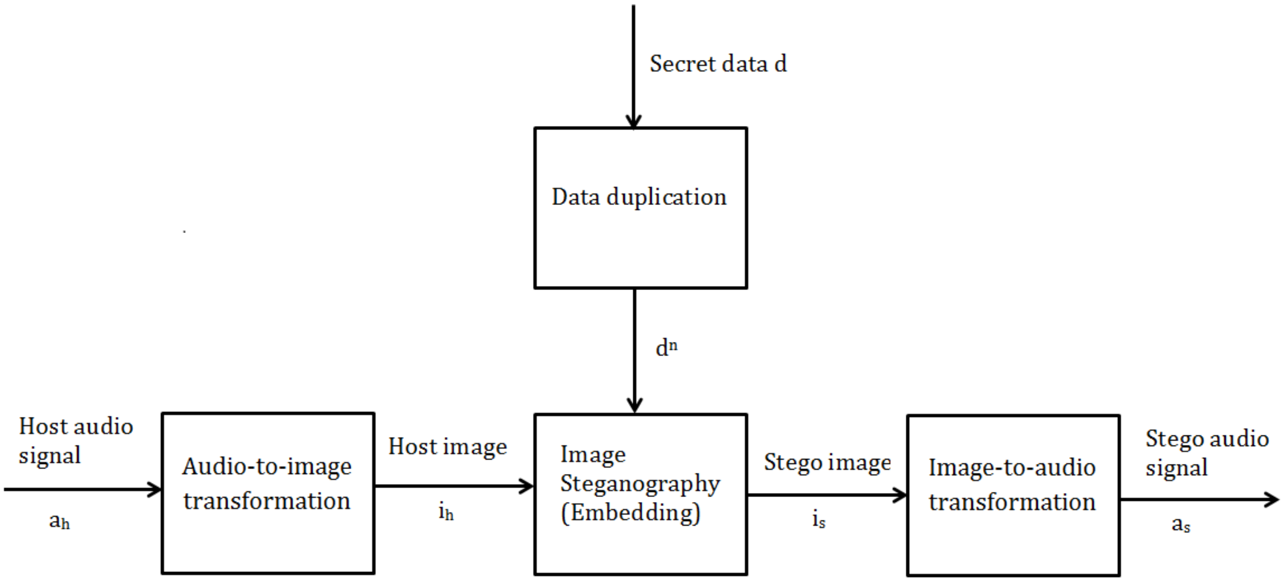


Figure 1: The embedding scheme

2.2 Secret keys & Extraction procedure

Let the signal sent to the extracting block be a'_s. To be able to extract the secret data from the stego audio input a'_s, the receiver needs the parameters used during the embedding scheme. These are the required secret keys. They include

(a) The specific wavelet used for audio signal decomposition

(b) The set of detail coefficients chosen to embed the secret data(cd_1, cd_2 or cd_3)

(c) The wavelet coefficient sampling interval Δ

(d) The height and width of the image r

(e) The seed used by the image steganography algorithm to select the blocks

(f) The threshold TH used by the image steganography algorithm

Once we have these keys, the extraction procedure of the audio steganographic algorithm is as follows : -

1. Conversion of the stego audio signal a'$_s$ into the stego image i$_s$ using Discrete Wavelet Transform.
2. Extraction of the secret duplicated data d'$_n$ from the stego image i$_s$.
3. Estimating the secret data d' using majority voting on d'$_n$.

It can be graphically explained as below : -

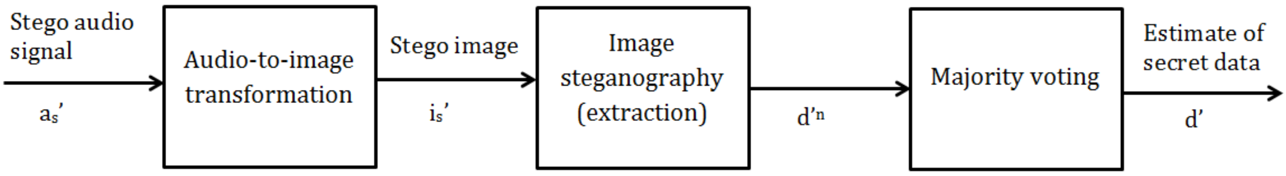


Figure 2: The extraction scheme

2.3 Details of the blocks used during embedding and extraction

The embedding and extraction procedures use the audio-to-image transform block, the image steganography block, and the image-to-audio transformation block. These are explained as follows : -

