HIGH CAPACITY DATA EMBEDDING SCHEMES FOR DIGITAL MEDIA

A DISSERTATION

SUBMITTED TO THE DOCTORAL PROGRAMME OF THE

UNIVERSITAT OBERTA DE CATALUNYA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Author: Mehdi Fallahpour

Thesis director: David Megías

October 2009

ACKNOWLEDGMENTS

This work is partially supported by the Spanish Ministry of Science and Innovation and the FEDER funds under the grants TSI2007-65406-C03-03 E-AEGIS and CONSOLIDER-INGENIO 2010 CSD2007-00004 ARES.

ABSTRACT

High capacity image data hiding methods and robust high capacity digital audio watermarking algorithms are studied in this thesis. The main results of this work are the development of novel algorithms with state-of-the-art performance, high capacity and transparency for image data hiding and robustness, high capacity and low distortion for audio watermarking.

- **Image contents**

Using the histogram of an image is a new idea in image data hiding [63]. The aim of this research is to obtain a narrow histogram after a modification step (such as prediction or tilling), resulting in an increased capacity. Reversibility is the main property in some kind of data hiding methods such as medical applications, since it is required that not only the hidden message but also the original (unmarked) image can be extracted at the detector side.

The peak point of the histogram of the image is used for embedding information. The value of this peak identifies the capacity payload. Based on the idea of using the histogram to embed secret information, prediction and tilling techniques are used to achieve good capacity and transparency, which have led to the published papers [68, 92, 69] contributed by the author of this thesis.

- **Audio contents**

The use of a psychoacoustic model is quite useful to design watermarking algorithms, as shown in the designed FFT-based schemes [78, 80]. In addition, to achieve robustness, the frequency domain is also a better choice compared to the time domain. In this thesis, the frequency domain, and more precisely, the Fast Fourier Transform (FFT) and the Digital Wavelet Transform (DWT) have been chosen since they are more efficient than other transforms and provide more robustness than the methods designed in the time domain.

Different attacks produce various changes in audio. One of the most important attacks in audio is compression. To defeat MPEG compression (MP3), comparing the original with a compressed/decompressed signal to find a safe area for embedding is a convenient possibility, and this idea is exploited in the proposed algorithm [80]. In

addition, the use of different techniques like differences, predictions, interpolation and spatial transforms in FFT have resulted in high capacity methods leading to the publication of three different papers [78- 80]. The fourth paper takes advantage of the DWT to design a high capacity, transparent and robust watermarking scheme [81]. The comparisons provided in these papers [79-81] with the existing audio watermarking schemes show that the suggested schemes have excellent capacity, about ten times than existing algorithms, whilst keeping transparency in the high quality area (ODG in [–1, 0]) and robustness against common attacks.

List of Contributions

The validation of the research is carried out with the publication of the results of this thesis on the seven original papers (and additional four papers are under review). All analysis and simulation results presented in publications or this thesis have been produced by the author (except the unpublished paper 91, for which the author of this thesis is not the main contributor).

- **Published papers**

68. **M. Fallahpour**; D. Megías. "Reversible Data Hiding Based On H.264/AVC Intra Prediction", *International Workshop on Digital Watermarking (IWDW 2008). Lecture Notes in Computer Science* 5450, pp. 52–60, 2009.

69. **M. Fallahpour**, D. Megías, M. Ghanbari, "High capacity, reversible data hiding in medical images", *IEEE International Conference on Image Processing, ICIP*2009, in press.

78. **M. Fallahpour,** D. Megías, "High capacity method for real-time audio data hiding using the FFT transform" *Advances in Information Security and Its Application* Third International Conference, ISA 2009, Springer, Seoul, Korea, June 25-27, 2009.

79. **M. Fallahpour**, D. Megias, "High capacity audio watermarking using FFT amplitude interpolation", *IEICE Electron. Express*, Vol. 6, No. 14, pp.1057-1063, 2009, (Impact factor 0.48 in 2008).

80. **M. Fallahpour,** D. Megías, "Robust high-capacity audio watermarking based on FFT amplitude modification" *IEICE Transactions on Information and Systems*, Vol.E93-D, No.01, pp.-, Jan. 2010, in press (Impact factor 0.36 in 2008).

81. **M. Fallahpour,** D. Megías, "DWT–based high capacity audio watermarking" *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol.E93-A, No.01, pp.-, Jan. 2010, in press (Impact factor 0.43 in 2008).

92. **M. Fallahpour**, "Reversible image data hiding based on gradient adjusted prediction", *IEICE Electron. Express*, Vol. 5, No. 20, pp.870-876, 2008. (Impact factor 0.48 in 2008).

- **Under review**

70. **M. Fallahpour**; D. Megías, Mohammad Ghanbari, "High capacity, reversible data hiding in medical images" Submitted revised version to *IET Image Processing*, August 2009, (Impact factor 0.51 in 2008).

89. **M. Fallahpour**; D. Megías, Mohammad Ghanbari, "Subjectively adapted high capacity lossless image data hiding based on prediction errors", submitted to *Multimedia Tools and Applications, Springer,* August 2009 (Impact factor 0.46 in 2008).

90. **M. Fallahpour,** D. Megías, “High capacity audio watermarking using the high frequency band of the wavelet domain” submitted to *Multimedia Tools and Applications, Springer,* August 2009, (Impact factor 0.46 in 2008).

91. D. Megías, J. Serra, **M. Fallahpour,** “A novel blind audio watermarking system with enhanced transparency and self-synchronisation” submitted to *Signal processing*, July 2009, (Impact factor 1.2 in 2008).

Contents

CHAPTER 1

INTRODUCTION

The fast growth of the Internet and the digital information revolution is causing significant changes in the global society, ranging from the influence on the world financial system to the way people communicate nowadays. Broadband communication networks and multimedia data existing in a digital formats (images, audio and video) have opened many challenges and prospects for innovation. Flexible and simple-to-use software and the decreasing prices of digital devices (*e.g.* digital photo cameras, camcorders, portable CD and MP3 players, DVD players, CD and DVD recorders, laptops, PDAs and so on) have made it possible for users from all over the world to produce, edit and exchange multimedia data. Broadband Internet connections and almost an errorless transmission of data make it possible to distribute large multimedia files and make exact digital copies of them. Digital media files do not suffer from any quality loss due to multiple copying processes, as occurs with analogue audio and VHS tapes.

1.1 Introduction to watermarking

The simplicity of content modification and a perfect reproduction in the digital domain have led the protection of intellectual ownership and the prevention of the unauthorised tampering of multimedia data to become an important technological and research issue [1, 2, 3, 4].

Simple protection methods which were based on the data embedded into the header bits of the digital file are useless since the header information can be easily removed by a change of data format, which does not affect the fidelity of media. Encryption of digital multimedia prevents access to the multimedia content to an individual without a suitable decryption key. Thus, content providers get paid for the delivery of perceivable multimedia and each client who has paid the royalties must be able to decrypt a received file properly. Once the multimedia has been decrypted, it can be repeatedly copied and distributed without any obstacles. Current software applications and broadband Internet provide the tools to carry out it quickly and without deep technical knowledge. One of

the most recent cases of breaking such a system is the hack of the Content Scrambling System for DVDs [5, 6].

It is obvious that the existing security protocols for electronic commerce secure only the communication channel between the content provider and the user. Digital watermarking has been proposed as a novel alternative scheme to protect the intellectual property rights and defend digital media from tampering [1, 2]. It involves a procedure of embedding a perceptually transparent digital signature into a host signal, carrying a message about the host signal in order to "mark" its ownership.

Generally, the term **"**watermarking**"** refers to the act of hiding a secret message signal into a host signal (usually digital media). It should be performed with no significant perceptual distortion of the host signal. As the word "watermarking" suggests, the mark itself is transparent or unnoticeable for the human perceptual system. As it is widely known, the Human Visual System (HVS) is not sensitive to small changes in the colour of the pixel. Hence, it is possible to modify the pixel values of images/videos by inserting a watermark without the modification being noticed by observers. Providing that a certain HVS threshold is not exceeded, the modified (marked) image/video will be undistinguishable from the original to the human eye. On the other hand, all the developed algorithms in audio take advantage of the perceptual properties of the Human Auditory System (HAS) in order to add a watermark into a host signal in a perceptually transparent manner. Embedding additional information into audio sequences is a more difficult task than that of images, due to dynamic supremacy of the HAS over human visual system [7].

Watermarking algorithms were primarily developed for digital images and video sequences [7, 8] and the interest and research in audio watermarking started slightly later [9, 10]. In the past few years, several algorithms for the embedding and extraction of watermarks in audio sequences have been presented. All of the developed algorithms take advantage of the perceptual properties of the Human Auditory System (HAS) in order to add a watermark into a host signal in a perceptually transparent manner. On the other hand, many attacks that are malicious against image watermarking algorithms (*e.g.* geometrical distortions, spatial scaling, etc.) cannot be implemented against audio watermarking schemes, although other specific attacks exist for digital audio contents.

Digital watermarking is considered as a secure, imperceptible and robust communication of data related to the host signal. The basic aim is that the embedded

(hidden) information follows the marked multimedia and stands unintentional modifications and intentional removal attempts. The most significant design challenge is to embed the watermark so that it is reliably detected by a watermark detector. The relative relevance of the watermarking properties depends significantly on the application for which the algorithm is designed [11]. For copy protection applications, the watermark must be recoverable even when the watermarked signal undergoes a considerable level of distortion.

The goal of digital image watermarking for copyright protection is to embed a mark into the image data which can identify the copyright holder of the work. Together with owner identification, one might also want to embed a mark identifying the buyer of a work for circulation tracking. The mark can be a registered number a text message or graphical logo, or some unique pattern.

The term "watermark**"** stems from the ancient art of marking paper with a logo for the same purpose. Digital watermarks can either be perceptible or imperceptible. Visible image watermarks, often the logo of the copyright holder, can be easily applied to the image but are hard to remove. Mintzer describes a successful implementation of visible image watermarking in [12]. On the other hand, many applications require the watermark to be invisible. Some invisible watermarking schemes are required to be robust against common image processing operations, like image compression (*e.g.* JPEG), image filtering (edge enhancement, contrast enhancement and others), and geometrical transform motions (*e.g.* cropping or scaling) Therefore, the watermark cannot be stored in the file format, but has to be embedded into the image data itself. In this regard, cryptographic techniques and statistical properties of pseudo-random numbers play an essential role.

Though the art of papermaking was invented in China over one thousand years earlier, paper watermarks did not appear until about 1282, in Italy [93]. The marks were made by adding thin wire patterns to the paper molds. The paper would be a little thinner where the wire was and thus more transparent. The meaning and reason of the earliest watermarks are uncertain. They may have been used for practical functions such as identifying the molds on those sheets of papers were made, or as trademarks to identify the paper maker. In contrast, they may have represented mystical signs, or might simply have served as decoration. By the eighteenth century, watermarks on paper made in Europe and America had become more obviously functional. They were

used as trademarks, to trace the date the paper was manufactured and to point to the sizes of original sheets. It was also about the time that watermarks began to be used as anti-counterfeiting measures on money and other documents. The term watermark is actually a misnomer, in that water is not especially important in the creation of the mark. It was probably given because the marks resemble the effects of water on paper.

1.2 Background and state of the art

The following sections present an overview of the applications of digital watermarking, the properties of watermarking systems and the state of the art of both image and audio watermarking.

1.2.1 Applications of digital watermarking

Watermarking can be the enabling technology for a number of applications. Each application requires different conditions on the watermarking system. As a result, watermarking algorithms targeting different applications might be very different in nature. Besides, devising an efficient watermarking method might be much more difficult for certain applications. In this section, some of the main application domains of watermarking are briefly reviewed.

1. Copy and access control

In copy control applications, the embedded watermark represents a copy or access control policy. After a watermark has been detected and the content decoded, the copy or access control policy is enforced by directing particular hardware or software operations such as enabling or disabling the record module [21]. These applications need watermarking algorithms robust against intentional attacks and signal processing modifications.

2. Authentication and tampering

In the content authentication applications, a set of secondary data is embedded into the host multimedia signal. This information is used later on to determine whether the host multimedia was tampered [22]. The robustness against removal of the watermark or making it undetectable is not a concern as there is no such motivation from the attacker’s point of view. Generally, the payload (capacity) needs to be high to satisfy the

demand for more additional data compared to ownership protection applications (see below). In addition, the detection must be a blind, *i.e.* detection must be performed without the original host signal because the original is unavailable.

3. Fingerprinting

Further data embedded by the watermarking scheme in fingerprinting applications is used to trace the creator or recipients of an exact copy of a multimedia file [23-30]. For example, watermarks carrying different serial or identity (ID) numbers are embedded in different copies of music CDs or DVDs, or multimedia files sold in on-line shops, before distributing them to a large number of recipients. The algorithms implemented in fingerprinting applications must show high robustness against intentional attacks and signal processing modifications such as lossy compression or filtering.

4. Ownership protection

In ownership protection applications, a watermark including ownership information is embedded into the multimedia host signal. The watermark, recognised only to the copyright holder, is supposed to be very robust and secure, enabling the owner to show the presence of this watermark in case of dispute to demonstrate his/her ownership. In this case, watermark detection must have a very low false alarm probability. On the other hand, ownership protection applications usually require a small embedding capacity.

5. Proof of ownership

This kind of application tries to solve the problem which occurs when an attacker uses editing software to replace the original copyright notice with his own one and, then, claims to own the copyright himself. In the case of early watermarking systems, the problem was that the detector was readily available to adversaries. To achieve the level of the security required for proof of ownership, it is indispensable to limit the availability of the detector. When an adversary does not have the detector, the removal of a watermark can be made really difficult. But, even if the owner’s watermark cannot be removed, an adversary might try to undermine the owner. As described in [2], an adversary, using his own watermarking method, might be able to make it appear as if his watermark data was present in the owner’s original host signal. This problem can be solved using a small change of the problem statement. Instead of a direct proof of ownership by embedding a watermark signature in the host object, the scheme will

instead try to verify that the adversary's object is derived from the original watermarked image. Such an algorithm presents indirect proof that it is more probable that the real owner owns the disputed image, because s/he is the one who owns the version from which the other two were created.

6. Information carrier

The hidden information in the application is accepted to have a high capacity and to be detected and decoded using a blind detection algorithm. While robustness against intentional attacks is not required, a certain degree of robustness against common processing like MPEG compression may be desired. A public watermark embedded into the host multimedia might be used as the link to external databases which contain certain additional information about the multimedia file itself, such as copyright information and licensing conditions. In audio applications, some metadata information may be embedded into the multimedia file (*e.g.* including information about the soloist, composer, genre of music, lyrics and others).

7. Broadcast monitoring

A variety of applications for audio watermarking are in the area of broadcasting [31-34]. Watermarking is a well-defined alternative method of coding identification information for broadcast monitoring. It has the benefit of being embedded within the multimedia host signal itself rather than exploiting a particular part of the broadcast signal. Hence, it is compatible with the already installed base of broadcast equipment, including digital and analogue communication channels. The main disadvantage of such systems is that the embedding procedure is more complex than a simple placing data into the file headers. There is also a concern, especially on the side of content creators, that the watermark would introduce distortions and degrade the visual or audio quality of multimedia. A number of broadcast monitoring watermark-based applications are already available on commercial basis. These consist of programme type identification, advertising research, broadcast coverage research and similar applications.

In general, the types of watermarks can be summarized as follows:

- *Robust watermarks* are designed to resist various manipulations; all applications presupposing security of the watermarking systems need this type of watermark.

- *Fragile watermarks* are embedded with very low robustness. Thus, this type of watermark can be destroyed even by the slightest manipulations.
- *Public* and *private watermarks* are differentiated according the secrecy requirements for the key used to embed and retrieve markings. In accordance with the basic principle of watermarking, the same key is used in the encoding and decoding process. If the key is known, this type of watermark is considered as public and, if the key is hidden, as private watermarks. Public watermarks can be used in applications which do not have security-relevant requirements.
- *Visible* or *localized watermarks* can be logos or overlay images in the field of image or video watermarking. Due to the implicit localization of the information, these watermarks are not robust.

1.2.2 Properties of digital watermarking

Watermarking systems can be characterised by a number of important properties. The relative significance of each property is dependent on the requirements of the application and the role the watermark will play. Indeed, even the interpretation of a watermark property can vary with the application.

Some properties associated to watermarking schemes are outlined below.

1. Perceptual transparency

The watermark pattern is embedded in the digital audio/image producing alterations. These alterations should not degrade the perceived quality of the marked object. In some of the applications, the watermark-embedding algorithm inserts additional data without affecting the perceptual quality of the audio/image host signal [13-17]. Large alterations are robust and can be detected with great certainty, but they produce degradation in the audio/image. The fidelity of the watermarking algorithm is defined as the perceptual similarity between the original and the marked multimedia sequence.

2. Payload capacity size

The capacity of the embedded watermark is the number of the embedded bits within a unit of time or space and is usually given in bits per second (bps) for audio and bits per pixel (bpp) for an image. Some audio watermarking applications, such as copy

control, need the insertion of a serial number or author ID, with the average bit rate of up to 0.5 bps. For a broadcast monitoring watermark, the bit rate is higher, due to the necessity of the embedding an ID signature of a trade mark within the first second at the start of the broadcast clip, with an average bit rate up to 15 bps. In image data hiding, depending on the application and method, the capacity varies from 2 to 40 kbit for a 512×512 greyscale image.

3. Complexity and cost

The implementation of a watermarking system is a complex task, and it depends on the application. In broadcast monitoring, embedding and detection must be performed in real time whereas in copyright protection applications, time is not a crucial factor for a practical implementation. One of the most relevant issues in the design of embedders and detectors, which can be implemented as hardware or software plug-ins, is the difference in processing power of different devices (laptops, PDA, mobile phones, and others). On the other hand, complexity is not a usual requirement for image applications. In the general case, the complexity in image applications is not comparable with audio and video applications since the amount of information in image files is much lower than that of audio and video files.

4. Robustness

The robustness of the method is defined as the ability of the watermark detector to extract the embedded watermark after common (signal processing) manipulations. Watermarking applications usually require robustness in the presence of a predefined set of signal processing modifications so that the watermark can be reliably extracted at the detection side. For example, in radio broadcast monitoring, the embedded watermark require only surviving distortions caused by the transmission process, including dynamic compression and low pass filtering, since the watermark detection is performed directly from the broadcast signal. Compression (JPEG for image and MP3 for audio) and filtering are well-known modifications which must be considered by robust watermarking applications [18-20].

5. Blind or informed watermark detection

In many applications, the detector needs the original host multimedia content to extract watermark from the marked sequence. It often significantly improves the detector performance, since the original multimedia can be subtracted from the

watermarked copy, resulting in the watermark sequence alone. However, the detection algorithm may not have access to the original signal (blind detection) and this inability substantially decreases the amount of data which can be hidden in the host signal.

6. Security

A watermarking algorithm must be secure in the sense that an adversary must not be able to detect the existence of embedded data, let alone remove the embedded data. The security of a watermarking system is interpreted the same way as the security of encryption methods, and it should not be broken unless an authorised user has access to a secret key that controls the watermark embedding (and extraction) process. An unauthorised user should be unable to extract the data in a reasonable amount of time even if s/he knows that the host signal contains a watermark and is familiar with the exact watermark embedding method. Security needs vary with the application and the tightest requirements are found in cover communications applications, where, in some cases, the embedded data is encrypted prior to embedding it into the host object.

