Analysis and Basic Implementation of Audio Watermark in Digital Image

P. S. Patel, N. M. Patel, U. K. Jaliya.
BVM Engineering college, V .V. Nagar.
Gujarat Technological University, Gandhi Nagar.
Pratiksha28785@gmail.com
nmpatel@bvmengineering.ac.in
udayjaliya@gmail.com

Abstract: **Image watermarking has a growing success in the image processing field in recent years. This paper presents analysis of watermark technique which uses audio as a watermark and image as a cover of audio watermark. Audio is used to increase robustness of watermark within the cover image. Here Audio samples are imperceptibility applied using MFCC on the most significant component of Image which give robustness. Real time applications of such audio hiding technique are covert communication (exchange a message secretly within image) and authentication or tamper proofing.**

Keywords MFCC--Mel Frequency Cepstral coefficient, PSNR--Peak Signal to Noise Ratio.

I. Introduction

Watermarking techniques are classified into two groups based on working domain. One type is spatial domain technique and other is frequency domain technique. The simplest method of watermarking is to embed a watermark into Least Significant Bit of the cover image.

$$Image_w(x, y) = Image(x, y) + k*W(x, y) \quad (1)$$

In equation (1) $Image_w$ is the watermarked image, Image is the original cover image, W is a watermark, k is a gain factor which is related to the strength of watermark. Value of k is chosen as tradeoff between Quality of watermarked image and robustness of watermark. If the strength of watermark increased then robustness increases and recovery of watermark is good but watermarked image suffer from visible degradation.

Audio watermark is embedded in digital image with using key to provide a security. Audio watermark becomes more robust with cryptographic key. Audio watermarking system is shown in figure 1. Communication channel is used to send a watermarked image which may suffer from noice or attack. Degradation in image or extracted watermark is possible due to noise and attack. Embedded watermark should have high capacity. It should be invisible and robust. Extracting watermark procedure has same key as used in embedding watermark. Secret key provides additional security but it is necessary to distribute a key with secure channel otherwise unauthorised user can extract a watermark.

To attach an audio watermark into digital image**,** we have to convert analog signal into digital. Two methods for the same are:

1) Watermark preparation based on decimation.
2) Watermark preparation by extracting MFCC from audio.

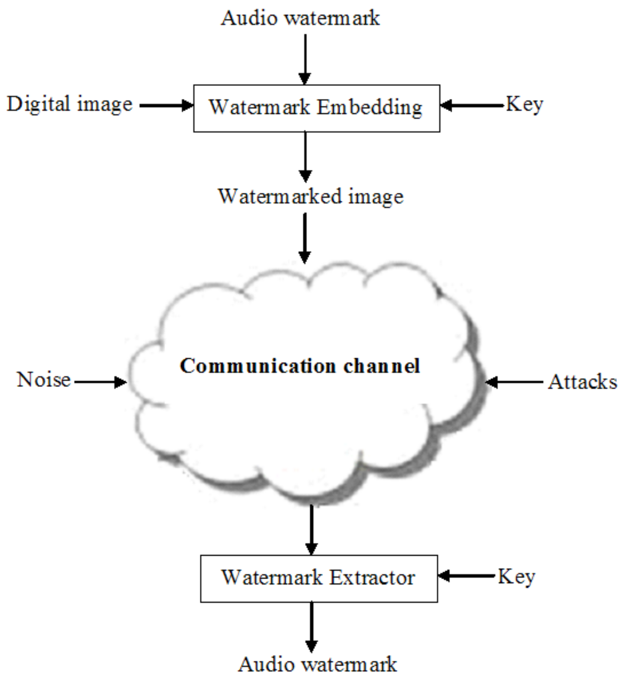

Fig 1 overview of audio watermarking system

MFCC calculation is explained in next section.

Many types of watermarks are used depend on its application.

1) According to the Type of document to be watermarked
   - Text Watermarking
   - Image Watermarking
   - Audio Watermarking
   - Video Watermarking
2) According to Human Perception
   - Visible watermarks
   - Invisible watermarks\
3) According to the necessary data for extraction
   - Informed or Private Watermarking
   - Blind or Public Watermarking
4) According to working domain
   - Spatial Domain
   - Transform Domain

Other issues for Watermark are Attacks. Here we summarize five types of attacks [2].