2.3.1 Audio-to-image transformation

First, the 3-level discrete wavelet decomposition of the audio signal a$_h$ is computed. This produces three sets of detail coefficients, at different resolution levels, denoted by ”cd$_n$”, where n = 1, 2, or 3. Also, we get one set of approximation coefficients denoted by ”ca”.
We choose one set of detail coefficients, downsample them by Δ, and convert them into an r x r image by reshaping them. Mathematically, we generate the r x r host image i$_h$ as follows : -

$$i_h(m,n) = cd_x(k); \quad 0 \le m,n \le r-1$$

$$k = (mr+n)\Delta; \quad 0 \le m,n \le r-1$$

2.3.2 Image Steganography (Embedding)

We use a pattern based image steganographic scheme as given in [2]. First of all, the host image (wavelet coefficients of the audio signal) are translated by t & scaled by f to lie in the range [0,255].
The r x r host image i$_h$ is the input to the image steganographic algorithm. Let S be the input secret data of bit length m.

$$i_h = \{\, i_h(i,j) \mid 0 \le i,j \le r-1 \,\}$$

$$S = \{\, s_k \mid 0 \le k \le m-1 \,\}$$

where

$$s_k \in \{0,1\}$$

DWT (Discrete Wavelet Transform) is used to decompose the cover image $\mathrm{i_h}$ into 3 scales. Let Y denote the transformation results of $\mathrm{i_h}$. We now have 1 smooth subband $\mathrm{Y_{s,3}}$ and 9 detailed subbands $\mathrm{Y_{k,n}}$ such that

$$Y_{\mathrm{k,n}} = \{\, y_{\mathrm{k,n}}(i,j) \mid k \in \{d,h,v\}, n = 1,2,3 \,\}$$

Now, we divide the wavelet coefficients of the DWT of $\mathrm{i_h}$ into non-overlapping 2x2 blocks. Each block contains 4 wavelet coefficients.

$$Y = \{\, B_i \mid 0 \leq i \leq n_{\text{blocks}} \,\}$$

We plan to hide 2 bits in each block. Hence, we need to select m/2 number of blocks. We do this by setting a secret seed for the random number generator, and then selecting m/2 number of blocks randomly. Note that we do not hide the data in the smooth subband $\mathrm{Y_{s,3}}$ as it contains the most important information of the cover image. Clearly, m/2 can't be greater than the number of blocks in the 3 detailed subbands.

We now use pattern-based classification to classify the selected m/2 blocks according to their coefficients and a threshold TH.

Suppose B = ($\mathrm{b_0}$, $\mathrm{b_1}$, $\mathrm{b_2}$, $\mathrm{b_3}$) is a block with four wavelet coefficients. The ”Pattern Matrix” T for the block B is constructed as follows : -

T = ($\mathrm{t_0}$, $\mathrm{t_1}$, $\mathrm{t_2}$, $\mathrm{t_3}$) where,

$$t_i = \begin{cases} 0 & \text{if } |b_i| < \mathrm{TH} \\ 1 & \text{otherwise} \end{cases}$$

We thus calculate the pattern matrix for each of the m/2 blocks. The pattern matrix has to be one of the 16 possibilities given in the table below : - We classify the pattern matrices into five classes by counting the number of 1's in them. Thus,

Class	Pattern matrix	Pattern value	Class	Pattern matrix	Pattern value
0	0000	(1,1)	2	1001	(0,0)
1	0001	(0,0)	2	1010	(0,1)
1	0010	(0,1)	2	1100	(1,0)
1	0100	(1,0)	3	0111	(0,0)
1	1000	(1,1)	3	1011	(0,1)
2	0011	(0,0)	3	1101	(1,0)
2	0101	(0,1)	3	1110	(1,1)
2	0110	(1,0)	4	1111	(1,1)

Table 1: All pattern matrices of a 2x2 wavelet block

$$\textit{Class of pattern matrix } T = \sum_{i=0}^{3} t_i$$

We now note that Class 0 and Class 4 have one pattern matrix each. There are four different pattern matrices in each of Class 1 and Class 3. Class 2 has six different pattern matrices.