1.2.3 Overview of existing watermarking methods

1.2.3.1 Audio contents

Watermarking of audio signals is more challenging compared to the watermarking of images or video sequences due to the wider dynamic range of the human auditory system (HAS) in comparison with human visual system (HVS) [7]. The HAS perceives sounds over a range of power larger than 10^9:1 and a range of frequencies greater than 1000:1. The sensitivity of the HAS to the Additive White Gaussian Noise (AWGN) is high as well. This noise in a sound file can be detected as low as 70 dB below ambient level. In contrast, opposite to its large dynamic range, the HAS contains a fairly small differential range, *i.e.* loud sounds generally tend to mask out weaker sounds. Additionally, the HAS is insensitive to a constant relative phase shift in a stationary audio signal and some spectral distortions are interpreted as natural, perceptually non-annoying ones. [7]. Two properties of the HAS frequently used in watermarking algorithms are frequency masking and temporal masking [35]. The concept of using the perceptual holes of the HAS is taken from wideband audio coding (*e.g.* MPEG 1 compression, layer 3, usually called MP3) [36]. In compression algorithms, the holes are used in order to decrease the amount of the bits needed to encode the audio signal

without causing any perceptual distortion to the coded audio. On the other hand, in the information hiding scenarios, masking properties are used to embed additional bits into an existing bit stream, again without generating audible noise in the audio sequence used for data hiding.

1. Frequency masking

Frequency masking is a frequency domain phenomenon for which a low level signal, *e.g.* a pure tone, can be made inaudible (masked) by a simultaneously appearing stronger signal, *e.g.* a narrow band noise, if the masker and the maskee are close enough to each other in frequency [35]. A masking threshold can be derived below which any signal will not be audible. The masking threshold depends on the masker and on the characteristics of the masker and the maskee (narrowband noise or pure tone).

2. Time masking

Besides frequency masking, two phenomena of the HAS in the time domain also play an important role in human auditory perception. Those are pre-masking and post-masking in time [35]. The time masking effects appear before and after a masking signal has been switched on and off, respectively. Both pre and post-masking have been used in the MPEG audio compression algorithm and several audio watermarking methods.

Selected audio watermarking algorithms

Some of the best known general audio watermarking algorithms are the following:

- Least Significant Bit(s) (LSB) coding

  One of the first few techniques studied in the information hiding and watermarking area of digital audio (as well as other media types [36-39]) is LSB coding [40]. A natural approach in the case of the audio sequences is to embed watermark data by changing the LSB of the individual samples of the digital audio stream having the amplitude resolution of 16 bits per sample. It usually does not exploit any psychoacoustic model to perceptually weight the noise introduced by the LSB replacement.

- Watermarking the phase of the host signal

Algorithms which embed the watermark into the phase of the host audio signal do not use masking properties of the HAS, but exploit the fact that the HAS is insensitive to a constant relative phase shift in a stationary audio signal. There are two main approaches used in the watermarking of the host signal's phase, namely, phase coding [41] and phase modulation [42].

- Echo hiding

A number of developed audio watermarking algorithms [43, 44] are based on the echo hiding method, described for the first time in [7]. Echo hiding schemes embed watermarks into a host signal by adding echoes to produce the marked signal. The nature of the echo is to add resonance to the host audio. Therefore, the problem of sensitivity of the HAS towards the additive noise is circumvented in this method.

- Spread spectrum watermarking

In a number of developed algorithms [45-47], the watermark embedding and extraction are carried out using spread-spectrum (SS) techniques. A SS sequence can be added to the host audio samples in time domain [47], to the FFT coefficients [48], in the sub-band domain [49], to the spectral coefficients [50] and in a compressed domain [51]. If the embedding takes place in a transformed domain, it should be located in the coefficients invariant to common watermark attacks as amplitude compression, re-sampling, low-pass filtering, and other common signal processing techniques.

- Methods using the patchwork algorithm

The patchwork technique was first presented for embedding watermarks in images. It is a statistical method based on hypothesis testing and relying on large data sets. As a second of CD quality stereo audio contains 88,200 samples, a patchwork approach is applicable for the watermarking of audio sequences as well. The watermark embedding process uses a pseudo-random process to insert a certain statistic into a host audio data set, which is extracted with the help of numerical indexes (like the mean value), describing the specific distribution.

- Methods using the Fast Fourier Transform

FFT is one of the best transform for audio watermarking because of low complexity and short computation time of FFT. David Megías *et. al.* proposed different methods based on FFT [52-55]. They analysed the MPEG 1 Layer 3 compression method to identify the frequencies at which the marking bits could be embedded [52, 53].

- Methods based on the Discrete Wavelet Transform

Among the existing transforms, the wavelet transform has several advantages in audio signal processing. Its inherent frequency multi-resolution and logarithmic decomposition of the frequency bands resemble the human perception of frequencies, since it provides the decomposition to mimic the critical band structure of the HAS [74].

- Methods based on interpolation

Interpolation techniques are often designed to provide a good perceptual quality from known sample values [83]. In [82], an original audio signal is divided into distinct frames and, then, a secret bit is embedded into each frame by using the spline interpolation. [88] proposes a spline interpolation-based watermarking scheme with more robustness against attacks compared to the one suggested in [82].

- Methods using various characteristics of the host audio

Several audio watermarking algorithms, developed in the recent years, use different statistical properties of the host audio and modify them in order to embed watermark data. Those properties are pitch values, number of salient points, difference in energy of adjacent blocks and others. However, modifications of the host signal statistical properties do influence the subjective quality of the audio signal and have to be performed in such a way that it does not produce distortions above the audible threshold. Usually, these methods are robust to signal processing modifications, but offer a low watermarking capacity. Some methods [56, 57] introduce content-adaptive segmentation of the host audio according to its characteristics in the time domain. Since the embedding parameters are dependent of the host audio, it is along the right direction to increase tamper resistance. The basic idea is to classify the host

audio into a predetermined number of segments according to its properties in time domain, and encode each segment with an embedding scheme which is designed to best suit this segment of audio signal, according to its features in the frequency domain.

1.2.3.2 Still images

Images can be represented in the spatial domain and in different transform domains. In transform domains, an image is represented in terms of its frequencies, while, in the spatial domain, it is represented by pixel values. In simple terms, the transform domain means that the image is segmented into various frequency bands. To map an image into its spectral representation, we can use several reversible transforms like the Discrete Cosine Transform (DCT), the Discrete Wavelet Transform (DWT), or the Discrete Fourier Transform (DFT). Simple watermarks can be embedded in the spatial domain of images by modifying the pixel values or the least significant bit (LSB) values, although more robust watermarks can be embedded in the transform domain by modifying some selected coefficients.

- DCT domain

DCT-based watermarking techniques are more robust compared to simple spatial domain watermarking techniques. Such algorithms are robust against simple image processing operations like low-pass filtering, brightness and contrast adjustment, blurring etc. DCT domain watermarking can be classified into Global DCT watermarking [58] and Block-based DCT watermarking [84].

- DWT domain

In the last few years, the wavelet transform has been widely studied in signal processing in general and image compression in particular. In some applications, wavelet-based watermarking schemes [59] overcome DCT-based approaches.

- DFT and FFT domain

The DFT domain [85] has been explored by researches because it offers robustness against geometric attacks like rotation, scaling, cropping, translation and other attacks. Pereira et al. (1999) proposed a watermarking algorithm based on FFT which is robust against compression, rotation, scaling and translation attacks. This

technique is also shown to be robust against cropping and print-and-scan attacks. However, it is very difficult to implement [60].

- DHT domain

DHT-based watermarking techniques rely on the Discrete Hadamard Transform. Falkowski and Lim [86] propose a watermarking technique based on multi-resolution transform and complex Hadamard transform.

1.3 Objectives of the thesis

Based on the type of media (image or audio) the objectives of this doctoral research can be divided into two parts:

1. Propose reversible high capacity image data hiding schemes with very low distortion.

2. Develop audio watermarking schemes with very high capacity, very low distortion and also robust against MPEG compression (MP3) and common signal processing attacks such as added noise or low-pass and high-pass filters. The methods must be blind an efficient such that they can be used in real-time applications (*e.g.* for broadcast monitoring).

With regards to image watermarking applications, a data embedding scheme should be chosen. For example, authentication watermarks must not be affected by legal operations, such as channel noise, as illegal attacks must destroy them. On the other hand, robustness is not a general requirement for data hiding techniques. A typical scenario for data hiding is the distribution of hidden information via newsgroups, bulletin boards, or simply by images on homepages.

In the image data hiding area, capacity and Peak Signal-to-Noise Ratio (PSNR) are the two main properties. It must be taken into account that capacity and PSNR are two parameters which change against another, *i.e.*, when capacity is increased the PSNR is decreased and vice versa.

In this thesis, we focus on reversible data embedding, also called lossless data embedding, which is a fragile technique in the sense that the embedded data will mostly be destroyed by small distortions of the image. Reversible data embedding allows one to embed a relatively large amount of data into an image in such a way that, apart from extracting secret information, the original image can be reconstructed from the marked image. This makes it an ideal technique for applications for which one wants to store metadata directly into the image and loss of quality is not always acceptable.

In some applications, the modification to the original image introduced by data hiding is tolerable in case that the original and the marked versions should be perceptually indistinguishable. However, in medical and military imaging, such an

imperceptible modification, or even a very small change in pixel values, is unacceptable since that may affect the right decision during diagnosis and analysis. Thus, using a reversible scheme for some applications is mandatory.

Watermarking techniques are applied to audio due to various reasons. Each of these possible applications involves typical processing operations which a watermarking technique must survive.

According to the deliberate application of watermarks in audio data, the algorithm as well as the watermark itself must fulfil a set of requirements. The system requirements can be divided into signal processing properties, security properties, and application-specific requirements of the algorithm. Embedding capacity, perceptual quality and robustness are the properties of the audio watermarking systems which can be selected according to the application requirements.

One of main challenges in audio processing is to estimate the behaviour of the human ear to audio distortion. To overcome this challenge, the use of psychoacoustic models is very useful. The human ear is sensitive to frequency and, for this reason, psychoacoustic models are based on frequency domain data. Using the time domain for embedding would lead to high capacity and low complexity, but to analyse the effect of the changes in the frequency domain and in distortion is the main challenge of this area.

Different attacks produce various changes in audio. One of the most important attacks in audio is compression. To defeat MP3 compression, comparing the original with a compressed/decompressed signal to find a safe area for embedding should be a convenient possibility [61]. On the other hand, frequency filters are other usual attacks to be taken into account. To overcome them, different areas of the spectrum could be used. To survive other attacks which distort a part of signal, repeating the secret information in different areas of the audio signal will be very useful.

1.4 Structure of the thesis

The rest of this thesis is organized as follows:

Chapter 2 contains the seven published contributions of this thesis.

In Section 2.1, the first published contribution of image data hiding [68] is presented. This paper proposes a novel high capacity reversible image data hiding scheme using a prediction technique which is effective for error resilience using the H.264/AVC standard. In the proposed method, which is based on H.264/AVC intra prediction, firstly the prediction error blocks are computed and then the error values are slightly modified through shifting the prediction errors. The modified errors are used for embedding the secret data. The experimental results show that the proposed method, called shifted intra prediction error (SIPE), is able of hiding more secret data than other schemes while the PSNR of the marked image is about 48 dB.

Section 2.2 presents the second published contribution of image data hiding [92]. This study illustrates a new lossless data hiding method for digital images using an image prediction technique. In the proposed method, the idea of shifting the prediction error is used. The gradient-adjusted prediction (GAP), which is one of the best casual predictors, is used resulting in excellent results. Predictions based on the neighbouring pixels, not neighbouring blocks, led to a narrower histogram even better than that obtained using the H264/AVC intra prediction error. Unlike H264/AVC, the GAP generates real numbers, not integer numbers, which leads to the use of intervals for embedding. This method is able to embed a huge amount of data (15-140 kbit for a 512 × 512 × 8 greyscale image) whilst the PSNR of the marked image versus the original image is very high.

In Section 2.3, the third published contribution of image data hiding [69] is presented. In this paper, we introduce a highly efficient reversible data hiding technique based on dividing the image into tiles and shifting the histograms of each image tile between its minimum and maximum frequencies. Data are then embedded into the pixel level with the largest frequency to maximise data hiding capacity. This method exploits the special properties of medical images, where the histogram of their non-overlapping image tiles mostly peak around some gray values and the rest of the spectrum is mostly empty. The zeros (or minima) and peaks (maxima) of the histograms of the image tiles

are then relocated to embed the data. The grey values of some pixels are therefore modified. High capacity, high fidelity, reversibility and multiple data insertions are the key requirements of data hiding in medical images. It is shown how the histograms of image tiles of medical images can be exploited to achieve these requirements. Compared with the data hiding method applied to the whole image, the suggested scheme can result in a 30%-200% capacity improvement with still better image quality, depending on the medical image content.

In Section 2.4, the first published contribution of audio data hiding [78] is presented. This paper describes a very efficient method for audio data hiding which is suitable for real-time applications. The FFT magnitudes which are in a band of frequencies between 5 and 15 kHz are modified slightly and the frequencies which have a magnitude lower than a threshold are used for embedding. Its low complexity is one of the most important properties of this method, making it appropriate for real-time applications. In addition, the suggested scheme is blind, since it does not need the original signal for extracting the hidden bits. The experimental results show that it has a very high capacity (5 kbps), without significant perceptual distortion and provides robustness against MPEG compression (MP3).

Section 2.5 contains the second published contribution of audio watermarking [79]. An audio watermarking technique in the frequency domain which takes advantage of interpolation is proposed. Interpolated FFT samples are used to generate imperceptible marks. The experimental results show that the suggested method has very high capacity (about 3 kbps), without significant perceptual distortion (Object Difference Grade, ODG about –0.5) and provides robustness against common audio signal processing such as echo, added noise, filtering, resampling and MPEG compression (MP3). Depending on the specific application, the tuning parameters could be selected adaptively to achieve even more capacity and better transparency.

In Section 2.6, the third published contribution of audio watermarking [80] is presented. This paper proposes a novel robust audio watermarking algorithm to embed data and extract it in a bit exact manner based on changing the magnitudes of the FFT spectrum. The key point is selecting a frequency band for embedding depending on the comparison between the original and an MP3 compressed/decompressed version of the signal and on a suitable scaling factor. The experimental results show that the method has a very high capacity (about 5 kbps), without significant perceptual distortion (ODG

about –0.25) and provides robustness against common audio signal processing such as added noise, filtering and MPEG compression (MP3). Furthermore, the proposed method has a larger capacity (number of embedded bits to number of host bits rate) than recent image data hiding methods.

In Section 2.7, the fourth published contribution of audio watermarking [81] is presented. This letter suggests a novel high capacity robust audio watermarking algorithm by using the high frequency band of the wavelet decomposition, for which the human auditory system (HAS) is not very sensitive to alteration. The main idea is to divide the high frequency band into frames and then, for embedding, the wavelet samples are changed depending on the average of the corresponding frame. The experimental results show that the method has very high capacity (about 5.5 kbps), without significant perceptual distortion (ODG in [–1 ,0] and SNR about 33 dB) and provides robustness against common audio signal processing such as added noise, filtering, echo and MPEG compression (MP3).

Finally, Chapter 3 presents the most relevant concluding remarks of this thesis, together with some possible directions for future research.

CHAPTER 2

CONTRIBUTIONS OF THE THESIS

2.1 Reversible Data Hiding Based On H.264/AVC Intra Prediction

68. **M. Fallahpour**; D. Megías. “Reversible Data Hiding Based On H.264/AVC Intra Prediction”, *International Workshop on Digital Watermarking (IWDW 2008). Lecture Notes in Computer Science* 5450, pp. 52–60, 2009. http://www.springerlink.com/content/e68001037m080825/

Reversible Data Hiding Based On H.264/AVC Intra Prediction

Mehdi Fallahpour and David Megías

Estudis d'Informàtica, Multimèdia i Telecomunicació
Universitat Oberta de Catalunya- Rambla del Poblenou, 156, Barcelona, Spain
Tel: (+34) 933 263 600, Fax: (+34) 933 568 822
E-mail: {mfallahpour,dmegias}@uoc.edu

Abstract. This paper proposes a novel high capacity reversible image data hiding scheme using a prediction technique which is effective for error resilience in H.264/AVC. In the proposed method, which is based on H.264/AVC intra prediction, firstly the prediction error blocks are computed and then the error values are slightly modified through shifting the prediction errors. The modified errors are used for embedding the secret data. The experimental results show that the proposed method, called shifted intra prediction error (SIPE), is able of hiding more secret data while the PSNR of the marked image is about 48 dB.

Keywords: Lossless data hiding, H.264/MPEG-4 AVC, intra prediction

1 Introduction

New findings of data hiding in digital imaging open wide prospects of new techniques in modern imaging science, secure communication and content management. Data hiding has been proposed as a promising technique used for security, authentication, fingerprint, video indexing, error resilient coding, etc.

H.264/AVC [1] is the newest international video coding standard providing many techniques to improve the coding efficiency of intra and inter frames. Among many new techniques, the intra prediction technique is considered as one of the most important features in the success of H.264/AVC. This technique, which is used in the proposed method, increases the dependence of the neighbouring blocks. An error resilient method that embeds information into image or video itself is another technique used in H.264/AVC. Once an error is detected, the error resilient technique extracts the hidden information and recovers the error block. Using reversible information embedding in the error resilient causes the original digital content to be completely restored in the decoder and also results in a lossless extraction of the embedded data.

Reversible data hiding [2] is a novel category of data hiding schemes. The reversibility is essential to some sensitive applications such as medical diagnosis, remote sensing and law enforcement. The methods reported in [4-10] are considered among the best schemes in lossless data hiding. In [3], a high capacity lossless data

hiding method was proposed based on the relocation of zeros and peaks of the histogram of the image blocks to embed the data. Recently, Lin and Hsueh [4] presented a reversible data hiding method based on increasing the differences between two adjacent pixels to obtain a stego-image with high payload capacity and low image distortion. Among recent lossless methods performed on the transform domain, the schemes based on the integer wavelet transform domain are more notable. Tian [5] used the integer Haar wavelet transform and embedded the secret message into high-frequency coefficients by difference expansion. Kamstra and Heijmans [6] improved Tian's method by using the information in the low-frequency coefficients to find suitable expandable differences in the high-frequency coefficients. Xuan *et al.* [8] reported the lossless embedding algorithm carried out in the integer wavelet transform domain. Xuan *et al.* [7] proposed a lossless data hiding scheme based on optimum histogram pairs in the wavelet domain. Recently, a few prediction based data hiding methods have been proposed [9-10]. Thodi et al. [9] expanded the difference between a pixel and its predicted value in the context of the pixel for embedding data. In Kuribayashi *et al.*'s algorithm [10], a watermark signal is inserted in the LSB of the difference values between pixels.

The method proposed in this paper, called SIPE, is based on increasing the differences between pixels of the cover image and their intra prediction values. The prediction error at which the number of prediction errors is at a maximum is selected to embed the message. The prediction errors larger than the selected error are increased by "1". Furthermore, the selected prediction error is left unchanged and increased by "1" if the embedded bit is "0" and "1", respectively.