1) Active attacks: Remove watermark
2) Passive attacks: Not remove but attempts to determine watermark

3) Forgery attacks: Embedding new watermark rather than removing
4) Lossy Compression: JPEG, MPEG etc.
5) Geometric Attacks: Cropping, scaling etc.

II. MFCC

Mel Frequency Cepstral Coefficients (MFCCs) are audio representation coefficients. The MFCCs analysis is derived from cepstral representation of audio data. The cepstrum is defined as the inverse Fourier transform of the log–spectrum [3]. The difference between Mel frequency cepstrum and standard cepstrum is that the MFCC frequency bands are positioned logarithmically. The coefficients generated by algorithm are fine representation of signal spectra with great data compression. The main use of the coefficients is audio compression, automatic speech recognition and signal classification.

This is expressed in the Mel-frequency scale, which is linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. The process of computing MFCCs is shown in Figure 2.

Pre-emphasis: the digitized speech waveform has a high dynamic range and suffers from additive noise. In order to reduce this range, pre-emphasis is applied.

In the time domain, with input $x[n]$ and $0.9 \leq \alpha \leq 1.0$, the filter equation

$$H(Z) = 1 - \alpha \cdot z - 1,\ 0.9 \leq \alpha \leq 1.0$$

Where α is the pre-emphasis parameter and z is sampled data. Pre-emphasis is implemented as a fixed coefficient filter, where the coefficient *a* is adjusted with time according to the auto-correlation values of the speech.

The pre-emphasis filter is applied on the input signal before windowing.

The steps of MFCC,

1) *Framing:*
Audio signal is blocked into frames of *N* samples, with adjacent frames being separated by *M* (*M* < *N*). The first frame consists of the first *N* samples.

2) *Windowing:*
A Hamming window w[n] of length N is applied to each frame x[k]

$$y_l(n) = x_l(n)w(n), \quad 0 \leq n \leq N-1$$

The *Hamming* window has the form

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \leq n \leq N-1$$

3) *FFT*:
An FFT of length L is applied to compute the magnitude spectrum of the short time signal. If N <L each frame of N samples is zero padded to form an extended frame of L samples. Only output of FFT for f = 0. . . L/2 − 1 are used for further processing.

4) *Absolute value*:
using the output of FFT calculate absolute value which is used as an input to calculate mels.

5) *Mel Scale Filter Bank:*
Compute mels for a frequency *f* in Hz.
mel(f)= 2595*log10(1+f/700)

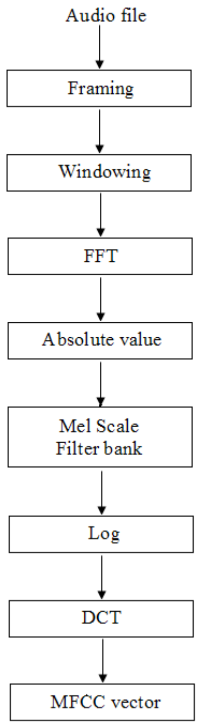

Fig 2 MFCC flow diagram

6) *Mel Scale Filter Bank:*
Compute mels for a frequency *f* in Hz.
mel(f)= 2595*log10(1+f/700)

7) *Log:*
Apply natural logarithm.

8) *DCT:*
MFCC are calculated from the output of non-linear transformation using the Discrete Cosine Transform.

MFCC vector is the digital form of analog audio signal, which is used to attach in digital image. In audio watermarking MFCC vector plays an important role.

III. Embedd & Extract Audio Watermark

Audio signals are analog and images are in digital form.

Figure 3 illustrates the block diagram of the watermark embedding process and watermark extracting process [4].

Steps of Embedding and extracting watermark

To embedd a watermark,

1) Take digital image as a watremark.
2) Apply DCT on cover image. apply DCT on gray image.
3) Remove DC coefficient which is located at (0,0) on (x,y) coefficient of image (MXN).
4) Take audio file which is used as a watermark.
5) Find MFCC coefficient as given in II section.
6) Find the length of MFCC. Lets say N.