We group the pattern matrices into four groups of four pattern matrices each as follows : -

- Group 1 : The 4 pattern matrices in Class 1
- Group 2 : The 4 pattern matrices in Class 3
- Group 3 : Class 0 & the 3 pattern matrices in Class 2 with leading bit 0

- Group 4 : Class 4 & the 3 pattern matrices in Class 2 with leading bit 1

Note that the Hamming distance between any two matrices in the same group is always 2.
We assign the four pattern values to the four pattern matrices in a group. The pattern value belongs to {(0,0),(0,1),(1,0),(1,1)}.
Now, we encode 2 bits in each 2x2 block as follows : Let (m_0,m_1) be the two bits which are to be embedded in to block B. Block B is modified to B' such that the pattern matrices of B and B' belong to the same group, and the pattern value of B' is (m_0,m_1). This distortion of B to B' has to be as small as possible. If pattern matrix of B belongs to Class 0 or Class 4, we need to modify just one of its four elements to get B'. If B belongs to other classes, we just need to exchange two of the original coefficients to get B'. Thus, the number of modifications in B is never greater than two.
Let the modified wavelet coefficient matrix be Y'. Y' is now transformed back to the spatial domain using IDWT (Inverse Discrete Wavelet Transform).

2.3.3 Image-to-audio transformation

Once we have the stego image i_s, the wavelet coefficients are rescaled by f and back translated by t to get the original wavelet coefficients. The stego image i_s is rearranged to form the audio signal's DWT coefficients as follows : -

$$cd_x(k) = \begin{cases} i(m,n) & \text{if } k = (mr+n)\Delta;\ 0 \leq m,n \leq r-1 \\ cd_x(k) & \text{otherwise} \end{cases}$$

Finally the 3-level discrete wavelet reconstruction is computed from cd_x, cd_n(n≠x), and cs to produce the stego audio signal a_s.

2.3.4 Image Steganography (Extraction)

The extraction procedure is same as the embedding procedure upto the selection of the $\mathrm{n}_{\mathrm{blocks}}$ number of blocks. Note that the random number generator used to select the blocks is set to the same seed as that used during the embedding process.
Once the blocks are selected, the pattern values of the blocks are the encrypted data. Thus, the hidden data is revealed.

3 Experimental results

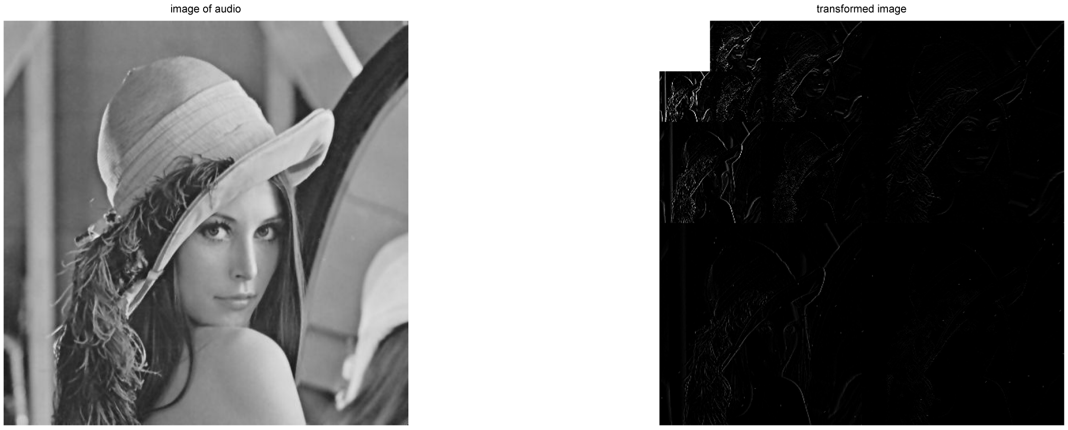
During the implementation of the algorithm, we choose Δ=2 and the maximum possible r for this delta. We then make this r a multiple of 16 by subtracting $mod(r, 16)$ from it. This is done in order to make the implementation more convenient. Let L be the length of the wavelet coefficient sequence for the audio signal. Thus,