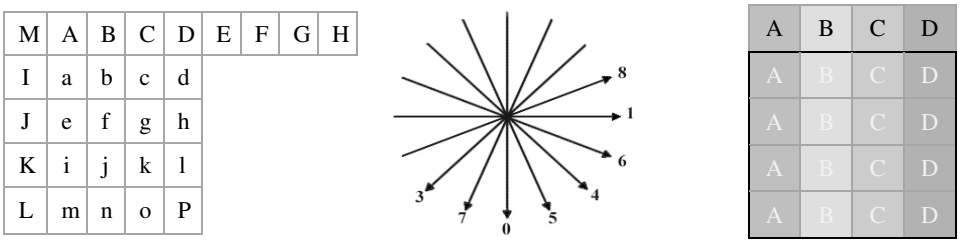
The SIPE method is able to embed a huge amount of data (15-120 kbits for a 512×512×8 greyscale image) while the PSNR of the marked image versus the original image is about 48 dB. In addition, simplicity and applicability to almost all types of images and H.264 video coding make this method superior to most of existing reversible data hiding techniques. Although the proposed lossless data hiding technique is applied to still images, it is very useful for H.264/AVC because the insertion of additional information only needs the shifting and embedding steps in coding steps. Furthermore, this lossless technique will not degrade the video quality.

2 H.264/AVC Intra prediction

In H.264/AVC intra prediction method [11], a prediction block is formed based on previously reconstructed blocks. There are nine prediction modes for each 4×4 block. The sixteen elements in the 4×4 block (labelled from *a* to *p* in Fig. 1.(a)) are predicted by using the boundary pixels of the upper and left blocks which are previously obtained (labelled from *A* to *M*). These boundary elements are therefore available in the encoder and decoder to form a prediction reference.

For each 4×4 block, one of nine prediction modes can be selected by the encoder. In addition to DC prediction type, numbered as mode 2, where all elements are predicted by $(A + B + C + D + I + J + K + L)/8$, eight directional prediction modes are specified as shown in Fig. 1.(b). For mode 0 (vertical prediction), the elements above the 4×4 block are copied into the prediction block as indicated by arrows, Fig. 1 (c).

Other modes copy adjacent pixels into the prediction block based on their prediction directions.



(a) (b) (c)

Fig. 1. 4×4 intra prediction mode (a) labeling of prediction samples. (b) 4×4 intra prediction mode direction. (c) Vertical (mode 0) prediction

The rate distortion optimisation (RDO) technique [12] is used to take full advantage of the mode selection regarding maximising coding quality and minimising data bits. The RDO is applied to all of the 4×4 block intra-prediction modes to find the best one. This approach can achieve the optimal prediction mode decision. The only drawback for using RDO is the computational complexity. Recently, there is more focus on developing the 4×4 intra-prediction mode decision techniques with lower complexity.

3 Suggested scheme

The SIPE method consists of an embedding and a extracting procedure. The embedding process includes both computing the prediction errors and embedding the information bits in the shifted prediction errors. Moreover, the data extraction is the reverse of data embedding. The proposed method is explained in the following two subsections.

3.1 Embedding

The embedding algorithm is as follows.

1. The prediction type is selected. It can be selected by RDO or other 4×4 intra-prediction mode decision techniques.
2. The prediction blocks are computed from the cover image by using an intra prediction algorithm (as described above). For the blocks without upper or left blocks, the coder uses their upper or left pixels for prediction.
3. The prediction error (PE) blocks are calculated by subtracting the predicted blocks from the cover image block, $e = I - P$.
4. The number of prediction errors in PE blocks equal to d is denoted by $D(d)$. The value M is found such that $D(M)$ is at a maximum. The following steps

(5-6) are carried out for each 4×4 block completely and then iterated for the next block.

5. In the shifting stage, the modified PE block is derived from the PE block by this approach: for each PE block element $e_{i,j}$ (expect top-most row and the left-most column of the cover image; $i \neq 1$ and $j \neq 1$), if $e_{i,j}$ is larger than M, then the modified PE $e'_{i,j}$ equals $e_{i,j} + 1$, otherwise $e'_{i,j} = e_{i,j}$.
6. In the embedding stage, each $e'_{i,j}$ $(i \neq 1 \text{ and } j \neq 1)$ with a value of M is increased by one if the corresponding bit of the data (to be embedded) is one, otherwise it will not be modified. After concealing data in $e'_{i,j}$, the embedded PE $e''_{i,j}$ is obtained.
7. Finally, the marked image I' is achieved by $I' = P + e''$.

In fact, the pixels in the top-most row and the left-most column of a cover image are preserved without carrying any hidden data. These pixels are used for recovering the original image and extracting the embedded data from the marked image. These row and column are the same in both the cover and the marked images. It is worth mentioning that, in the coder and the decoder, the raster scan order is used. The gray value of M, the prediction mode and the block size will be treated as side information that needs to be transmitted to the receiving side for data retrieval.

3.2 Detection

The following process is used for extracting the secret message from a marked image and also losslessly recovering the cover image. Let the marked image I' be the received image at the decoder. The following steps (1-4) are carried out for each block completely and then iterated for the next block.

1. The prediction block P of I can be obtained by using the intra prediction algorithm using its upper and left blocks which have been already restored.
2. If the embedded PE block element, $e''_{i,j} = I'_{i,j} - P_{i,j}$, is equal to $M + 1$, it is concluded that the embedded bit is "1". In this case, $e''_{i,j}$ should be decreased by one to obtain the modified PE block element, $e'_{i,j} = e''_{i,j} - 1$. If $e''_{i,j}$ is equal to M the embedded bit is "0" and $e'_{i,j} = e''_{i,j}$, otherwise there is no embedded data bit and again $e'_{i,j} = e''_{i,j}$.
3. If $e'_{i,j} > M$, then the prediction error $e_{i,j}$ is calculated by decreasing $e'_{i,j}$ by one, $e_{i,j} = e'_{i,j} - 1$, otherwise $e_{i,j} = e'_{i,j}$.
4. Finally, the $e_{i,j}$ should be added to prediction value $P_{i,j}$ to recover the original cover image pixel, $I_{i,j} = P_{i,j} + e_{i,j}$.

Fig. 2 shows an example of a 4×4 block greyscale image. The encoder scans the cover image block, Fig. 2(c1) pixel by pixel and subtracts the vertical prediction pixels, Fig. 2(c2), from the cover image pixels. In the PE (prediction error) block, Fig. 2(c3), following the computation of all prediction blocks the obtained M is equal to 0 and D(M) in the current block equals 6. Suppose that the bit stream to be embedded is

101001. The encoder scans the PE block and all elements larger than 0 are increased by one, Fig. 2(c4), then modified prediction errors equal to 0 are chosen for embedding data.

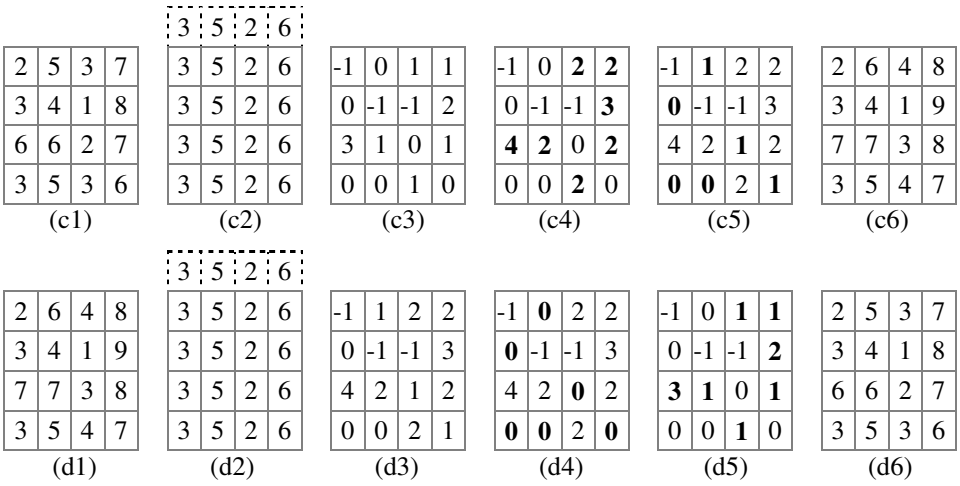


Fig. 2. Embedding steps (c1) - (c6) , Detection steps (d1) – (d6)

If the corresponding bit of the secret data is one, the modified prediction value is added by one, otherwise it will not be modified, as shown in Fig. 2(c5). The marked image, Fig. 2(c6), is obtained by adding the embedded prediction errors, Fig. 2(c5), to the prediction pixels, Fig. 2(c2). As described above, the value *M* (in this example equals to zero) and the prediction mode are treated as side information. The decoder extracts the secret data and obtains the original image block from the marked image block, Fig. 2(d1). The prediction block is computed based on the restored cover image blocks, Fig. 2(d2). The decoder scans the embedded PE block, Fig. 2(d3), which is obtained by subtracting the prediction block from the marked block. If the embedded PE element is equal to 1 and 0, the embedded data bit is "1" and "0", respectively. In case the embedded PE is equal to "1" or "0", the modified PE Fig. 2(d4), equals "0", otherwise it equals the embedded PE. In order to get PE, Fig. 2(d5), if the modified PE is larger than 0, it must be decreased by one, otherwise the PE equals the modified PE. Finally, restored cover image pixel, Fig. 2(d6), is computed by adding PE, Fig. 2(d5), to the prediction pixel, Fig. 2(d2).

This example clarifies the simplicity and reversibility of this method. In fact we have three steps in embedding and detecting: calculating prediction error, shifting and embedding. In the detector the prediction block is calculated based on restored cover image blocks and, thus, the prediction blocks in the embedding and detecting procedures are the same. In the encoder, the marked image is achieved by adding the embedded PE to the prediction image block and then, in the decoder, the embedded PE is achieved by subtracting the prediction block from the marked image. When we have the embedded PE block, the modified PE and the PE can be obtained very easily by decreasing the value of the block pixels. Finally, adding PE to the prediction block results in the restored cover image. Hence, all steps are reversible.

4 Experimental results

The SIPE algorithm was implemented and tested on various standard test images from the UWaterloo database [13]. Also, the results of Lin and Hsueh's [4], Kamstra and Heijmans's [6], and Xuan et al's. [7] methods were compared with the results obtained with this method. This comparison shows that the SIPE method is able of hiding more secret data than almost all methods mentioned in the literature for the same (above 45dB) PSNR. The experimental results of the SIPE method show that the embedded data remains invisible, since no visual distortion can be revealed. It is worth pointing out that the embedded data were generated by the random number generator in MATLAB on a personal computer.

Table 1 summarizes the experimental results obtained by this method for the Mandrill, Lena, Peppers, Zelda, Goldhill, Barbara, and Boat images using vertical prediction. For the sake of brevity, only the simple case of one maximum of $D(d)$ is described, because the general cases of multiple maxima of $D(d)$ can be decomposed as a few maximum cases.

The payload capacity is increased by using multiple maximum of $D(d)$. In the Right Shifted type, M is equal to 0 and the prediction errors larger than 0 are increased by one. In the Left Shifted type, M is equal to –1 and the prediction errors smaller than 0 are reduced by one. In addition, in the L&R Shifted type, Left Shifted and Right Shifted types are used simultaneously.

Table 1. PSNR (dB) and payload capacities (bits) of the test images of the UWaterloo database [13] with different block size ($N \times N$)

Shift type		Right shifted $N = 1$	Right Shifted $N = 4$	L&R shifted $N = 1$	L&R shifted $N = 4$	Ratio
Mandrill	PSNR	51.13	51.1	48.3	48.24	2.12
	Payload	7630	6137	15088	12031	
Lena	PSNR	51.12	51.16	48.58	48.49	2.58
	Payload	28932	22616	52664	42334	
Peppers	PSNR	51.13	51.17	48.49	48.43	2.70
	Payload	21158	17810	40929	34775	
Zelda	PSNR	51.16	51.13	48.61	48.48	2.75
	Payload	27120	19977	53474	39201	
Goldhill	PSNR	51.26	51.34	48.43	48.36	2.63
	Payload	17431	13063	34492	25948	
Barbara	PSNR	51.15	51.19	48.51	48.40	2.33
	Payload	22409	15680	43682	31167	
Boat	PSNR	51.12	51.06	48.59	48.49	2.60
	Payload	27522	21735	52543	41591	

The experimental results of the L&R Shifted type exhibit that the PSNR of all marked images is above 48 dB.

Although in the description of the method only 4×4 blocks were used for embedding and detecting data, other block sizes could be easily used. Fig. 3 shows how the performance of the proposed scheme varies with different block sizes: 1×1, 2×2, 3×3, 4×4, and 8×8. As shown in the figure, the performance of the proposed scheme is increased when the block size is the smallest one (1×1). The smaller the block size, the larger the amount of the embedding information and the more computation time is required. Since in most of image data hiding applications the computation time is not critical, using 1×1 blocks to achieve high capacity and PSNR is desirable. In Table 1, the "Ratio" is the ratio of computation time for embedding process using 1×1 blocks to embedding time using 4×4 blocks. In H.264/AVC, using 4×4 blocks is more effective because the frames are divided into 4x4 blocks for coding and decoding and, furthermore, the execution time is a critical issue in video coding.

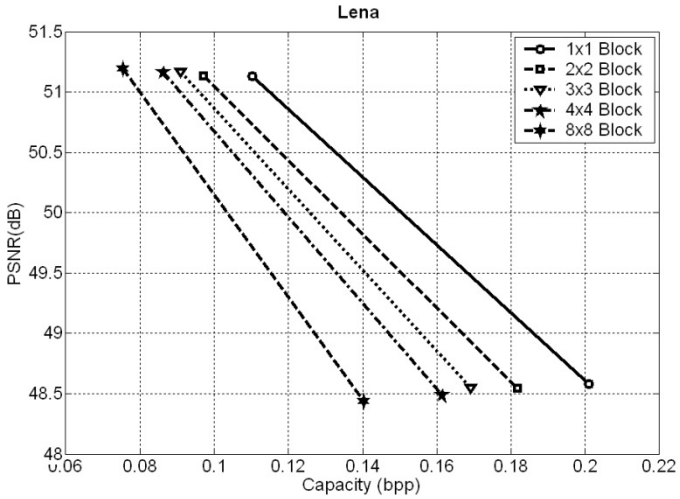


Fig 3. Comparison between the embedding capacity in bpp and distortion in PSNR with different block sizes for the Lena image.

Fig. 4 illustrates the performance comparison of the SIPE with the methods reported in [4], [6], and [7] for the Lena image in terms of PSNR (dB) and payload (bpp: bits per pixel). As shown in Fig 4, the SIPE scheme provides high enough bound of the PSNR (above 48dB) with a quite large data embedding capacity, indicating a fairly better performance of the SIPE method.

It is worth to mentioning that the selection of the prediction mode is an important step. Fig. 5 demonstrates the effects of the selection mode on capacity and distortion. As an example for the Lena image, vertical prediction has higher capacity with lower distortion.

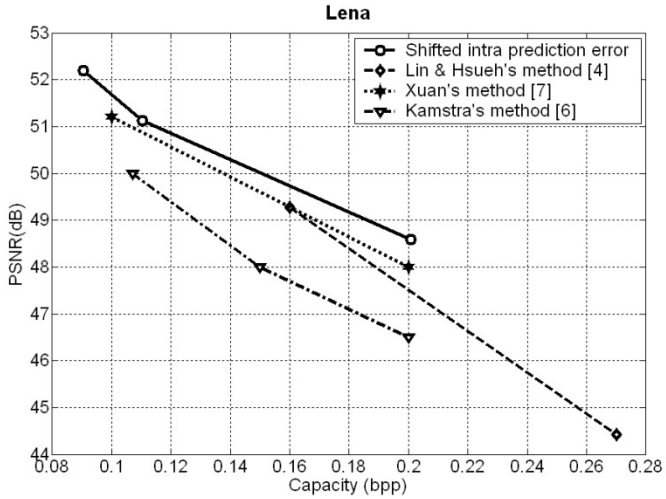


Fig. 4. Comparison between the embedding capacity (bpp) and distortion (PSNR) for the Lena image

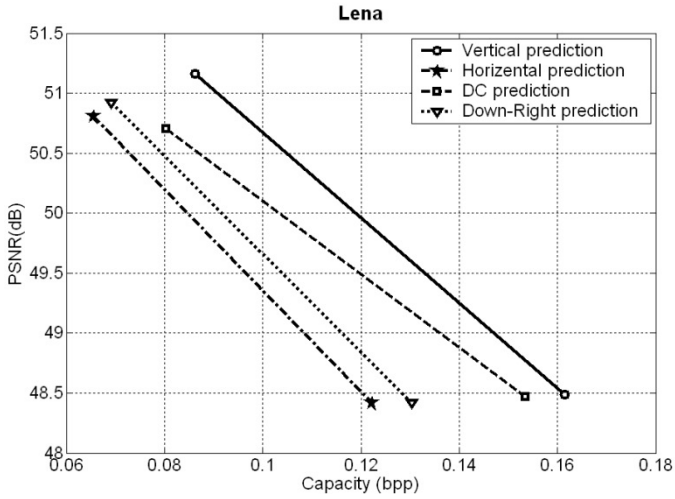


Fig. 5. Comparison of embedding capacity versus distortion with different prediction modes for the Lena image

5 Conclusion

This paper presents a novel high-capacity reversible data hiding algorithm, called shifted intra prediction error (SIPE), which is based on shift differences between the cover image pixels and their predictions. Large capacity of embedded data (15-120 kbits for a 512×512 greyscale image), PSNR above 48 dB, applicability to almost all types of images, simplicity and short execution time are the key features of this

algorithm. The SIPE method is applicable for error resilient solutions in H.264 advanced video coding. Therefore, the SIPE method has several advantages with respect to the methods reported in [4], [6] and [7], in which the suggested algorithms are considered among the best methods in lossless data hiding.

Acknowledgement

This work is partially supported by the Spanish Ministry of Science and Innovation and the FEDER funds under the grants TSI2007-65406-C03-03 E-AEGIS and CONSOLIDER-INGENIO 2010 CSD2007-00004 ARES.