7) Find N number of AC coefficient which has largest value to attach a watermark.
8) Attach watermark into cover image. watermark value is inserted into DCT using.

$$V_i' = V_i(1 + \alpha X_i)$$

where,
x_i = i^{th} watermark value
$v_{i'}$ = modified DCT element
v_i = original DCT element
α = strength of watermark

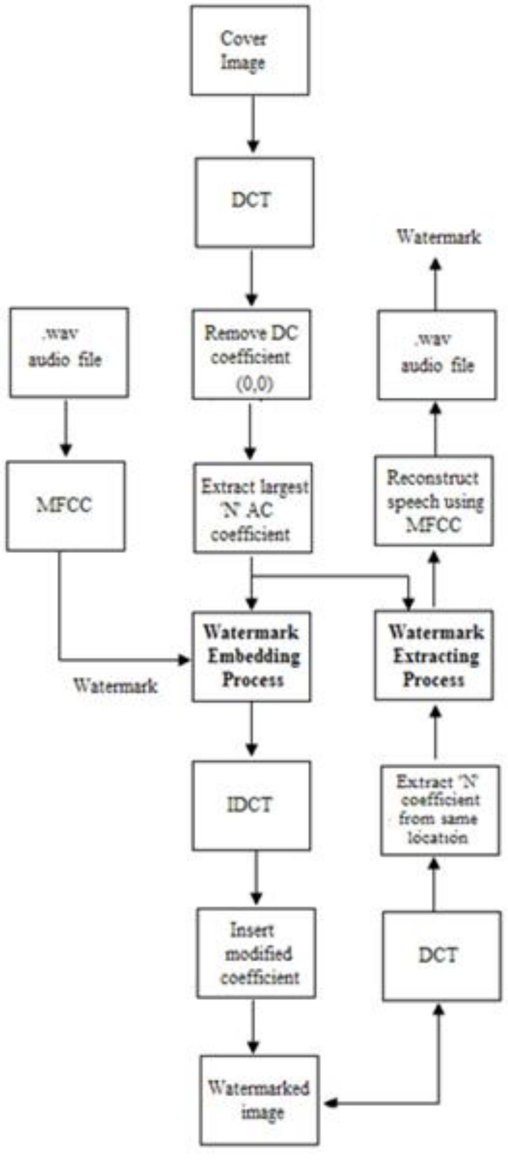

Fig 3 Embedding and Extracting audio watermark

9) Apply inverse DCT on modified DCT elements..
10) Insert modified coefficient and get watermarked image.

To extract a watermark,

11) Apply DCT on watermarked image.
12) Extract 'N' coefficient from same location as watermark insert in the embedding process.
13) Extract watermark by using the inverse of embedded equation.

$$X_i^* = [(V_i'/\ V_i)-1]/\ \alpha$$

where,
x_i^*= i^{th} recovered watermark value
$v_{i'}$ = recovered DCT element
v_i = original DCT element
α = strength of watermark

14) Use MFCC to reconstruct the audio (reconstruction process is shown in next section).
15) After reconsrtruction process we get audio file which is the hidden message.

The implemented algorithm represents a scheme for embedding the audio as a watermark. The decoding stage of the diagram helps in properly retrieving the audio watermark from watermarked image.

To check the performance of algorithm find **Peak Signal-to-Noise Ratio (PSNR),** which is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. Because many signals have a very wide dynamic range, PSNR is usually expressed in terms of the logarithmic decibel scale. It is most easily defined via the mean squared error (**MSE**) which for two *m*×*n* monochrome images *I* and *K* where one of the images is considered a noisy approximation of the other is defined as:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\|I(i,j)-K(i,j)\|^2$$

The PSNR is defined as:

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$

Here, MAX_i is the maximum possible pixel value of the image. When the pixels are represented using 8 bits per sample, this is 255.

IV. Audio Reconstruction

Audio can be reconsructed from MFCC Vector as shown in fig 4. It is the inverse procedure of MFCC calculation.

To get audio as a result we have to convert MFCC vector into analog form with using inverse calculation procedure of MFCC.

We get the recovered audio from the watermarked image in digital form. We can hear audio if we convert it into analog form using reconstruction process.

The steps are given below.

1) Take inverse DCT.
2) Apply exponential on inverse DCT.
3) Apply inverse Mel matrix.
   f=700*(10^(e/2595)-1)
   where f=freqency

e=number of mels.

4) Apply inverse power spectrum or inverse FFT on the output of inverse Mel Matrix to get reconstructed audio.