$$\Delta = 2$$

$$r = \sqrt{\frac{L}{2\Delta}} - r\ modulo\ 16$$

The image steganography algorithm uses the transform domain image to encrypt messages into it. Shown below is the image and its transform domain image with 3 level decomposition 3. The high frequency regions clearly show edges in the horizontal, vertical and diagonal edges.

Audio file is converted into images because image steganography techniques are very well studied. The image of a 10 seconds clip from the audio is shown 4.



(a)

(b)

Figure 3: Image in spatial and transform domain

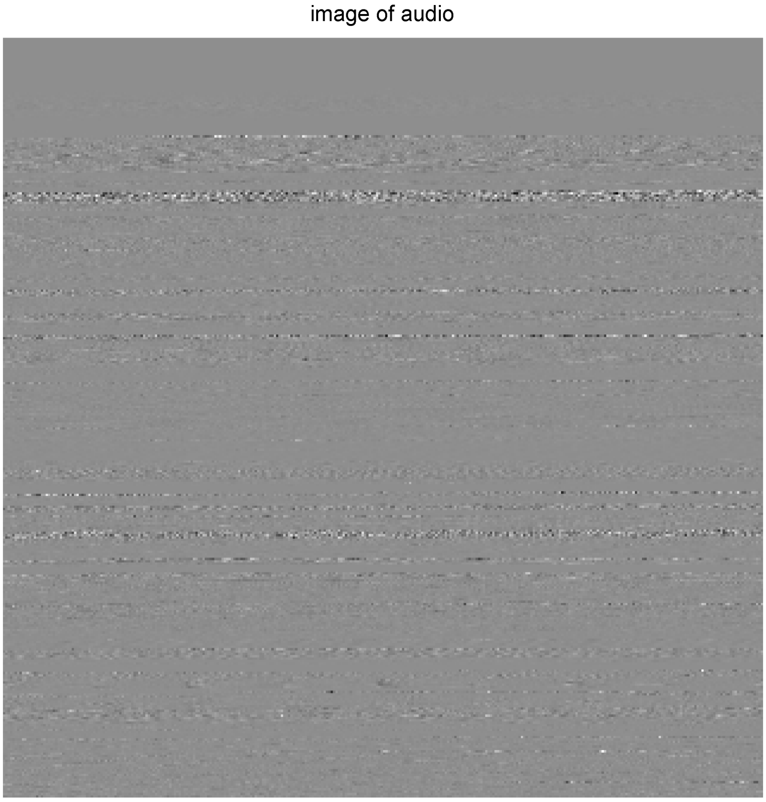


Figure 4: Image of a 10 seconds sound clip that is used in further experiments. There is no clear pattern in the image as in expected.

We gave as input a song of 40 seconds. The song was of the rock genre. It had a sampling rate of 44.1 kHz. We encrypted messages ranging from 200 bits to 1600 bits. The PSNR of the resulting audio file were calculated. The results can be seen in 5. The PSNR comes down as the number of bits increase. But there is not much change in the audio quality as perceived by the ear. However, if the number of bits is increased to 2000, there is a slight hum in the background of the track. Bit error is 0 in that case too.