References

1. ITU-T Rec. H.264/ISO/IEC 14496-10, "Advanced Video Coding," Final Committee Draft, Document JVTG050, Mar. 2003
2. Shi, Y.Q., Ni, Z., Zou, D., Liang, C., and Xuan, G., "Lossless data hiding: Fundamentals, algorithms and applications," in Proc. IEEE Int. Symp. Circuits Syst., Vancouver, BC, Canada, , vol. II, pp. 33–36, 2004
3. Fallahpour, M., Sedaaghi M.H., "High capacity lossless data hiding based on histogram modification," IEICE Transactions on Electronics Express Vol. 4, No. 7 pp.205-210, 2007
4. Lin , C.C, and Hsueh N.L, "Hiding Data Reversibly in an Image via Increasing Differences between Two Neighboring Pixels," IEICE TRANS. INF. & SYST., Vol.E90–D, NO.12 pp 2053-2059, Dec. 2007
5. Tian, J., "Reversible data embedding using a difference expansion," IEEE Transactions on Circuits and Systems for Video Technology, pp.890-896, 2003.
6. Kamstra, L., Heijmans, H. J.A.M., "Reversible data embedding into images using wavelet techniques and sorting", IEEE transactions on image processing vol. 14, no. 12, pp.2082-2090, 2005
7. Xuan, G., Shi,Y. Q., Chai, P., Cui, X., Ni, Z., Tong, X., "Optimum Histogram Pair Based Image Lossless Data Embedding," Proc. International Workshop on Digital Watermarking (IWDW07), Guangzhou, China, 2007
8. G. Xuan, Y. Q. Shi, C. Yang,, Y Zheng, D. Zou, P. Chai, "Lossless data hiding using integer wavelet transform and threshold embedding technique," IEEE International Conference on Multimedia & Expo (ICME05), Amsterdam, Netherlands, July 6-8, 2005
9. D.M. Thodi and J.J. Rodriguez, "Expansion embedding techniques for reversible watermarking," *IEEE Trans. Image Process*., vol.16, no.3, pp.723–730, 2007
10. M.Kuribayashi, M.Morii, and H.Tanaka, "Reversible watermark with large capacity based on the prediction error expansion," *IEICE Trans. Fundamentals*, vol.E91-A, no.7, pp.1780-1790, Jul. 2008
11. Richardson, I. E. G., "H.264 and MPEG-4 Video Compression", Wiley, pp. 120-145, 2003.
12. Sullivan, G. J. and Wiegand, T., "Rate-distortion optimization for video compression," IEEE Signal Process. Mag., vol. 15, pp. 74–90, Nov. 1998
13. Waterloo Repertoire GreySet2, http://links.uwaterloo.ca/greyset2.base.html. Last checked on October 27th , 2008

2.2 Reversible image data hiding based on gradient adjusted prediction

92. **M. Fallahpour**, “Reversible image data hiding based on gradient adjusted prediction”, *IEICE Electron. Express*, Vol. 5, No. 20, pp.870-876, (2008). (Impact factor = 0.48 in 2008). http://www.jstage.jst.go.jp/article/elex/5/20/5_870/_article

Reversible image data hiding based on gradient adjusted prediction

Mehdi Fallahpour
Fallahpour@Gmail.com

Abstract: The present study illustrates a new lossless data hiding method for digital images using image prediction technique. In the proposed method which is based on gradient-adjusted prediction (GAP), first prediction errors are computed and then the error values are slightly modified through shifting the prediction errors. The modified errors are used for embedding the data. Experimental results of present research have demonstrated that the proposed method called shifted gradient-adjusted prediction error (SGAPE) is capable of hiding more secret data with absolutely high PSNR.

Keywords: Gradient-adjusted prediction, reversible data hiding.
Classification: Science and engineering for electronics

References

1. Y. Q. Shi, Z. Ni, D. Zou, C. Liang, and G. Xuan, "Lossless data hiding: Fundamentals, algorithms and applications," in *Proc. IEEE Int. Symp. Circuits Syst.*, Vancouver, BC, Canada, May 2004, vol. II, pp. 33–36.
2. Ni Zhicheng, Y.Q. Shi, N.Ansari, and W. Su, "Reversible data hiding," *IEEE Trans. on Circuits and Systems for Video technology*, 16(3):354–362, March 2006.
3. M.Fallahpour, M.H. Sedaaghi, "High capacity lossless data hiding based on histogram modification," *IEICE Transactions on Electronics Express* Vol. 4, No. 7 pp.205-210, April 2007.
4. C.C Lin, N.L Hsueh, "Hiding Data Reversibly in an Image via Increasing Differences between Two Neighboring Pixels," *IEICE TRANS. INF. & SYST.*, Vol.E90–D, NO.12 pp 2053-2059, December 2007.
5. X. Wu, N. Memon, "Context-based, Adaptive, Lossless Image Codec" *IEEE Trans. On Communications*, vol. 45, no. 4, 1997.
6. UWaterloo Image Database [Online]. Available: http://links.uwaterloo.ca/greyset2.base.html

[DOI: **.****/**********]
Received July **, 2003;
Accepted July **, 2003;
Published in August **, 2003 issue

1. Introduction

In parallel with the development of information technologies, security becomes an important issue. Data hiding methods can conceal additional information in media. Most data hiding schemes distort the cover media in order to insert the secret data. Although the distortion is often small and imperceptible to human visual system, the reversibility is crucial to some sensitive applications, such as medical diagnosis, remote sensing and law enforcement. Therefore, it is desired to invert the marked media back to the original cover media after the hidden data have been retrieved.

Reversible data hiding is a novel category of data hiding schemes. For a survey on reversible methods, readers are referred to [1]. Ni et al. [2] introduced the lossless data embedding algorithm based on the spatial domain histogram shifting. In [3] a high capacity lossless data hiding method was proposed based on the relocation of zeros and peaks of the histogram of image blocks to embed the data. Recently, Lin and Hsueh [4] presented a reversible data hiding method based on increasing the differences between two adjacent pixels to obtain a stego-image with high payload capacity and low image distortion.

The proposed method, shifted gradient-adjusted prediction error (SGAPE), is based on increasing the differences between pixels of cover image and their prediction values. The prediction error at which the number of prediction errors is at a maximum is selected to embed the message. The prediction errors larger than the selected error are increased by "1". Furthermore, the selected prediction error is left unchanged and increased by "1" if the embedded bit is "0" and "1", respectively.

SGAPE method is able to embed a huge amount of data (15-140 kb for a 512 x 512 x 8 grayscale image) while the PSNR of the marked image versus the original image is very high. In addition, simplicity, short execution time and applicability to almost all types of images make this method superior than most of existing reversible data hiding techniques.

2. The proposed method

SGAPE method contains embedding and extracting procedures. The embedding process includes both computing the prediction errors and embedding the information bits in the shifted prediction errors. However, the data extraction is the reverse of the data embedding. The proposed method is explained in the three following subsections.

2.1 The Prediction Algorithm

The gradient variations of the neighboring pixels are used for estimating the pixel value. The gradient-adjusted prediction (GAP) algorithm [5] operates on seven neighbors of the current pixel of a cover image $I_{i,j}$. By applying GAP prediction for $I_{i,j}$, its predictive value $\hat{I}_{i,j}$ can be computed as follows:

$$d_h = \left| I_{i-1,j} - I_{i-2,j} \right| + \left| I_{i,j-1} - I_{i-1,j-1} \right| + \left| I_{i,j-1} - I_{i+1,j-1} \right|$$

$$d_v = \left| I_{i-1,j} - I_{i-1,j-1} \right| + \left| I_{i,j-1} - I_{i,j-2} \right| + \left| I_{i+1,j-1} - I_{i+1,j-2} \right|$$

[DOI: **.****/**********]
Received July **, 2003;
Accepted July **, 2003;
Published in August **, 2003 issue

IF ($d_v - d_h > 80$) {sharp horizontal edge} $\hat{I}_{i,j} = I_{i-1,j}$
ELSE IF ($d_v - d_h <$ -80) {sharp vertical edge}$\hat{I}_{i,j} = I_{i,j-1}$
ELSE {
$\hat{I}_{i,j} = (I_{i-1,j} + I_{i,j-1})/2 + (I_{i+1,j-1} - I_{i-1,j-1})/4$
IF ($d_v - d_h > 32$) {horizontal edge} $\hat{I}_{i,j} = (\hat{I}_{i,j} + I_{i-1,j})/2$
ELSE IF ($d_v - d_h > 8$) {weak horizontal edge} $\hat{I}_{i,j} = (3\hat{I}_{i,j} + I_{i-1,j})/4$
ELSE IF ($d_v - d_h <$ - 32) {vertical edge} $\hat{I}_{i,j} = (\hat{I}_{i,j} + I_{i,j-1})/2$
ELSE IF ($d_v - d_h <$ - 8) {weak vertical edge} $\hat{I}_{i,j} = (3\hat{I}_{i,j} + I_{i,j-1})/4$
}
The GAP predictor results in a new image with predicted pixel values.

2.2 Embedding

The embedding algorithm is as follows:

1) Prediction image is computed from the cover image by GAP algorithm (as described above).
2) The prediction error (PE) matrix elements are calculated by subtracting the predicted image from the cover image, $e_{i,j} = I_{i,j} - \hat{I}_{i,j}$.
3) The number of prediction errors inside the interval [d , d+1) is denoted by D(d). S value is found such that D(S) is at a maximum. As the GAP is a good predictor, in most images S value is equal to zero.
4) To prevent from overflow and error in extracting the embedded data, the position of all pixels with value 255 are recorded as side information. Also steps 5 and 6 are carried out for elements with $I_{i,j}$<255.
5) In shifting stage, the modified PE matrix is derived from the PE matrix by this approach: For every $e_{i,j}$ $(i > 2 \; and \; j > 2)$, if $e_{i,j}$ is larger or equal to S+1, then the modified PE $e'_{i,j}$ equals $e_{i,j} + 1$, otherwise $e'_{i,j} = e_{i,j}$.
6) In embedding stage, each $e'_{i,j}$ $(i > 2 \; and \; j > 2)$ inside the interval [S , S+1) is increased by one if the corresponding bit of the data (to be embedded) is one, otherwise it will not be modified. After concealing data to $e'_{i,j}$, embedded PE $e''_{i,j}$ is obtained.
7) Finally, marked image pixel $I'_{i,j}$ is achieved by $I'_{i,j} = \hat{I}_{i,j} + e''_{i,j}$. If $I_{i,j}$=255 then $I'_{i,j}$=255.

In fact, the pixels in the two top-most rows and the two left-most columns of a cover image are preserved without carrying any hidden data. These rows and columns are the same in cover and marked images. Thus, the first pixel that can hold the secret message is at position (3,3) of the marked image and all scan processes in coder and decoder should start from this pixel. It is worth mentioning that in coder and decoder raster scan order is used. The gray value of S and position of all pixels with value 255 will be treated as side information that needs to be transmitted to the receiving side for data retrieval.

2.3 Detection

The following process is used for extracting secret message from a marked image and lossless recovery of the host image and. Let the marked image $I'_{i,j}$ be the received image at the decoder.

[DOI: **.****/**********]
Received July **, 2003;
Accepted July **, 2003;
Published in August **, 2003 issue

1) As the pixels in the two top-most rows and the two left-most columns do not carry any secret data, we can readily restore them by $I_{i,j} = I'_{i,j}$ for i=1,2 or j=1,2. Beginning from pixel $I'_{3,3}$, the following steps (2-5) are performed for each pixel completely and then iterated for the next pixel. If $I_{i,j}$ was recorded as side information then $I_{i,j} = I'_{i,j}$ and steps (2-5) are performed for next pixel.

2) The prediction pixel value $\hat{I}_{i,j}$ of $I_{i,j}$can be obtained by GAP algorithm with its seven adjacent pixels which have been already restored.

3) If embedded prediction error value, $e''_{i,j} = I'_{i,j} - \hat{I}_{i,j}$, is inside the interval [S+1 , S+2) , then it is concluded that the embedded data bit is "1". In this case, $e''_{i,j}$ should be decreased by one to obtain modified prediction error value, $e'_{i,j} = e''_{i,j} - 1$. If $e''_{i,j}$ is inside the interval [S , S+1), then the embedded data bit is "0" and $e'_{i,j} = e''_{i,j}$, Otherwise, there is no embedded data bit and again $e'_{i,j} = e''_{i,j}$.

4) If $e'_{i,j}$ is larger or equal to S+2 then prediction error value $e_{i,j}$ is calculated by decreasing $e'_{i,j}$ by one, $e_{i,j} = e'_{i,j} - 1$, otherwise $e_{i,j} = e'_{i,j}$.

5) Finally, the $e_{i,j}$ should be added to prediction value $\hat{I}_{i,j}$ to get original cover image pixel, $I_{i,j} = \hat{I}_{i,j} + e_{i,j}$.

Fig. 1 shows an example of a 5×5 grayscale image. The encoder scans the cover image in Fig. 1(c1) pixel by pixel and subtracts the prediction pixels, Fig. 1(c2), from the cover image pixels. In the prediction error matrix, Fig. 1(c3), the S can be found which in this example, S = 0 and D(S) = 3. The two top-most rows and the two left-most columns are not used for embedding. Suppose that the bit stream to be embedded is 101. The encoder scans the prediction error matrix and all values larger or equal to 1 are increased by one, Fig. 1(c4). The elements obtained from the previous stage which are inside the interval [0, 1) are chosen for embedding data. If the corresponding bit of the secret data is one, the modified prediction value is added by one, otherwise it will not be modified, Fig. 1(c5). Marked image, Fig. 1(c6), is obtained by adding embedded prediction errors, Fig. 1(c5), to prediction pixels, Fig. 1(c2). As said before S, in this case equals to zero, is treated as side information.

It is already explained that the pixels with $i > 2$ *or* $j > 2$ are the same in marked and cover images. The decoder scans the marked image, Fig. 1(d1), starting from pixel at position (3,3), and does all steps pixel by pixel as follows. Based on restored cover image pixels, Fig. 1(d2), the prediction pixel value is computed, Fig. 1(d3). If the embedded PE, Fig. 1(d4), which is obtained by subtracting prediction pixel from marked image pixel, is inside the interval [0 , 1), the embedded data bit is "0" and modified PE is equal to embedded PE. In case the embedded PE is inside the interval [1 , 2), the embedded data bit is "1" and to get the modified PE, the embedded PE should be decrement. In order to get PE, Fig. 1(d6), if modified PE is larger or equal to "2", it has to be decreased by one, and otherwise PE equals modified PE. Finally, restored cover image pixel, Fig. 1(d2), is computed by adding PE, Fig. 1(d6), to prediction pixel, Fig. 1(d3).

[DOI: **.****/**********]
Received July **, 2003;
Accepted July **, 2003;
Published in August **, 2003 issue

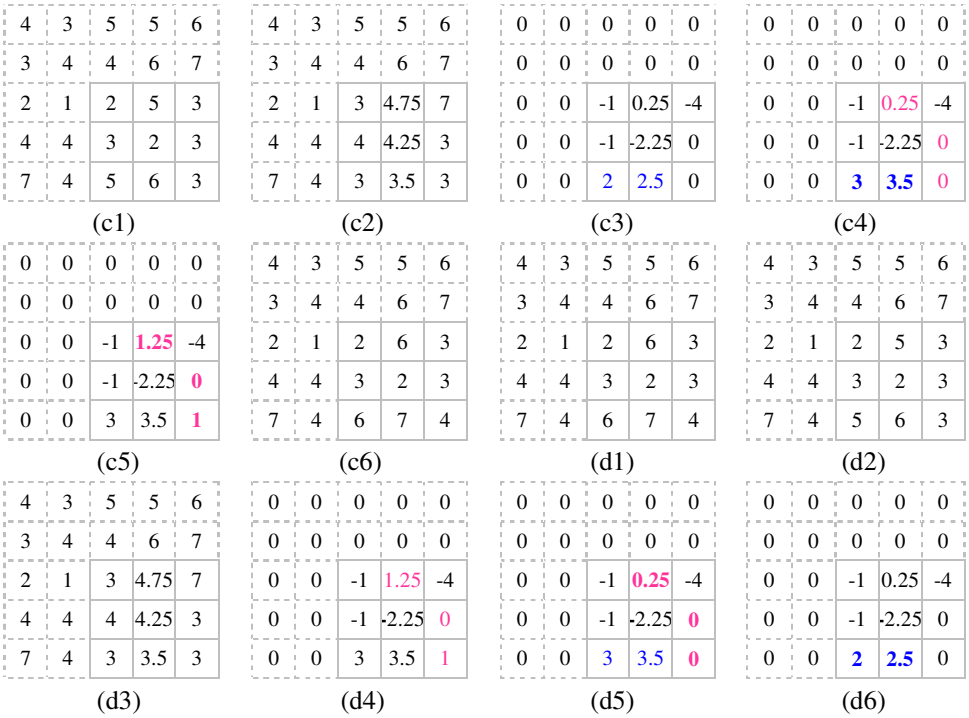


Fig. 1. Embedding steps (c1) - (c6) and Detection steps (d1) – (d6)

3. Experimental results and evaluations

The Ni et al.'s[2], Fallahpour and Sedaaghi's (F&S's)[3] , Lin and Hsueh's [4] and SGAPE algorithms was implemented and tested on various general test images of UWaterloo database[6].

The comparison between SGAPE and all methods mentioned in this literature proves that SGAPE method is capable of hiding more secret data than almost all compared methods at the same (above 40dB) PSNR. The experimental results of SGAPE method show that the embedded data remains invisible, besides no visual distortion can be revealed. It is noteworthy that the embedded data was generated by the random number generator in MATLAB.

Table I, summarizes the experimental results obtained by SGAPE method. In Right Shifted type, interval [0 , 1) is used for embedding data and the prediction errors larger or equal to "1" are increased by one. In RL Shifted type, after right shifting, left shifting is used. In left shifting, the prediction errors smaller than "0" are reduced by one and interval [-2, -1) are used for embedding. The payload capacity is enhanced and the distortion is increased by increasing numbers of interval [S , S+1).

Fig. 2 illustrates the performance comparison of SGAPE with the methods reported in [2], [3] and [4] for Lena, Barbara, Baboon, and Boat images in terms of PSNR and payload (bpp: bits per pixel). As shown in Fig. 2 SGAPE scheme provides high enough bound of the PSNR (above 40dB) with a quite large data embedding capacity, indicating fairly better performance of SGAPE method. Fig. 2 confirms better performance of gradient-adjusted shifted prediction error method.

[DOI: **.****/**********]
Received July **, 2003;
Accepted July **, 2003;
Published in August **, 2003 issue

Table I. PSNR (dB) and payload (bits) of the test images of UWaterloo database.

Shift type		Right Shifted	RL shifted	RLR shifted	RLRL shifted
Lena	Payload	28971	57949	83937	108540
	PSNR	52.1	49.2	45.9	44.1
Barbara	Payload	23816	47363	68743	89296
	PSNR	51.9	49	45.5	43.7
Boat	Payload	28941	56850	82322	105900
	PSNR	52.1	49.2	45.8	44
Goldhill	Payload	20837	41466	60207	79110
	PSNR	52	48.9	45.4	43.5
Baboon	Payload	9629	19277	28584	37678
	PSNR	51.5	48.5	44.7	42.7
Peppers	Payload	22841	45669	66106	86373
	PSNR	51.9	49	45.5	43.6
Zelda	Payload	28906	57908	84636	110798
	PSNR	52.2	49.2	45.9	44.2

It is observed that this scheme provides high enough bound of the PSNR with a quite large data embedding capacity.