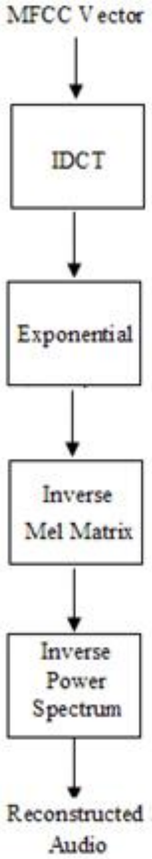

Fig 4 Reconstruction of audio from MFCC vector

V. RESULTS

Algorithm is applied on many images and some of the results are shown in Figure. Figure 5 shows the original image, Figure 6 shows the original audio signal, which is used as watermark in original image. watermarked image shown in figure 8 and extracted audio is shown in figure 9.

Fig 5 Original Image

Figure 5 shows original image which is used as a cover image. Audio watermark is embedded in the original image. The dimension of Original Image is 251 X 201 in jpg format.

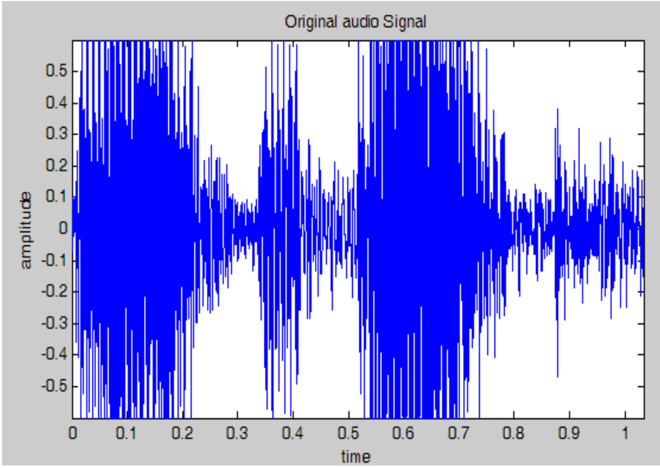

Fig 6 Original Audio Signal

Graph in Figure 6 shows 1 Sec audio in time domain with amplitude. It is original audio signal, which is hidden in image as a secret message. Receiver receives watermarked image and extract this audio from watermarked image.

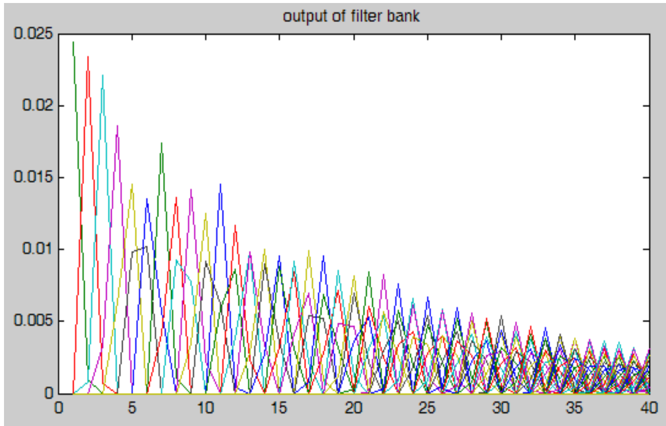

Fig 7 Output of Filter Bank

Figure 7 shows Mel Filter bank, which is a bank of band pass filters, with center frequencies linearly space in perceptual scale. The most common shape for filters is triangular. Mel Filter bank is used to calculate MFCC vector from audio signal.

Fig 8 Watermarked Image with strength 0.8

Figure 8 shows the result of watermared image. Audio is hidden in this Watermarked image. Distortion in image is dependent on the strength of watermark. As the strength of watermark increases, the robustness of watermark increases and also it makes recovery good but the image suffers visible degradation. So, the value of the watermark strength is chosen as a trade off value between perceptibility and robustness.

Fig 9 Watermarked image with strength 6.8

Strength is high so the robustness is high and the recovery is good but the image shows visible degradation in figure 9.

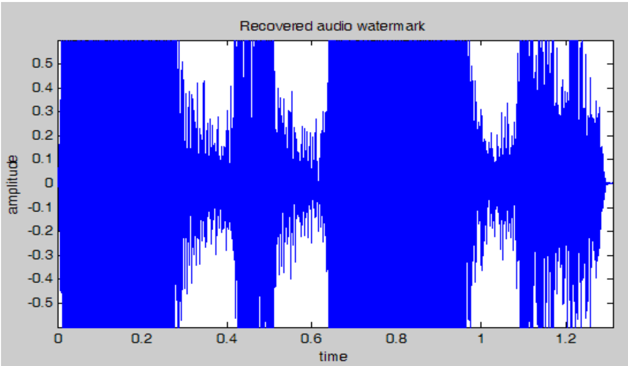

Fig 10 Recovered Audio Watermark

Figure 10 shows Recovered audio watermark from watermarked image. See the difference with figure 6 which is original audio.