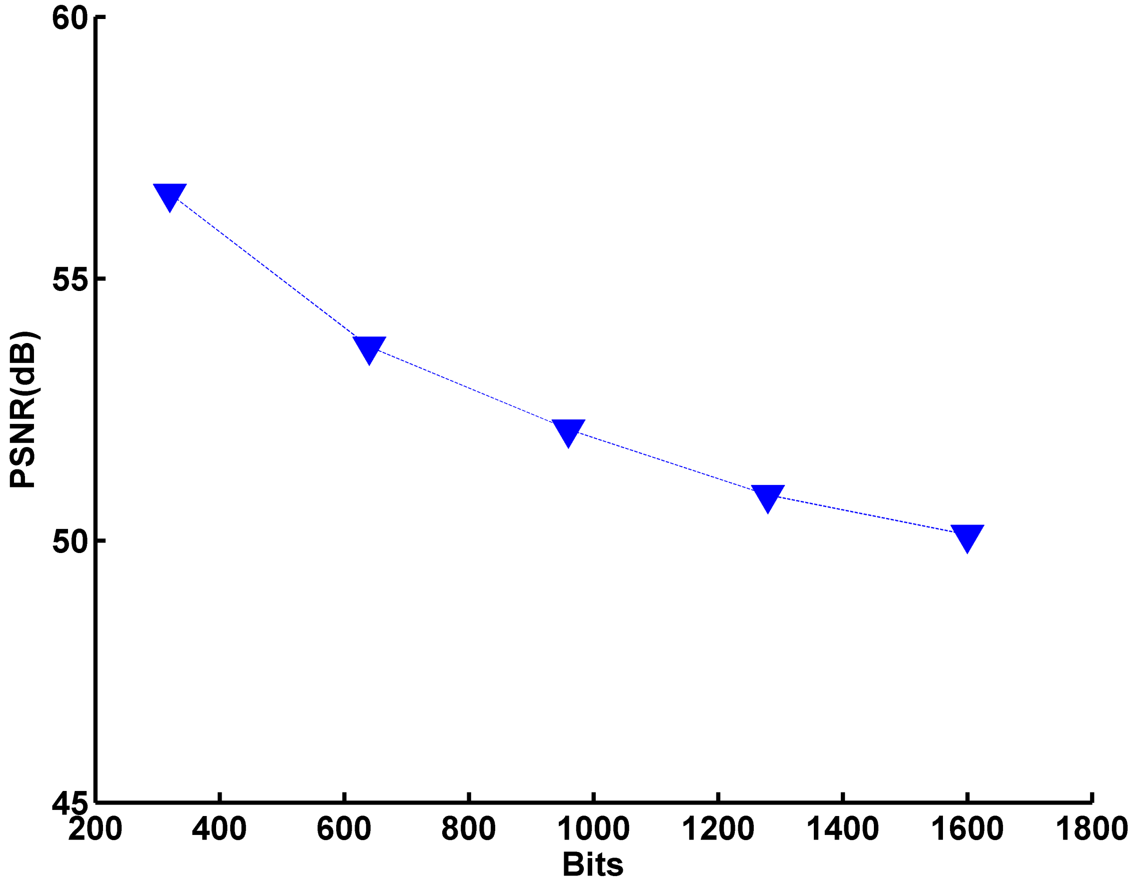


Figure 5: PSNR v/s number of bits in the in the message. The bit error in decoding for these message lengths was 0.

4 Future work

The scheme can be experimented with by trying different wavelet transforms. While converting the wavelet coefficients to an image, random sampling with a secret seed can be used to make this scheme more robust.

5 Conclusion

We studied techniques for image steganography like LSB method and pattern-based steganography. Their appilications in cryptography were also studied. Implementation of an audio-image based method was done. The robustness of the method as a function of message length was studied. The method gives high PSNR for upto 2000 bits and 0 bit error. However, the audio quality as perceived by the ear degrades drastically after a point, clearly beating the aim of steganogrpahy. The scheme however can be used for cryptography purposes since it provides a various tuning parameters which can be used as a secret key.

References

[1] Santosa, Rully Adrian, and Paul Bao. "Audio-to-image wavelet transform based audio steganography." ELMAR, 2005. 47th International Symposium. IEEE, 2005.

[2] C. C. Chang, T. S. Chen, and H. S. Hsia, "An Effective Image Steganographic Scheme Based on Wavelet Transform and Pattern-Based Modification", IEEE Proceedings of the 2003 International Conference on Computer Networks and Mobile Computing, 2003

[3] El Safy, R. O., H. H. Zayed, and A. El Dessouki. "An adaptive steganographic technique based on integer wavelet transform." Networking and Media Convergence, 2009. ICNM 2009. International Conference on. IEEE, 2009