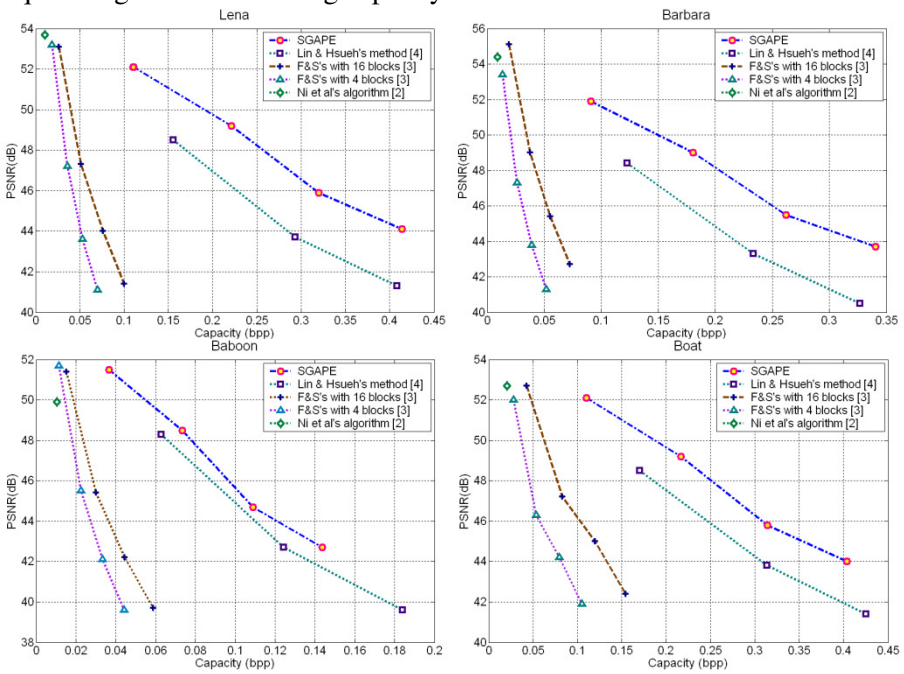


Fig. 2. Comparison among methods in [2], [3], [4] and SGAPE for Lena, Barbara, Baboon, and Boat images.

4. Conclusion

This paper presents a novel high-capacity reversible data hiding algorithm called shifted gradient-adjusted prediction error (SGAPE) which is based on shift differences between cover image pixels and their predictions. Large capacity of embedded data (15-140 kb for a 512 x 512 grayscale image), very high PSNR, applicability to almost all types of images, simplicity and short execution time are key features of this algorithm. Therefore, SGAPE method has advantages to the methods reported in [2] and [4] where used algorithms are considered as among the best methods in lossless data hiding.

[DOI: **.****/**********]
Received July **, 2003;
Accepted July **, 2003;
Published in August **, 2003 issue

Acknowledgment

Much gratitude to Fatemeh Zahra for her support and special thanks to Behi for proofreading the article.

[DOI: **.****/**********]
Received July **, 2003;
Accepted July **, 2003;
Published in August **, 2003 issue

2.3 High capacity, reversible data hiding in medical images

69. M. Fallahpour, D. Megias, and M. Ghanbari, “High capacity, reversible data hiding in medical images," in IEEE International Conference on Image Processing (ICIP2009). Los Alamitos, CA, USA: IEEE Computer Society, 2009, in press.

It is worth pointing out that the International conference on image processing is in a category A in the CORE Computer Science Conference Rankings[87].

HIGH CAPACITY, REVERSIBLE DATA HIDING IN MEDICAL IMAGES

M. Fallahpour[1], D.Megias[1] and M. Ghanbari, FIEEE[2]
[1] Universitat Oberta de Catalunya, Spain
[2] University of Essex , United Kingdom

ABSTRACT

In this paper we introduce a highly efficient reversible data hiding technique. It is based on dividing the image into tiles and shifting the histograms of each image tile between its minimum and maximum frequency. Data are then inserted at the pixel level with the largest frequency to maximize data hiding capacity. It exploits the special properties of medical images, where the histogram of their non-overlapping image tiles mostly peak around some gray values and the rest of the spectrum is mainly empty. The zeros (or minima) and peaks (maxima) of the histograms of the image tiles are then relocated to embed the data. The grey values of some pixels are therefore modified.
High capacity, high fidelity, reversibility and multiple data insertions are the key requirements of data hiding in medical images. We show how histograms of image tiles of medical images can be exploited to achieve these requirements. Compared with data hiding method in the whole image, our scheme can result in 30%-200% capacity improvement with still better image quality, depending on the medical image content.

Index Terms— data hiding, Image watermark, medical images

1. INTRODUCTION

Data hiding is the insertion of a message into a host document or cover media. In applications where additional information is required to describe another information media, such process can be very useful. For instance, in medical images, patients' details and the doctors' views can be inserted into the medical images to form a comprehensive data bank. However, data hiding in medical images, due to their specific requirements impose certain constraints, which set some specific requirements. In fact high quality (fidelity), authentication, high capacity, frequent insertions and reversibility are the main requirements of medical files. Various kinds of data hiding for medical images that may meet some but not all the requirements can be categorized into three requirements of high quality, reversibility and high capacity.
To preserve high quality, one may embed information in the region of non-interest (RONI) [1]. The main drawback of this method is the ease of introducing copy attack on the non-watermarked regions. Various experiments suggest that RONI corresponds in general to the black background of the image, but sometimes RONI can include gray-level parts of little interest [2], thus leaving some area for embedding on the gray level image itself. For the reason that there is no interference with the invisibility, image content is less strict; consequently one can revert to methods with higher robustness and capacity [3]. Another medical image watermarking system embeds information in bit planes, which results in stego images with very low normalized root mean square errors (NRMSE), indicating that the watermark is practically invisible [4]. On reversible data hiding, where the embedded content can be added or removed without affecting the original image quality, [5], a vast attempt has been recently provided. However the capacity is still way below the embedding capacity of nonreversible data hiding technique. However, if capacity is of prime importance, then quality can be sacrificed for capacity. For instance, the embedded data may replace some image details such as the least significant bit of the image [6] or details are lost after lossy image compression [7]. For a survey on medical watermarking application, the readers may refer to [8].
Perhaps the histogram-shifted-based lossless data hiding algorithm proposed by Ni et. al [9] is one the most capacity efficient data hiding system that suits medical images well. Since in this method, at most the intensity of all the watermarked pixels are shifted by one quantum level, then for an 8-bit image with the mean squared error (MSE) of 1, the PSNR of the watermarked image, at the worse case PSNR=10xLog10(255x255/MSE)=48.13dB, which is regarded a very high quality and is suitable for medical images. In this paper we show how by applying shifted-histogram, not only the watermarked image quality can be improved, but more importantly, the data hiding payload can be significantly increased.
The rest of the paper is as follows. Characteristics of the proposed algorithm and its details are described in Section II. Experimental results are presented in Section III, and conclusions are drawn in Sections IV.

2. PROPOSED METHOD

The main idea in the shifted-histogram data hiding method is to find a pair of maximum and minimum in the image pixel intensity histogram and then shift the intensity of those pixels within the maximum and minimum frequency range

by one level, towards the minimum frequency level. This creates an empty space on the shifted histogram at the vicinity of the maximum pixel density. To embed a data stream, the modified image is re-scanned and when the pixel of maximum frequency is encountered if the corresponding bit in the embedding stream is "1" its gray level is incremented by one level otherwise it is unaltered. Thus the maximum number of bits that can be hidden into the image is equal to the maximum frequency of the original histogram. Due to the created gap, the data hiding mechanism is reversible. The values of the pixels with maximum and minimum frequency are also recorded as side information. If the minimum frequency is non-zero, then their numbers also need to be embedded as the side information, which reduces the data hiding capacity of the system.

Although Ni et. al. have shown that their algorithm for a vast variety of images outperforms almost all the known reversible data hiding methods so far, we believe for medical images it has two drawbacks:

1. If the intensity of the pixels in a region of interest lay in the maximum and minimum range of the histogram, then their values are also modified.
2. If the minimum frequency of the histogram is non-zero, the coordinates of all the pixels with minimum frequency have to be embedded as side information. This restricts the data hiding capacity of the system.

Now if the image is partitioned into sub-images, the so-called image tiles, and the histogram shifting is applied to each image tile, not only the above shortfalls are overcome, but some additional benefits can be gained. These include:

1. Region of Interest: The image can be divided into parts such that, only the histograms of the non-region of interest image tiles are modified and the data is hidden.
2. High payload: In the shifted-histogram based data hiding method, the maximum number of hidden bits is equal to the maximum frequency of the pixel intensity histogram. When the histograms of the image tiles are considered separately, it is intuitive that the sum of individual maxima is greater than the maximum of the original image intensity histogram. Hence shifted-histograms of the image tiles can hide more data.
3. Higher objective quality: In the shifted-histogram method, the marked image quality depends on the number of pixels whose intensity lay between the maximum and minimum frequency pixels, irrespective of the number of hidden bits. That is, image quality due to embedding of one bit of data is as bad/good as if the maximum payload (equivalent to the maximum of histogram) is embedded. On the other hand, with the histograms of image tiles, they may be first prioritized, in the order of their least intensity distance between the maximum and minimum frequency. Data are embedded in the ordered image tiles till it is fully loaded, and the left over data will be carried over to the next image tile, and so on. In this way, for a given payload, the intensity of the smallest number of pixels is modified and hence image quality will be at its best.
4. Higher subjective quality: Rather than prioritizing the image tiles as in 3 above, they may be prioritized based on their spatial content. Data hiding can then start from those image tiles that have the highest spatial details. In this case, due to spatial masking of the human visual system, the subjective quality of the watermarked image will be at its best.
5. Narrower histogram: Some image tiles have much narrower histograms than that of the whole image. This is particularly true for medical images that leads to the flowing useful properties for data hiding:
   a) In the broader histogram of the whole image the minimum frequency may not be zero. Hence for reversible data hiding, their positions need to be identified and given as side information, which greatly reduce the data hiding capacity. On the other hand, in the narrower histograms of the image tiles, the minimum frequencies are more likely to be zero.
   b) Narrower histograms provide the opportunities of selecting the most suitable pairs of peaks-zeros that will increase the quality of the marked images.

The two steps of our embedding of watermark and its detection will be as follows:

2.1 Embedding

1. The image is first divided into N_b non-overlapping image tiles (e.g. N_b =4, 16). The intensity histogram of each image tile is generated and, the following steps (2-4) are iteratively executed for each image tile.
2. In each image tile, for a given number of n (peak, zero) pairs, the pairs are chosen such that the image quality is either maximized (least distances between the chosen pairs), or according to any other criteria such as perceptual quality. The (P_i, Z_i) pairs are then prioritized either based on objective or subjective quality, as explained above, with P_i and Z_i as the intensity of the peak and zero.
3. The following iteration is executed n times for i =1: n.
4. For pair (P_i, Z_i) the image tile is scanned and if:
   a) $P_i > Z_i$, the gray values of the pixels between Z_i+1 and P_i are reduced by one (shifting the range of the histogram [Z_i +1, P_i] by 1 to the left). This creates a gap at gray level P_i. The image tile is re-scanned and the gray values of the pixels with gray value of P_i -1 are incremented by one if the bits of the to be embedded data are "1", otherwise they will not be modified.
   b) $Z_i > P_i$, the gray values of the pixels between P_i +1 and Z_i -1 are increased by one. This creates a gap at gray value P_i +1. Then image tile is re-scanned and the gray values of the pixels with gray value of P_i are increased by one if the corresponding bits of to be embedded data are "1", otherwise they will not be altered.

The number of image tiles, N_b, their priority order, number of (peak, zero) pairs n, their positions will be treated as side information that needs to be transmitted to the receiving side for data retrieval.

2.2 Detection

For the given N_b, their embedding order and n, the following process is used to extract the secret message from a marked image and the lossless recovery of the host image.

1. Firstly, the image is divided into N_b image tiles. They are then rank ordered in their order of priority. Then steps 2-3 are repeatedly executed for each image tile.
2. The following iteration is done n times for i =1: n.
3. For pair (P_i, Z_i) the image tile is scanned and if:
   a) $P_i > Z_i$, the pixel with gray value P_i indicates that the embedded data bit was 1 and it should not be modified. Otherwise, if it is equal to P_i-1, it indicates that the embedded data bit was 0. In this case, its gray value has to be increased by 1. Later on the gray values of all pixels with gray values between Z_i and P_i-2 need to be increased by one.
   b) $Z_i > P_i$, the pixel with gray value P_i indicates that the embedded data bit was 0 and they do not need to be modified. However, if it is equal to P_i +1, it indicates the embedded data bit was 1. Then, its gray value is reduced by 1. Therefore, the gray values of all pixels with gray values between P_i +2 and Z_i are reduced by 1.

The shift of the peaks and zeros should not lead to loss of information about the location(s) of peaks and zeros. It is noteworthy that, in any case, if there is no sufficient number of zeros the minima are used instead of zeros.

3. EXPERIMENTAL RESULTS AND EVALUATIONS

We have tested the performance of the proposed method for 4 and 16 image tiles, on a variety of medical images. The original image sizes were 512 ×512 pixels with 8 bit resolution. Table 1 summarizes the results of Cancer tissue image(im2 of Fig 1) tiles, with respect to that of the whole image. As the table shows, tiled images have higher data hiding capability and still do have a better watermarked image quality. By using four tiles the overall capacity is 14,247 bits/whole image, which is almost 119% extra payload compared to using the whole image and average marked image quality of 45.3 dB, or 0.7 dB improvement in quality. The number of peak-zero pairs can vary from one tile to another and they can be arranged such that for instance, the watermarked image quality is uniform across the whole image. The extra payload is image dependent. For example, in Image number 6 (Im6), which shows the least average percentage of the watermarked image quality under tiling is always better than the whole image itself. On the other hand, in image 2 (Im2), albeit at a slightly lower quality, the degree of improvement in payload varies with marked image quality. The 4-tile and 16 tile images can hide on average more than 100% and 100-220% respectively over the whole improvement of 1-5 % for 4-tile and 30-40% improvement for the 16 tile images over the whole image.

Table. 1 Maximum capacity and marked quality of whole and 4 tiles of Cancer tissue image (im2 of Fig 1).

Image	Whole image	4 tiles image			
		Tile 1	Tile 2	Tile 3	Tile 4
Capacity	6492	4362	3794	4217	1874
		14,247			
PSNR	44.6	48	42.22	45.26	49.06
		45.3			

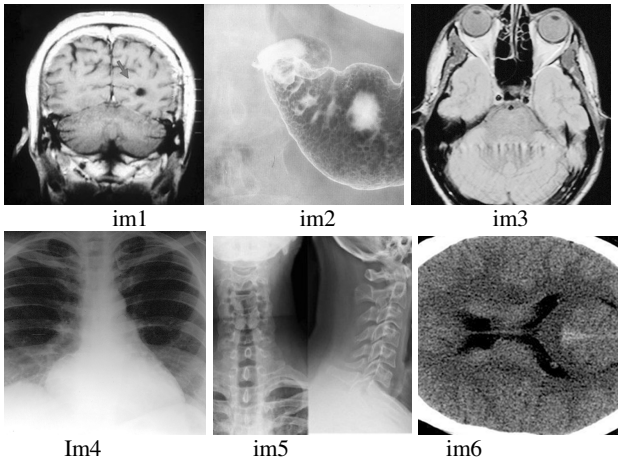


Fig. 1 Six original medical images

Table 2 shows the maximum payload and the quality of 6 various medical images which are shown in Fig 1, under shifted histogram of whole (WSH), 4-tile(TSH-4) and 16-tile (TSH-16) versions. In each experiment the results were the average of 60 embedded sets of random bit stream messages. In each experiment, data were embedded at the full capacity of each image, without use of any priority in image tiles, their spectral density or number of peak-zero pairs. Up to 4 pairs of peak-zero has been used but the PSNR may not be acceptable at some higher number of pairs. The first column shows the number of peak-zero pairs and the maximum payload of the whole image, 4-tile and 16-tile images are respectively depicted in columns, 2, 3 and 5. The percentage of increase in payload for 4 and 16 tile images over the whole image, for similar number of peak-zero pairs are respectively shown in columns 4 and 6. Finally, the watermarked quality of each method is shown in columns 7, 8 and 9. Marked image quality greater than 40 dB are highlighted. As in all cases quality can be traded for capacity. For equal quality, the improvement in payload capacity can be better judged from Fig 2, for two extreme images of Im2 and Im6. For instance for a 42 dB image quality, in Im6 while 4-tile image has 5% larger capacity over the whole image, in 16-tile version, this extra capacity is about 42%. This extra capacity, for the most favorable image, Im2 of the above data base is 110% and 200% for the 4tile and 16-tile.

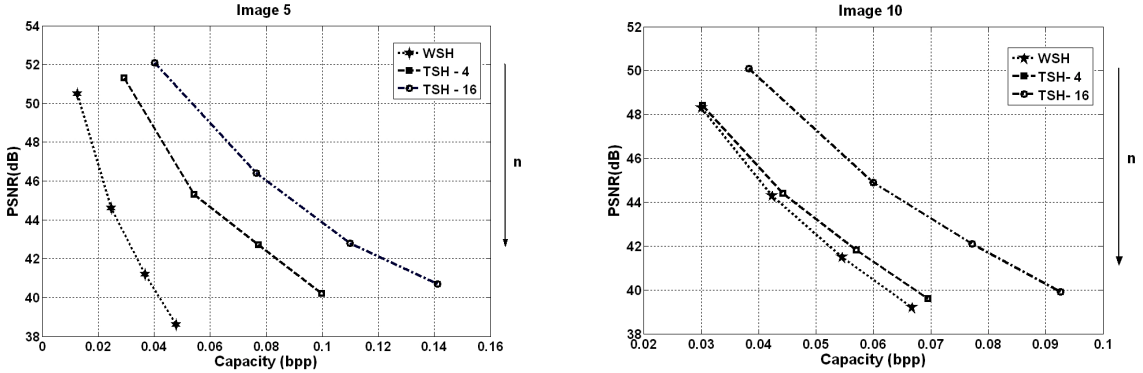


Fig. 2 PNSR versus capacity of the two extreme tiled-images

Table. 2 Maximum capacity and marked quality of whole, 4 and 16 tiles of 6 medical images of Fig 1.

Image	Pure payload						PSNR of marked images (dB)		
	WSH		TSH-4		TSH-16		WSH	TSH-4	TSH-16
	n	bits	bits	%	bits	%			
Im 1	1	9834	10931	11	16668	69	**49.0**	**50.4**	**53.0**
	2	19515	20874	7	31385	61	**43.0**	**43.9**	**47.3**
	3	26720	29810	12	44201	65	39.4	**40.6**	**43.9**
	4	33918	37967	12	55528	64	37.0	38.1	**41.2**
Im 2	1	3276	7694	135	10552	222	**50.5**	**51.3**	**52.1**
	2	6492	14247	119	20070	209	**44.6**	**45.3**	**46.4**
	3	9629	20243	110	28841	200	**41.2**	**42.7**	**42.8**
	4	12528	26153	109	36988	195	38.6	**40.2**	**40.7**
Im 3	1	11181	11506	3	15208	36	**49.0**	**50.2**	**51.3**
	2	21833	22465	3	29425	35	**42.9**	**44.1**	**45.2**
	3	30184	31702	5	41715	38	39.3	**40.6**	**41.5**
	4	38181	40333	6	53286	40	36.9	38.0	39.0
Im 4	1	2729	4817	77	6560	140	**48.5**	**54.2**	**53.1**
	2	5413	9560	77	12916	139	**42.5**	**49.1**	**47.3**
	3	8096	14155	75	19107	136	38.9	**44.5**	**43.5**
	4	10652	18698	76	25010	135	38.0	**41.4**	**41**
Im 5	1	2121	4382	107	10711	405	**57.3**	**56.1**	**54.3**
	2	3978	8433	112	16776	322	**52.3**	**49.5**	**48.2**
	3	5825	12116	108	22612	288	**48.8**	**46**	**44.8**
	4	7664	15555	103	28050	266	**44.1**	**43.3**	**41.5**
Im 6	1	7861	7929	1	10045	28	**48.3**	**48.4**	**50.1**
	2	11083	11607	5	15724	42	**44.3**	**44.4**	**44.9**
	3	14280	14944	5	20242	42	**41.5**	**41.8**	**42.1**
	4	17473	18217	4	24262	39	39.2	39.6	39.9

The image quality would have been improved, as the results of Fig 2 had indicated.