Table 1 PSNR

WM	IMAGES		
In sec	259X194 PSNR	251X201 PSNR	512X514 PSNR
1	35.9658	43.8669	29.4961
2	34.7510	40.9272	28.7807
3	34.5411	40.8877	28.6752
4	13.6967	6.3281	17.1829
5	13.5685	6.2696	20.1854

Table 2 PSNR

WM	IMAGES	
In sec	1024X768 PSNR	1024X1024 PSNR
1	31.7253	39.4857
2	29.1832	36.6263
3	29.9967	36.4898
4	18.6894	27.0345
5	22.0255	29.9350

Table 1 and Table 2 shows the PSNR value for different size of images and different length of watermark.
PSNR is used to evaluate the performance. How much length of watermark should be attached is depends on the size of image.

VI. CONCLUSIONS

The performance of this scheme is illustrated by embedding audio data within images to produce watermarked images. When these watermarked images are decoded, the audio data are recoverable.

Listening test is performed to evaluate the intelligibility contained in the extracted audio. This paper includes basics of watermarking, the procedure of embedding watermark and extracting audio from the watermarked image. It also includes implementation results after implementation of embedding audio watermark and extracting watermark from cover image.

We conclude that audio can be used as a WATERMARK and gives capacity, robustness, imperceptibility and security.

VII. FUTURE WORK

1) Key can be used to randomize MFCC vector to get better security.
2) To Apply attacks on watermark and check the robustness of algorithm.
3) Use other methods like DFT or DWT and check the performance.

ACKNOWLEDGMENT

We would like to express our sincere thanks to **Prof. P. B. Swadas** for giving us such an opportunity to explore new ideas. We are also very much thankful to all **Faculties of Birla Vishvakarma Mahavidhyalaya** as they have been great sources of inspiration to us and we thank them from the bottom of heart.

The support and encouragement of my family members and friends cannot be acknowledged enough. Our eternal gratitude goes to parents.

REFERENCES

[1] N. Nikolaidis, I. Pitas, “Digital Image Watermarking: an Overview”, proc. Artificial Intelligence and Information Analysis Laboratory Department of Informatics, Aristotle University of Thessaloniki 1999.

[2] Sviatolsav Voloshynovskiy, Shelby Pereira, and Thierry Pun, “Attacks on Digital Watermarks”, IEEE Comunication magazine, Aug. 2001.

[3] Ben Milner and Xu Shao, “Speech Reconstruction from MFCCs using a source-filter model”, school of information systems, University of East Anglia, Norwich, UK. Proc. ICSLP, 2002.

[4] Priti.P.Rege, “Speech as Watermark Embedding in Digital Images”, NCSIA 2009, National conference on signal and image processing applications.

[5] C. Neubauer, J. Herre, “Digital watermarking and its influence on audio quality”, Fraunhofer Institute for integrated circuits IIS, D-91058 Erlangen, Germany. Print no 4823.

[6] Priya Gupta, Derek Justice, and Matt Pfenninger, “Digital Watermarking: Techniques and Attacks”, Sept. 2002.

[7] I. Cox, M. Miller, and J. Bloom, “Watermarking applications and their properties”, Proc. Intl. Conf. Inf. Tech.: Coding & Computing 2000, pp. 6-10, Mar. 2000.

[8] A. Piva, M. Barni, F. Bartolini, V. Cappellini, “DCT-based watermark recovering without resorting to the uncorrupted original image”, preprint, 1997.

[9] Xu Shao and Ben Milner, “Clean Speech Reconstruction from Noisy Mel-Frequency Cepstral Coefficients using a Sinusoidal Model”, School of Information Systems, University of East Anglia, Norwich, UK.

[10] Sirko Molau, Michael Pitz, Ralf Schl ¨uter, and Hermann Ney "Computing Mel-Frequency Cepstral Coefficients on The Power Spectrum", Lehrstuhl f¨ur Informatik VI, Computer Science Department, RWTH Aachen – University of Technology, 52056 Aachen, Germany.