4. CONCLUSION

We have shown that data-hiding based on the shifted histogram is better to be applied to image tiles than the image itself. This not only improves the data-hiding capacity, but also improves the marked image quality. This is mainly due to the fact that sum of the peaks of the individual pixel intensity histograms is greater than the single peak of the image histogram itself. Besides, the individual histograms are much narrower and sharper than the histogram of the image itself, creating more possibility for zeros, as well as making distances between the peaks tiles, such that while the region-of-interest data can be free from disturbance, they can also be hidden according to the perceptual characteristics of the human visual system and zeros in each image tile shorter. All in all improving the marked image quality, while maintaining high data embedding capability. Finally individual histograms make it possible to distribute the embedded bits among the image.

5. ACKNOWLEDGMENTS

This work is partially supported by the Spanish Ministry of Science and Innovation and the FEDER funds under the grants TSI2007-65406-C03-03 E-AEGIS and CONSOLIDER-INGENIO 2010 CSD2007-00004 ARES.

REFERENCES

[1] Wakatani, A.: "Digital watermarking for ROI medical images by using compressed signature image". *Hawaii International Conference on System Sciences*, 2002, pp.2043-2048.

[2] Shih, F. Y., Wu, Y-.Ta : "Robust watermarking and compression for medical images based on genetic algorithms". *Journal of Information Sciences*, vol. 175, pp.200-216, 2005.

[3] Coatrieux, G., Sankur, B., and Maître, H.: "Strict Integrity Control of Biomedical Images". In Proc. Electronic Imaging, Security and Watermarking of Multimedia Contents, SPIE, USA, pp.229-240, 2001.

[4] Jagadish, N., Bhat , P. S., Acharya, R., and Niranjan, U. C.: 'Simultaneous storage of medical images in the spatial and frequency domain'. June 2004, a comparative study. Biomedical Engineering Online, 3(1):record 17.

[5] Macq, B., Dewey F.: "Trusted Headers for Medical Images". DFG VIII-DII Watermarking Workshop, Germany, 1999.

[6] Zhou, X. Q., Huang, H. K., and Lou, S. L.: "Authenticity and integrity of digital mammography images," IEEE Trans. on Medical Imaging, 2001, vol. 20, n°8, pp.784–791.

[7] Li, M., Poovendran, R., and Narayanan, S.: "Protecting patient privacy against unauthorized release of medical images in a group communication environment". 2005Computerized Medical Imaging and Graphics, vol. 29, n°5, pp. 367-383,

[8] Coatrieux, G., Lecornu, L., Sankur, B., and Roux Ch.: 'A Review of Image Watermarking Applications in Healthcare'. In Conference of the IEEE-EMBS, pp. 4691-4694, 2006.

[9] Ni, Z., Shi, Y. Q., Ansari, N., and Su, W.: 'Reversible data hiding'. IEEE Trans. on Circuits and Systems for Video technology, vol. 16, no. 3, pp.354-362, 2006.

2.4 High capacity method for real-time audio data hiding using the FFT transform

78. **M. Fallahpour,** David Megías, "High capacity method for real-time audio data hiding using the FFT transform" *Advances in Information Security and Its Application* Third International Conference, ISA 2009, Springer, Seoul, Korea, June 25-27, 2009. http://www.springerlink.com/content/k941412658442824/

High capacity method for real-time audio data hiding using the FFT transform

Mehdi Fallahpour and David Megías

Estudis d'Informàtica, Multimèdia i Telecomunicació
Universitat Oberta de Catalunya- Rambla del Poblenou, 156, Barcelona, Spain
Tel: (+34) 933 263 600, Fax: (+34) 933 568 822
E-mail: {mfallahpour,dmegias}@uoc.edu

Abstract. This paper presents a very efficient method for audio data hiding which is suitable for real-time applications. The FFT magnitudes which are in a band of frequencies between 5 and 15 kHz are modified slightly and the frequencies which have a magnitude less than a threshold are used for embedding. Its low complexity is one of the most important properties of this method making it appropriate for real-time applications. In addition, the suggested scheme is blind, since it does not need the original signal for extracting the hidden bits. The Experimental results show that it has a very good capacity (5 kbps), without significant perceptual distortion and provides robustness against MPEG compression (MP3).

Keywords: Audio data hiding, Fast Fourier Transform (FFT), Real time

1 Introduction

By the growth of audio production and broadcasting, security issues have arisen for audio producers. Recently, watermarking has been suggested as a good means for protecting audio against illegal tasks mainly for ownership proof and copy/modification detection. While most multimedia watermarking researches have been devoted to images so far, in the last decade many audio watermarking schemes have been suggested. These methods use the weakness of the human auditory system to embed the hidden data in the regions of the audio signals at which human ears are unable to perceive the distortion caused by the data embedding process.

Considering the embedding domain, audio watermarking techniques can be classified into time domain and frequency domain methods. In time domain schemes, the hidden bits are embedded directly into the time signal samples. These methods are easy to implement and are usually very efficient but they tend to be weak against common signal processing attacks. Phase modulation [1] and echo hiding [2] are well known methods in the time domain.

In frequency domain watermarking, after taking one of the usual transforms such as FFT, MDCT and WT from the signal, the hidden bits are embedded into the resulting transform coefficients [3-6]. In [6] the low-frequency coefficients of the wavelet transform are used for embedding the watermark. The robustness of this scheme

against various attacks is high but this method is not adaptive. Consequently, this scheme cannot be used in real-time software and users have to determine the embedding strength by using subjective audio quality tests for each audio signal. This process is very time-consuming and costly. In other group of schemes, the hidden data are embedded into the audio signal by changing middle frequency components in the frequency domain [7] or in the time domain [8].

Using methods based on transforms provides a better perception quality and robustness against common attacks at the price of increasing the computational complexity. For real-time applications complexity is the most important issue to be considered. In [9], an efficient audio watermarking method suitable for real-time applications is proposed. In this method, the hidden bits are embedded into the scale factor values during the MP3 encoding process. Also, [10] proposes a method to embed and extract the watermarks into and from digital compressed audio and therefore is applicable only to compressed audio. J. Garcia [11] proposed a real-time method which use spread spectrum algorithm in the modulated complex lapped transform (MCLT) domain to embed the watermark. In fact, just a few algorithms for an efficient real-time audio watermarking have been proposed so far.

In this paper we present a very efficient method for audio watermarking which is suitable for real-time applications. This scheme has been implemented taking special care for the efficient usage of the two restricted resources of computer systems: memory space and CPU time. It offers to the industrial user the capability of watermark embedding and detection in time immediately comparable to the real playing time of the original audio file, while the end user/audience does not find any artifacts or delays hearing the watermarked audio file. In the proposed algorithm, the FFT magnitudes of the selected clip which are in a band of frequency between 5 and 15 kHz are slightly distorted and magnitudes which have a value lower than a chosen threshold are used for embedding. This frequency band is scanned and when we meet the magnitude with the value lower the threshold it is increased if the corresponding embedding bit is ‘1’, otherwise the magnitude is not altered.

Low complexity is one of the most important properties of this method, making it appropriate for real-time applications. In addition, the suggested scheme is blind one since it does not need the original signal for extracting the hidden bits. The experimental results show that the method has a very high capacity (above 5 kbps) and provides robustness against MPEG compression.

The rest of the paper is organised as follows. In Section 2 the suggested scheme is presented. In Section 3, the experimental results are shown. Finally, Section 4 summarises the most relevant conclusions of this research.

2 Suggested scheme

We have chosen the Fast Fourier transform (FFT) domain to embed the hidden data to exploit the translation-invariant property of the FFT transform such that small distortions in the time domain can be resisted. Compared to other schemes, such as quantisation or odd/even modulation, keeping the relationship of FFT coefficient pairs is a more realistic scheme under distortions. We select a band of frequency between 5

kHz and 15 kHz to embed the data, choosing the frequencies for which the value of the magnitude is near one. Since we need integer values for the FFT magnitude in the embedding step, these magnitudes are multiplied by a reference level q and then rounded to the nearest integer. This scaling and rounding process generates a great deal of zero values in the magnitudes, so the signal will result slightly distorted. Changing the parameter q, both the capacity and the perceptual distortion will be affected. When q decreases, the capacity is increased but the perceptual distortion also enlarges and vice versa (capacity and distortion can decrease by rising q).

Figure 1 shows an example of selecting a band to embed data. The experimental results presented in Section 3 show how this choice affects the properties of the watermarking scheme (capacity and transparency). The choice of the band of frequencies and the reference level q depends on the application. For example, if the marked audio should have a very high capacity a wide frequency band could be selected but this would increase distortion.

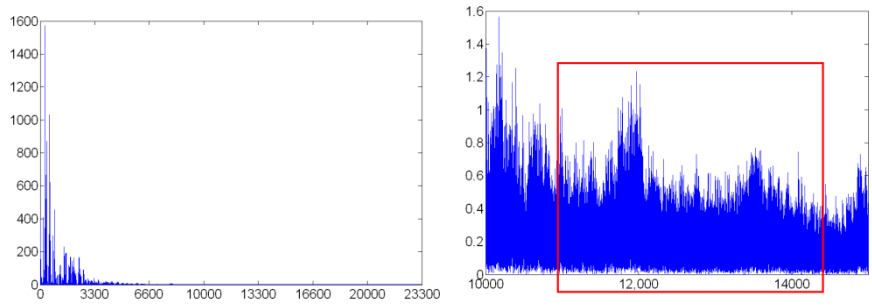


Fig 1. (a) Magnitude of the spectrum for the violoncello wave file and (b) selected band for embedding

The watermarking scheme presented here is positional. This means that the detector must be synchronised in order to recover the embedded bits correctly. In a real application, the cover signal would be divided into several blocks of a few seconds and it is essential that the detector can determine the position (the beginning sample) of each of these blocks. One of the most practical solutions to solve this problem is to use synchronisation marks such that the detector can determine the beginning of each block. Several synchronisations strategies have been described in the literature (for example [15, 16]) and any of them could be used together with the method described here in order to produce a practical self-synchronising solution.

The embedding and detection methods are described in the following sections.

2.1 Watermark embedding

The embedding steps are as follows:

1- Based on the computation processor (speed and memory) select the length of the segment of the audio file.

2- Calculate the FFT of the audio segment.

3- Select the band of frequencies between 5 kHz and 15 kHz for which the magnitudes are near 1.
4- Using q as a parameter, convert the FFT magnitudes to integer values (multiplying them by q and then rounding).
5- Expanding step: scan all these integer FFT magnitudes in the selected band. If a magnitude is larger than zero then increase it by 1. After this step we have no magnitude with the value 1.
6- Embedding step: scan again all integer FFT magnitudes in selected band. When a zero magnitude is found, if the corresponding embedded bit is '1' add one to the magnitude. Otherwise, the magnitude is not changed. After this step all magnitudes with value zero or one represent an embedded bit.
7- The marked (FFT) signal is achieved by dividing all the magnitudes by q.
8- Finally, use IFFT to achieve the marked audio segment in the time domain.

2.2 Watermark extracting

The watermark detection is performed by using the FFT transform and the embedding parameters. Since the host audio signal is not required in the detection process, the detector is blind. The detection process can be summarised in the following steps:
1- Calculate the FFT of the marked audio segment.
2- To achieve scaled FFT magnitudes multiply them by q.
3- Detection step: scan all scaled FFT magnitudes in selected band. If a magnitude with value in the interval [0, 1/2) is found, then the corresponding embedded bit is equal to '0' and the restored magnitude equals to zero. If the magnitude value is in the interval [1/2, 3/2), then the corresponding embedded bit is equal to '1' and the restored magnitude equals to zero.
4- Scan all scaled FFT magnitudes in the selected band. If each magnitude value in interval [k+1/2, k+3/2), then the restored magnitude equals to k.
5- The restored magnitudes are achieved by dividing them by q.
6- Finally, use IFFT to achieve the restored audio segment.

3 Experimental results

To evaluate the performance of the proposed method, two pieces of audio signals have been selected from the Sound Quality Assessment Material (SQAM) corpus audio files [12], used for watermarking. All audio clips were sampled at 44.1 kHz with 16 bits per sample resolution. The experiments have been performed for each channel of the audio signals separately. Hence, we have four test mono signals.
The Objective Difference Grade (ODG) is used to evaluate the transparency of the proposed algorithm. The ODG is one of the output values of the ITU-R BS.1387 PEAQ [13] standard, and its description is shown in Table 1. Additionally, the software OPERA [14] based on ITU-R BS. 1387 was used to provide an objective measure of quality.

Table 1. ODG description

Impairment	ODG
Imperceptible	0.0
perceptible, not annoying	−1.0
slightly annoying	−2.0
Annoying	−3.0
very annoying	−4.0

Tables 2, 3, 4, and 5 show the effect of changing the scaling coefficient q and the width of the selected frequency band (B) on capacity, transparency (ODG) and the BER of detection after MPEG compression for the Trumpet (trpt21_2.wav) and Violoncello (vioo10_2.wav) wave files.

In Table 2, the first and second rows show the results obtained with the same q and different B. When using a wider frequency band both the capacity and the perceptual distortion increase. Note that BER is provided for various MPEG compression rates. The fourth and fifth rows show that when q increases, capacity decreases but perceptual distortion is reduced. Finally the sixth and seventh rows show that the BER is affected by the MP3 compression rate.

Table 2. Trumpet.wav left channel

q	Frequency band (kHz)	Capacity (bps)	ODG	Mp3 rate (kbps)	BER %
5	8.5-16	7205	−0.73	128(> 128)	1.0 (0.0)
5	11-16	5038	−0.67	128(> 128)	1.2 (0.0)
4	6.5-16	8454	−0.95	128(> 128)	1.0 (0.0)
4	5-11	6009	−0.92	≥ 128	0.0
3	5-11	6050	−1.71	≥ 128	0.0
3	3-14	7449	−1.94	128	1.7
3	3-14	7449	−1.94	> 128	0.0

Table 3. Trumpet.wav right channel

q	Frequency band (kHz)	Capacity (bps)	ODG	Mp3 rate (kbps)	BER %
5	8.5-16	5860	−0.67	128	1.0
5	11-16	2534	−0.71	128	0.0
4	6.5-16	6083	−0.93	128	1.0
4	5-11	3954	−1.06	≥ 128	0.0
3	5-11	4418	−1.51	≥ 128	0.0
3	3-14	7560	−1.75	128	1.1

Table 4. Violoncello left channel

q	Frequency band (kHz)	Capacity (bps)	ODG	Mp3 rate (kbps)	BER %
1.5	11-14	4074	–1.06	≥ 128	0.0
1.75	11-14	3570	–0.91	128 (> 128)	0.5 (0.0)

Table 5. Violoncello right channel

q	Frequency band (kHz)	Capacity (bps)	ODG	Mp3 rate (kbps)	BER %
1.5	11-14	3155	–0.93	128 (>128)	3.0 (0.0)
1.25	11-14	3788	–1.17	≥ 128	0.0

It is worth to mentioning that Reed-Solomon Codes (or other error correction codes) could be used to improve BER rates if required (at the price of reducing the capacity). Similar results are shown in the other tables.One of important issue in audio watermarking is computation time. As FFT is a fast transform this method is very useful for real-time applications. Table 6 illustrates the embedding and extracting times. It worth to mention that these computation times have been achieved with an Intel (R) core (TM) 2 Duo 2.2GHz CPU and 2 GB of RAM memory. It can be noticed that the extracting time is one order of magnitude smaller than the file playing time. Thus, it is perfectly possible to recover the embedded data in a real-time scenario.

Table 6. Computation time

Audio File	Time (sec)	Embedding time (sec)	Extracting time (sec)
Violoncello	30	4.15	4.01
Trumpet	17.8	1.82	1.76

5 Conclusion

In this paper, we describe a high capacity data hiding algorithm for digital audio which is appropriate for real-time applications. A scaling coefficient (q) and the selected frequency band to embed the hidden information in it are the two main parameters of this method. The experimental results show that using different values for q and the frequency band lead to different capacity, perceptual distortion (ODG) and bit error rates (BER) of detection after MP3 compression. Low complexity is one of the most important properties of this method, making it suitable for real time applications. Furthermore, the suggested scheme is blind, since it does not need the original signal for extracting the hidden bits. The experimental results show that this scheme has a very good capacity (5 kbps or above), without significant perceptual distortion and provides robustness against MPEG compression (MP3).

Acknowledgement

This work is partially supported by the Spanish Ministry of Science and Innovation and the FEDER funds under the grants TSI2007-65406-C03-03 E-AEGIS and CONSOLIDER CSD2007-00004 ARES.

References

1. W. N. Lie, L. C. Chang, "Multiple Watermarks for Stereo Audio Signals Using Phase-Modulation Techniques", *IEEE Trans. Signal Processing*, Vol. 53, No. 2, pp. 806–815, Feb. 2005.
2. H. J. Kim, Y. H. Choi, " A novel echo hiding scheme with backward and forward kernels", *IEEE Trans. Circuit and Systems*, Aug. 2003, pp. 885-889.
3. S. Esmaili, S. Krishnan, K. Raahemifar, "A novel spread spectrum audio watermarking scheme based on time - frequency characteristics". *IEEE Conf. Electrical and Computer Engineering*, Vol. 3, pp. 1963-1966, May 2003.
4. D. Megías, J. Herrera-Joancomartí, and J. Minguillón. "A robust audio watermarking scheme based on MPEG 1 layer 3 compression", *Communications and Multimedia Security - CMS 2003*, Lecture Notes in Computer Science 2828, pages 226–238, Turin (Italy), October 2003. Springer-Verlag.
5. N. Sriyingyong, K. Attakitmongcol, "Wavelet-Based Audio Watermarking Using Adaptive Tabu Search", *IEEE Int. symp. Wireless Pervasive Computing,* pp. 1-5, Jan. 2006.
6. S. Wu, J. Huang, D. Huang, Y. Q. Shi, "Efficiently Self-Synchronized Audio Watermarking for Assured Audio Data Transmission", *IEEE Trans. Broadcasting*, Vol. 51, No. 1, pp. 69-76, Mar. 2005.
7. W. Li, X. Xue, P. Lu, "Localized Audio Watermarking Technique Robust Against Time-Scale Modification", *IEEE Trans. Multimedia*, Vol. 8, No. 1, pp. 60-69, Feb. 2006.
8. A. N. Lemma, J. Aprea, W. Oomen, L. v. d. Kerkhof, "A Temporal Domain Audio Watermarking Technique", *IEEE Trans. signal processing*, Vol. 51, No. 4, pp. 1088- 1097, Apr. 2003.
9. D. Koukopoulos, Y. Stamatiou, "An Efficient WatermarkingMethod for MP3 Audio Files", *Transaction on Engineering, Computer and Technology* V7, August 2005.
10. C.Xu,Y.Zhu,D.D.Feng, "A robust and fast watermarking scheme for compressed audio", *IEEE International Conference on Multimedia and Expo* ISBN 0-7695-1198-8/01 , 2001
11. J. Garcia, M. Nakano, and H. Perez. "Real time implementation of low complexity audio watermarking algorithm", *Proc. Third International Workshop on Random Fields and Processing in Inhomogeneous Media*, October 2005.
12. SQAM Sound Quality Assessment Material, http://sound.media.mit.edu/mpeg4/audio/sqam/
13. T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerens, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, and B. Feiten, "PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality," *Journal of the AES*, vol. 48(1/2), pp. 3–29, 2000.
14. OPTICOM OPERA software site. http://www.opticom.de/products/configurations.html
15. X.-Y. Wang and H. Zhao. A novel synchronization invariant audio watermarking scheme based on DWT and DCT. IEEE Transactions on Signal Processing, 54(12):4835–4840, December 2006.
16. Y. Lin and W. Abdulla. A secure and robust audio watermarking scheme using multiple scrambling and adaptive synchronization. In Proceedings of the 6th International Conference on Information, Communications & Signal Processing, pages 1–5, 2007.

2.5 High capacity audio watermarking using FFT amplitude interpolation

79 M. Fallahpour and David Megías, “High capacity audio watermarking using FFT amplitude interpolation”, *IEICE Electron. Express*, Vol. 6, No. 14, pp.1057-1063, (2009). (Impact factor 0.48) http://www.jstage.jst.go.jp/article/elex/6/14/6_1057/_article

High capacity audio watermarking using FFT amplitude interpolation

Mehdi Fallahpour,[1a)] and David Megias[1b)]
[1]*Universitat Oberta de Catalunya, Rambla del Poblenou, 156, Barcelona, Spain*
a) MFallahpour@UOC.edu b) DMegias@UOC.edu

Abstract: An audio watermarking technique in the frequency domain which takes advantage of interpolation is proposed. Interpolated FFT samples are used to generate imperceptible marks. The experimental results show that the suggested method has very high capacity (about 3 kbps), without significant perceptual distortion (ODG about –0.5) and provides robustness against common audio signal processing such as echo, add noise, filtering, resampling and MPEG compression (MP3). Depending on the specific application, the tuning parameters could be selected adaptively to achieve even more capacity and better transparency.

Keywords: Watermarking, FFT, Spline interpolation, Frequency domain
Classification: Science and engineering for electronics

References

1. J. J. Garcia-Hernandez, M. Nakano-Miyatake and H. Perez-Meana, "Data hiding in audio signal using Rational Dither Modulation", *IEICE Electron. Express*, Vol. 5, No. 7, pp.217-222, 2008.
2. H. Kang, K. Yamaguchi, B. Kurkoski, K. Yamaguchi, and K. Kobayashi, "Full-Index-Embedding Patchwork Algorithm for Audio Watermarking", *IEICE TRANS. on Information and Systems*, E91-D(11):2731-2734, 2008
3. D. Megías, J. Herrera, J. Minguillón, "Total Disclosure of the Embedding and Detection Algorithms in a Robust Digital Watermarking Scheme for Audio". *LNCS*. Volume 3783, pages 427-440, 2005
4. S. Xiang, H.J. Kim, J. Huang, "Audio watermarking robust against time-scale modification and MP3 compression" Signal Processing,v.88, 2372-2387, 2008
5. M. Fan and H. Wang, "Chaos-based discrete fractional Sine transform domain audio watermarking scheme" *Computers and Electrical Engineering*, *Elsevier* Volume 35 , Issue 3, Pages 506-516, May 2009.
6. R. Fujimoto, M. Iwaki, T. Kiryu, "A Method of High Bit-Rate Data Hiding in Music Using Spline Interpolation". *IIH-MSP* , 11-14, 2006
7. A. Deshpande, K. M. M. Prabhu, "A substitution-by-interpolation algorithm for watermarking audio". *Signal Processing, Elsevier* 89(2): 218-225, 2009
8. C J Weinstein, "Programs for Digital Signal Processing", IEEE Press, 1979.
9. T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerens, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, and B. Feiten, "PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality,"

[DOI: **.****/**********]
Received September **, 2008;
Accepted September **, 2008;
Published in October **, 2008 issue

Journal of the AES, vol. 48(1/2), pp. 3–29, 2000.
10. http://www.jamendo.com/en/album/7365
11. http://www.opticom.de/products/opera.html
12. http://wwwiti.cs.uni-magdeburg.de/~alang/smba.php

1. Introduction

With the broad use of information security applications and the rising growth of the watermarking schemes, various signal processing techniques are being used to improve audio watermarking methods. Audio watermarking methods exploit the insensitivity of the human auditory system (HAS) in various techniques such as embedding algorithms based on low-bit coding, echo, rational dither modulation [1], patchwork [2], Fourier transform [3, 5], wavelet transform [4] or spread spectrum and interpolation [6, 7].

Interpolation techniques are often designed to provide a good perceptual quality from known sample values [8]. In [6], an original audio signal is divided into distinct frames and then a secret bit is embedded in each frame by using spline interpolation. [7] proposes a spline interpolation-based watermarking scheme with more robustness against attacks than the one suggested in [6].

The aim of the proposed method is to develop a high-bit-rate audio watermarking technique with robustness against common attacks and good transparency. This algorithm is based on the difference between the original and the interpolated amplitudes of the FFT samples as obtained by spline interpolation. If the difference is lower than a given fraction of the interpolated value, it is selected for embedding secret information. To obtain the marked FFT samples, the interpolated value is changed according to the secret bit. The experimental results show that the method provides high data bit rate, about 3 kbps, with good perceptual transparency (ODG about –0.5) and robustness against common attacks. Better capacity and transparency may be achieved with appropriate tuning for specific applications.

2. Proposed method

Interpolation [8] is a technique of constructing new data points within the range of a discrete set of known data points. Linear interpolation is obtained by passing a straight line between two data points. Polynomial interpolation is the best known one-dimensional interpolation scheme. Its advantages consist of its simplicity of implementation and the good quality of the interpolants obtained from it. However, it has a relatively low performance. In the spline interpolation, the interpolation interval is divided into small subintervals and each of these is interpolated by using a third-degree polynomial. The main advantages of the spline interpolation, which is used in the proposed algorithm, are its stability and calculation simplicity.

2.2 Embedding algorithm

We use some of the original FFT samples as source data in the coder and do not alter them. Hence, these original values can be used in decoder. In the algorithm,

[DOI: **.****/**********]
Received September **, 2008;
Accepted September **, 2008;
Published in October **, 2008 issue

we use the odd FFT samples to generate the interpolated values of the even samples which are used for embedding the secret bits. In the receiver, the original values of the odd FFT samples and the interpolated even samples are the same as in the coder. The embedding steps are follows:

$k = 1$;
for $i = low_{band}$ **to** $high_{band}$
    if $\mathrm{mod}(i, 2) == 0$
        $e_i = f_i - I_i$;
        if $|e_i| > 2\alpha\, I_i$
            $f_i' = f_i$;
        else if $b_k == 0$
            $f_i' = I_i; k = k + 1$;
        else if $\{(b_k == 1)$ **and** $(e_i \geq 0)\}$
            $f_i' = I_i(1 + \alpha); k = k + 1$;
        else
            $f_i' = I_i(1 - \alpha); k = k + 1$;
        end;
    end;
end;

where f_i is the magnitude of the i^{th} sample of the FFT spectrum, low_{band} *and* $high_{band}$ are the lower and higher limits of a selected band for embedding secret information , I_i is the interpolated value of f_i, α is a threshold, $b_k$is the k^{th} bit of secret bit stream, and f_i' is the marked value of f_i.

Adaptation to capacity, transparency and robustness against attacks is the most relevant advantage of the proposed method. *i.e.* the results are altered by changing the selected frequency band and the threshold. The selected frequency band is the area which is used for embedding secret information. Another main parameter of this algorithm is threshold α, which defines the maximum allowed ratio of interpolation error (difference between original FFT sample and its interpolant) to the interpolated sample for embedding. The interpolated sample can be used for embedding secret information when the ratio between the interpolated error to the interpolated sample is less than the threshold. The frequency band 0.5 – 5 kHz and α = 0.3 are suitable fixed parameters which have been selected for this algorithm after different experiments. However, depending on the application, these parameters could be modified to obtain better capacity and/or distortion.

The FFT samples with an amplitude value lower than one (when the original time-domain audio signal is normalised between –1 and 1) are sensitive to a few attacks such as MP3 compression. Thus, interpolation is not used for even FFT bins with value less than one. These FFT amplitudes are fixed to 0 instead and, if the corresponding embedding bit is 0, it is not changed. Otherwise, if the secret bit is one, the amplitude is increased to 0.5.

The effect of some attacks [12] on the embedding scheme has been analysed. For example, the echo attack is one of the most difficult ones to survive. Fig. 1 (a) shows the FFT spectrum of the marked signal and Fig.1 (b) displays the echo-attacked spectrum. Fig.1 (c) illustrates the difference between marked and echo-attacked spectra. This spectrum shows how the echo-10 attack repeats the signal after 10 samples and destroys the signal. In this case, we can use three secure spaces which are indicated by red rectangles for which the difference is

[DOI: **.****/**********]
Received September **, 2008;
Accepted September **, 2008;
Published in October **, 2008 issue

very low and the echo attack does not destroy the FFT samples. Fig 1. (d) – (f) shows the marked, attacked, error of attack and the secure embedding area (the region for which the difference between the marked and attacked signals is low) for the BassBoost attack. To find a secure area, we should consider the ratio between the error and the attacked samples, not the amplitude of the error, since we use the ratio to extract information. Different applications have different requirements. In copyright protection, robustness is more relevant than capacity, whereas for transmission of secret information it is the opposite. Thus, depending on the application, the frequency band and the threshold could be changed. In this paper, we have used a fixed frequency band and threshold for various natural audio files and different attacks to show the widespread use of the proposed method.

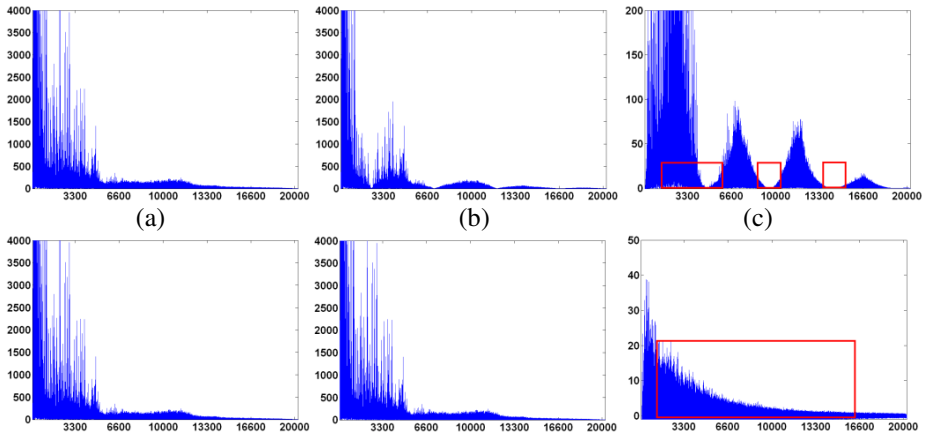


Fig. 1. FFT spectrum of (a),(d) marked audio signal (b) echo-10 attacked (c) error of echo-10 attacked (e) BassBoost attacked (f) error of BassBoost attacked

2.3 Extraction algorithm

As mentioned for the embedding steps, in the receiver the original values of odd FFT samples and the interpolated value of the even samples, I_i, are the same as in the coder. The following algorithm describes how to obtain the secret bit sequence, $b_k^{'}$, by using the marked FFT sample , $f_i^{'}$, the frequency bands and the threshold , α, as input values.

$k = 1;$
for $i = low_{band}$ **to** $high_{band}$
  if $\mathrm{mod}(i, 2) == 0$
    $e_i' = f_i' - I_i$;
    if $|e_i'| < 0.5\alpha\, I_i$
      $b_k' = 0;\ \ k = k + 1;$
    else if $(|e_i'| \geq 0.5\alpha\, I_i)$ **and** $(|e_i'| \leq 1.5\alpha\, I_i)$
      $b_k' = 1;\ \ k = k + 1;$
    end;
  end;
end;

3. Experimental results

To evaluate the performance of the proposed method and to consider the applicability of the scheme in a real scenario, five songs included in the album Rust by No, Really [10] have been selected. All audio clips are sampled at 44.1 kHz with 16 bits per sample and two channels. The Objective Difference Grade

[DOI: **.****/**********]
Received September **, 2008;
Accepted September **, 2008;
Published in October **, 2008 issue

(ODG) [9] is used to evaluate the transparency where ODG = 0 means no degradation and ODG = –4 means a very annoying distortion.

Table I illustrates the perceptual distortion and the payload obtained for these songs and compares the performance of the proposed watermarking algorithm and several recent audio watermarking strategies robust against many attacks. [1] Evaluates distortion by mean opinion score (MOS), which is a subjective measurment, and achieves transparency between imperceptible and perceptible but not annoying, MOS = 4.7. [4, 5] have a low capacity but are robust against most of common attacks. [6] proposes a high bit rate data hiding, but only considers MP3 compression attacks. We have used several random bits for embedding leading to different transparency results which are shown in the ODG column. The comparison shows the superiority in both capacity and imperceptibility of this method with respect to other schemes in the literature. Note that all the results have an ODG between 0 (not perceptible) –1 (not annoying), and that capacity is around 3000 bps in all the experiments. The proposed method is thus able to provide large capacity whilst keeping imperceptibility in the admitted range (–1 to 0). We provide imperceptibility results both as SNR and ODG. SNR is provided only for comparison with other works, but ODG is a more accurate measurement of audio distortions, since it is assumed to provide a good model of the subjective difference grade (SDG) results which may be obtained by a group of human listeners. The SNR results are computed using the whole (original and marked) files, whereas the ODG results are provided using the advanced ITU-R BS.1387 standard as implemented in the Opera software [11] (the average of measurements taken in frames of 1024 samples).

Table II shows the effect of various attacks provided in the Stirmark Benchmark for Audio v1.0 [12] on ODG and BER for the five audio signals of Table I. The embedding method has been applied, the SMBA software has been used to attack the marked files and, finally, the detection method has been performed for the attacked files. The ODG in Table II is calculated between the marked and the attacked-marked files. The parameters of the attacks are defined based on SMBA web site [12]. For example, in AddBrumm, 1-7000 shows the strength and 0-1500 shows the frequency. This row illustrates that any value in the range 1-7000 for the strength and 1-1500 for the frequency could be used without any change in BER. In fact, this table shows the worst and best results for the five test signals based on

Table I. Results of 5 signals (robust against table II attacks) and comparison between literature schemes

Algorithm	*Audio File*	*Time (m:sec)*	*SNR (dB)*	*ODG of marked*	*Payload (bps)*
proposed	Beginning of the End	3:16	26.5 to 33	–0.6 to –0.9	3142
	Breathing On Another Planet	3:13	26 to 35.5	–0.3 to –0.9	3047
	Citizen, Go Back to Sleep	1:57	24.5 to 30	–0.1 to –0.7	2825
	Go	1:51	29 to 37.5	–0.3 to –0.8	2938
	Thousand Yard Stare	3:57	27 to 36.5	–0.3 to –0.9	3030
	average	**2:50**	**30.55**	**– 0.58**	**2996**
[1]	Song	-	Not reported	Not reported	689
[3]	One instrument	~0:20	18 to 40	–0.5 to –2	61
[4]	One instrument	~0:20	42 to 45	–1.66 to –1.88	2
[5]	Song	-	30 - 45	Not reported	86
[6]	Classical music	0:10	25	Not reported	~ 1k

[DOI: **.****/**********]
Received September **, 2008;
Accepted September **, 2008;
Published in October **, 2008 issue

Table II. Robustness test results for five selected files

Attack name	*State*	ODG of attacked file	*BER %*	*parameters*
AddBrumm	best	–3.5	0	1-7000, 1-1500
	worst	–3.6	0.5	1-9000,1-3000
AddDynNoise	best	–3.25	0	1-5
	worst	–0.27	8	1
ADDFFTNoise	best	–0.59	0.5	2048, 2000
	worst	–0.68	1.5	2 , 50
Addnoise	best	–0.28	0	1-80
	worst	–0.64	0.5	1-40
AddSinus	best	–0.2	0	No Limitation
	worst	–0.1	0	
Amplify	best	–0.2	0	10-100
	worst	–0.1	0	10-100
BassBoost	best	–3.6	0	1-50 and 1-45
	worst	–3.5	0	1-45 and 1-20
Echo	best	–3.43	0	1-4
	worst	–2.93	0.5	1-10
FFT_HLPassQuick	best	–3.2	1	1024,1,7k-15k
	worst	–3.51	3.2	1024,1,9k-15k
FFT_Invert	best	–3.54	1.5	1024
	worst	–3.59	2	1024
invert	best	–3.5	0	No Limitation
	worst	–2.8	0	
Resampling	–	–1.4	4.5	44.1 to 22.05 to 44.1(kHz)
LSBZero	Best	– 0.2	0	–
	worst	– 0.1	0	
MP3	Best	– 0.2	0 (0.5,3)	>160 (160,128)
	worst	– 0.1	0.5 (1.5,4.5)	>160 (160,128)
Noise_Max	best	–3.5	0	1-10,1-500,1-300
	worst	–1.8	0.5	1-14,1-100,1-100
Pitchscale	best	– 0.1	0	1
	Worst	– 0.3	1	1
RC_HighPass	best	– 2.9	0	21k>
	worst	– 2.7	0.5	21k>
RC_LowPass	best	–3.4	0	700<
	worst	–2.9	0	1000<

BER and, in the case with the same BER, based on limitation of parameters. When the BER is (slightly) greater than zero, it can be made zero by using Error Correction Codes at the price of reducing the capacity. As mentioned above, depending on the specific application the parameters can be changed. E.g. by using frequency band 4–16 kHz and α =0.25 for the clip "Citizen, Go Back to Sleep", the obtained capacity is 6460 bps and ODG= –0.7, but robustness is decreased.

A very relevant issue in audio watermarking is computation time. As FFT is a fast transform, this method is useful for real-time applications. The extracting time is about 20% of the file playing time. Thus, it is perfectly possible to recover the embedded data in a real-time scenario. It is worth mentioning that these computation times have been obtained with an Intel core (TM) 2 Duo 2.2 GHz CPU and 2 GB of RAM memory with MATLAB.

4. Conclusion

In this paper, we describe a robust high-capacity watermarking algorithm for digital audio which is robust against common signal processing attacks. The ratio between the interpolated error to the interpolated sample and the selected frequency band are the two parameters of this method which can be selected

[DOI: **.****/**********]
Received September **, 2008;
Accepted September **, 2008;
Published in October **, 2008 issue

adaptively to regulate the capacity, the perceptual distortion and the robustness of the scheme. Furthermore, the suggested scheme is blind. The experimental results show that this scheme has a high capacity (about 3 kbps) without significant perceptual distortion (ODG about –0.5) and provides robustness against common signal processing attacks such as echo, noise, filtering, resampling, and MPEG compression (MP3). Besides, the CPU time required by the scheme is short enough (about 20% of the playing time) to use it in real-time applications.

[DOI: **.****/**********]
Received September **, 2008;
Accepted September **, 2008;
Published in October **, 2008 issue

2.6 Robust high-capacity audio watermarking based on FFT amplitude modification

80. **M. Fallahpour,** David Megías, "Robust high-capacity audio watermarking based on FFT amplitude modification" *IEICE Transactions on Information and Systems*, Vol.E93-D, No.01, pp.-, Jan. 2010. (Impact factor=0.3), in press.

IEICE Transactions on Information and Systems., VOL. E85-A/B/C/D, No. xx JANUARY 20xx

PAPER

Robust high-capacity audio watermarking based on FFT amplitude modification

M. Fallahpour,† *student member* and D. Megias†

Summary This paper proposes a novel robust audio watermarking algorithm to embed data and extract it in a bit-exact manner based on changing the magnitudes of the FFT spectrum. The key point is selecting a frequency band for embedding based on the comparison between the original and the MP3 compressed/decompressed signal and on a suitable scaling factor. The experimental results show that the method has a very high capacity (about 5 kbps), without significant perceptual distortion (ODG about –0.25) and provides robustness against common audio signal processing such as added noise, filtering and MPEG compression (MP3). Furthermore, the proposed method has a larger capacity (number of embedded bits to number of host bits rate) than recent image data hiding methods.

Key words: *Audio watermarking, Fast Fourier Transform (FFT).*

1. Introduction

The easy transmission and manipulation of digital media has led to a strong demand for watermarking schemes. Since the human auditory system is more sensitive than the visual system, to develop a high-performance audio watermarking technique is a challenging task. Considering the embedding domain, audio watermarking techniques can be classified into time domain and frequency domain methods. Phase modulation [1] and echo hiding [2] are well known methods in the time domain.

In frequency domain watermarking [3-8], after taking one of the usual transforms such as the Discrete/Fast Fourier Transform (DFT/FFT), the Modified Discrete Cosine Transform (MDCT) or the Wavelet Transform (WT) from the signal, the hidden bits are embedded into the resulting transform coefficients. In [6, 8] the FFT domain is selected to embed watermarks for making use of the translation-invariant property of the FFT coefficients to resist small distortions in the time domain. In particular, [8] shows that the FFT domain provides excellent robustness against MP3 compression. In fact, using methods based on transforms provides better perceptual quality and robustness against common attacks at the price of increasing the computational complexity.

In the algorithm suggested in this paper, selecting the frequency band and a scaling factor are the tuning steps. We consider that a safe area for embedding information is the frequency range at which the difference between the FFT magnitudes of the original and the MP3 compressed/decompressed signals is lower than a threshold. Moreover, to strengthen the robustness against attacks, a scaling factor is used. This factor adjusts the value which is added to the FFT magnitudes in the embedding step. For embedding, the FFT magnitudes are first scaled and rounded to the nearest integer. Then, the selected frequency band is scanned and when we meet a magnitude with the value larger than one, it is incremented. If the magnitude is equal to zero it is incremented if the corresponding embedding bit is '1', otherwise the magnitude is not altered. The experimental results show that this method has a very high capacity (about 5 kbps), provides robustness against common signal processing attacks, and entails very low perceptual distortion. Using FFT magnitudes, $\sqrt{\text{real}^2 + \text{imag}^2}$, results in better robustness against attacks compared to using the real or the imaginary parts.

The rest of the paper is organized as follows. In Section 2, the proposed method is presented. In Section 3, the experimental results are shown. Finally, Section 4 summarizes the most relevant conclusions of this research.

2. Proposed method

In this scheme, we use the following method to embed a bit stream (secret bits) into a set of various numbers (FFT coefficients). In the set of numbers, the number which is most frequently encountered than others in the set is selected to embed the hidden bits. For example, in A = {0 2 3 4 7 1 0 2 0 2 1 0}, the zero value is selected. Then, all numbers larger than the selected value (in this case zero) are incremented (shifted) A' = {0 **3 4 5 8 2** 0 **3** 0 **3 2** 0}. Note that, in the shifted A', there is no number with value equal to 1. Finally, in the embedding step, the stream is scanned and the secret bits ("0110") are embedded. When we meet "0" in the stream, if corresponding secret bit is "0", it will not be changed, but if it is equal to "1" it should be incremented. The marked set of values is then A^* = {**0** 3 4 5 8 2 **1** 3 **1** 3 2 **0**}. At the detector side, the secret bits are extracted and, then, all values larger than the selected value are decremented

As mentioned above, we have chosen the FFT domain to embed the hidden data in order to exploit the translation-invariant property of the FFT transform such that small distortions in the time domain can be resisted. Compared to other schemes, such as quantization or

Manuscript received January xx, 20xx.
Manuscript revised March xx, 20xx.

† The authors are with Estudis d'Informàtica, Multimèdia i Telecomunicació, Universitat Oberta de Catalunya, Rambla del Poblenou, 156, 08018 Barcelona, Spain.

odd/even modulation, keeping the relationship of FFT coefficient pairs is a more realistic scheme under several distortions.

2.1 Tuning

This method is based on using near zero values of FFT magnitudes in a selected frequency band. The method needs integer values for the FFT magnitude in the embedding step. Thus, the magnitudes in the selected band are multiplied by a scaling factor *s* and then rounded to the nearest integer. This scaling and rounding process generates a great deal of zero values in the integer magnitudes. The frequency band and the scaling value (*s*) are the two parameters of this method which adjust capacity, perceptual distortion and robustness.

To select the frequency band, the considered points are as described below:

1. The selected band should have as many zeros as possible after scaling (multiplying by *s*) and rounding. The number of zeros identifies the capacity.
2. In the selected band, the difference between the magnitudes of the FFT coefficients of the original and the MP3 compressed/decompressed signals should be small.

To select the scaling factor *s,* the following points should be considered:

1. By increasing the scaling factor *s*, the error of the scaling and rounding step decreases and results in better perceptual distortion and lower capacity.
2. After the embedding steps, the magnitudes with value zero at which "1" is embedded are changed to 1/*s*. To obtain the secret bits after attacks, 1/*s* should be larger than difference between the original magnitudes and the attacked magnitudes.

Fig. 1 shows the flowchart for the selection of the tuning parameters. We select the frequency band from 15 kHz as low frequency and the cut-off frequency of MP3 as the high frequency. For example, the cut-off frequency of MP3-128 is around 17 kHz. If there are not enough magnitudes with zero value (after scaling and rounding), the selected frequency band should be expanded by decreasing the low frequency band. Since 8 kHz is start of high frequencies it is selected as the limitation. In the flowchart, the required capacity is denoted by *cap*, N_z is the number of zeros in selected frequency band and L_f is low frequency of the selected band. As the flowchart shows, decreasing *s* increases 1/*s* which is used for detecting secret information. In the initialization, the parameters *s* is equal to 10.0. Most FFT magnitudes in the selected frequency band are between 0 and 10, hence with *s* in the interval [0.1, 10.0] with 0.1 steps we do not miss significant magnitudes and increase the number of zeros in the frequency band simultaneously.

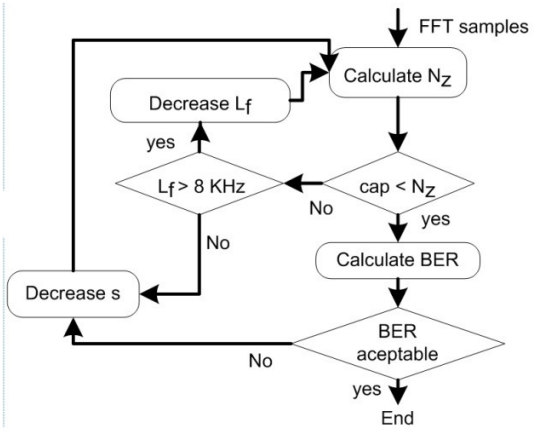


Fig. 1. Flowchart of tuning steps

The watermarking scheme presented here is positional. This means that the detector must be synchronized in order to recover the embedded bits correctly. In a real application, the cover signal would be divided into several blocks of a few seconds and it is essential that the detector can determine the position (the beginning sample) of each of these blocks. One of the most practical solutions to solve this problem is to use synchronization marks such that the detector can determine the beginning of each block. Several synchronization strategies have been described in the literature (for example [10, 11]) and any of them can be used together with the method described here in order to produce a practical self-synchronizing solution. A self-synchronized version of the proposed scheme (using the synchronization approach described in [10]) has been implemented and the results are shown in Section 3.

2.2 Embedding algorithm

The embedding steps are as follows:

1. Calculate the FFT of the audio signal. We can use the whole file (for short clips, *e.g.* with less than one minute) or blocks of a given length (*e.g.* 10 seconds) for longer files.
2. Use the *s* selected in the tuning step as a parameter to convert the FFT magnitudes in the selected frequency band to integer values (multiplying them by *s* and then rounding).
3. Scan all the integer FFT magnitudes in the selected band. If a magnitude is larger than zero, then increase it by one. After this step we have no magnitude with the value one.
4. Scan again all integer FFT magnitudes in selected band. When a zero magnitude is found, if the corresponding embedded bit is "1" add one to the magnitude. Otherwise, the magnitude is not changed. After this step all magnitudes with value zero or one represent an embedded bit.

5. Embed *s* and the frequency band limits in their reserved positions as described in the end of this section. This step must take into account security concerns, as detailed below.
6. The marked (FFT) signal is achieved by dividing all the magnitudes by *s*.
7. In the previous embedding steps, the FFT phases are not altered. The marked audio signal in the time domain is obtained by applying the inverse FFT with the new magnitudes and the original FFT phases.

2.3 Extraction algorithm

The watermark extraction is performed by using the FFT transform and the tuning parameters. Since the host audio signal is not required in the detection process, the detector is blind. The detection process can be summarized into the following steps:

1. Calculate the FFT of the marked audio signal.
2. Extract the *s* and the frequency band from special positions (this step requires the use of a secret key).
3. To achieve the scaled FFT magnitudes in selected frequency band, multiply them by *s*.
4. Scan all the scaled FFT magnitudes in the selected band. If a magnitude with value in the interval [0, 1/2) is found, then the corresponding embedded bit is equal to "0" and the restored magnitude equals to zero. If the magnitude value is in the interval [1/2, 3/2), then the corresponding embedded bit is equal to "1" and the restored magnitude equals to zero.
5. Scan all the scaled FFT magnitudes in the selected band. For each magnitude value in the interval [*k*+1/2, *k*+3/2), the restored magnitude equals to *k* (for $k > 1$).
6. The restored magnitudes are achieved by dividing them by *s*.
7. Finally, use the IFFT to achieve the restored audio signal.

For example, assume that the magnitudes at the selected frequency band are (0.9 0.4 0.2 0.1 1.4 0.15), $s = 2$ and the secret bit stream is "010". Table 1 summarizes all steps of embedding and extracting.

Table 1. embedding and extracting steps

steps		FFT magnitudes					
embedding	calculate FFT	0.9	0.4	0.2	0.1	1.4	0.15
	scaling	1.8	0.8	0.4	0.2	2.8	0.3
	rounding	2	1	0	0	3	0
	shifting	3	2	0	0	4	0
	embedding	3	2	**0**	**1**	4	**0**
	Dividing by *s*	1.5	1	0	0.5	2	0
extracting	Scaling	3	2	0	1	4	0
	Detecting	3	2	0	0	4	0
	Shifting back	2	1	0	0	3	0
	Dividing by *s*	1	0.5	0	0	1.5	0

It is worth pointing out that the tuning parameters (*s* and the frequency band) should be used in receiver to detect the secret information. In the embedding steps, a few special spaces are kept for saving tuning parameters. The FFT magnitudes in special frequencies such as 12, 13, 14, 15 and 16 kHz, which are reserved for scaling factor, are changed by the value of *s*. Consequently, in the receiver *s* will be available. Similarly, we use a 16-bit space available for embedding secret information which begins after the first FFT coefficient with a zero magnitude from a selected frequency (*e.g.* 15 kHz) to embed the values of the low and high frequencies of the selected frequency band. For example, if we embed at the frequency band from 12.3 to 16.7 kHz, we multiply them by ten and change them to binary values

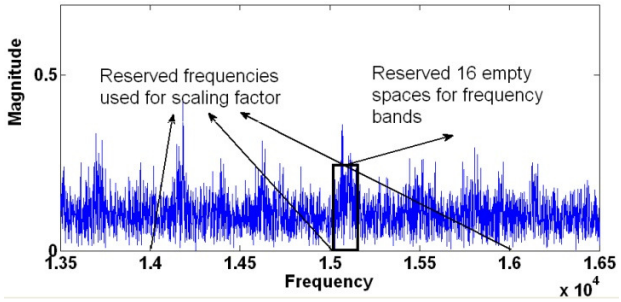


Fig. 2. Reserved positions for *s* and frequency bands

$(01111011)_2 = 123$ and $(10100111)_2 = 167$. After that, these binary streams are embedded in the free space found next to the selected frequency (15 kHz). The first bit of 123, is embedded in first available space after the selected position, and so on. Fig. 2. shows an example of the reserved positions for *s* and the frequency band.

The security of this method requires that the frequency band and the scaling factor *s* are not known by an attacker. Note that if an attacker does not know the scaling factor *s*, it will not be possible for him or her to analyze the values of the FFT magnitudes to determine the position of the embedded bits. For example, the rounded FFT magnitudes after scaling by $s = 0.2$ or $s = 0.4$ are completely different. If the attacker does not know the frequency band either, it becomes even more difficult for him or her to try to determine the interval of the FFT spectrum which carries the secret information.

In order to keep both the frequency band and the scaling parameter secret, there are two possibilities. The first one would be to consider both *s* and the frequency band as part of the secret key. In that case, the values of these parameters should **not** be embedded in the marked audio sequence and they should be transmitted as side information over a secured channel. At the receiver side, this information would be given to the extractor in order to recover the hidden data.

A second possibility introduces security even if the frequency band and the scaling parameter are embedded as suggested above. The following security measures are required:

1. The values of *s* and the frequency band should not be embedded as clear text. The bits which form the values must be scrambled using a Pseudo Random Binary Sequence (PRBS) generated through a secret key (seed) and the embedded values would be the result of an XOR sum of the bits of the original parameters (*s* and the frequency band limits) and the bits of the PRBS. The secret key would be also needed at the detector side in order to unscramble the values of *s* and the frequency band.
2. The FFT positions for embedding the values of *s* and the frequency bands should **not** be fixed. Instead of using fixed frequencies (like 12, 13, 14, 15 or 16 kHz) the position must be also generated with a Pseudo Random Number Generator (PRNG) in some interval (*e.g.* [12, 16] kHz) using a secret key (seed) which is required at both the sender and the receiver. This procedure makes it impossible for an attacker to destroy the values of *s* and the frequency band, since he or she cannot know the position of these data in the FFT spectrum. In order to destroy them, it would be required to disturb a wide interval of the spectrum and, thus, the quality of the attacked signal would also be damaged and would become unusable.

To increase security even further, a PRNG can also be used to change the secret bit stream to a scrambled stream. For example, the embedded bitstream can be constructed as the XOR sum of the real watermark and a PRBS. The seed of the PRNG would be required as a secret key both at the embedder and the detector. The usage of PRNG to increase the security of watermarking schemes is discussed in the literature (for example in [8]).

3. Experimental results

To evaluate the performance of the proposed method, male speech in English in *spme50_1*, violoncello in *vioo10_2*, trumpet in *trpt21_2*, soprano in *sopr44_1*, quartet in *quar48_1* have been selected from the Sound Quality Assessment Material (SQAM) [12]. Also, to consider the applicability of the scheme in a real scenario, the song "Thousand Yard Stare" (3:57) included in the album *Rust* by No, Really [13] has been selected. All audio clips are sampled at 44.1 kHz with 16 bits per sample and two channels. The experiments have been performed for each channel of the audio signals separately.

The Objective Difference Grade (ODG) is used to evaluate the transparency of the proposed algorithm. The ODG is one of the output values of the ITU-R BS.1387 PEAQ [14] standard, where ODG = 0 means no degradation and ODG = –4 means a very annoying distortion. Additionally, the OPERA software [15] based on the ITU-R BS.1387 has been used to compute this objective measure of quality.

Table 2 illustrates the tuning parameters, perceptual distortion and payload for six mono signals for BER equal zero under the MP3-128 attack. The tuning parameters have been chosen manually just to test the system for different tuning settings (*i.e.* we have not followed the flowchart depicted in Fig. 1). Table 3 shows the effect of various attacks, provided by the Stirmark Benchmark for Audio (SMBA) v1.0 [16], on ODG and BER for the five selected SQAM signals. *E.g.* the row "Amplify" shows that the changes in volume of the watermarked signal has BER equal to zero when the alteration of volume is within the interval [0.8, 1.45]. As described in Section 2, the frequency band and the scaling factor *s* are the two parameters of the method. These parameters were selected for each signal, then the embedding method has been applied, the Stirmark Benchmark for Audio (SMBA) software has been used to attack the marked files and, finally, the detection method has been performed for the attacked files. The ODG in Table 3 is calculated between the marked and the attacked-marked files. The parameters of the attacks are defined based on SMBA web site [16]. For example, in AddBrumm, 1-7000 shows the strength and 0-14000 shows the frequency. This row illustrates that any value in the range 1-7000 for the strength and 1-14000 for the frequency could be used without any change in BER. In fact, this table provides the worst and best results for the five test signals based on BER and, in the case with the same BER, based on the limitation of the parameters. The only attack in Table 3 which removes the hidden data is *FFT_Stat1*, which is able to remove the secret data for one of the SQAM files (BER = 27%). Note, however, that the ODG of this attack is extremely low (–4). This means that the attack does not only removes the hidden data, but also destroys the perceptual quality of the host signal.

Table 2: Parameters and results of 5 mono signals (BER=0 under MP3-128)

Audio File	***s***	***Frequency band (kHz)***	***ODG of marked signal***	***Payload (bps)***
spme50_1	0.3	15.5-16.9	– 0.18	1478
vioo10_2	0.5	8-16.9	– 0.34	8719
trpt21_2	3.5	8.5-16.9	– 0.39	7934
sopr44_1	0.35	10-16.9	– 0.30	7006
quar48_1	0.7	13.5-16.9	– 0.34	3177
Thousand Yard Stare	[0.1, 0.6]	15-16.9	– 0.78	1849

The SQAM files are short clips (30 seconds or less), and it is not necessary to use synchronization marks with them, since the whole file can be used in the embedding and extracting processes with short enough CPU time.

Table 3: Robustness test results for five SQAM selected files

Attack name	*State*	*ODG of attacked file*	*BER %*	*parameters*
No attack	–	0	0	–
MP3	–	–0.1	0	≥ 128
AddBrumm	best	–1.75	0	1-7000,1-14000
	worst	–3.58	0.5	1-7000,1-750
AddDynNoise	best	– 0.71	0	1-4
	worst	– 0.32	0.5	1
ADDFFTNoise	best	–1.47	0	4096,1-8800
	worst	–1.48	0	2,1-150
Addnoise	best	–1.46	0	1-115
	worst	–1.94	0	1-21
AddSinus	best	–1.25	0	No Limitation
	worst	–3.33	0	
Amplify	–	– 0.1	0	80-145
FFT_Invert	best	–1.61	0	No limitation
	worst	–3.66	0	
FFT_RealReverse	best	–3.56	0	No limitation
	worst	–3.96	0	
FFT_Stat1	best	–3.98	0	1024
	worst	-3.96	27	
Invert	best	–1.63	0	–
	worst	–3.63	0	
RC_LowPass	best	– 0.1	0	Selected frequency band in passing area
	worst	– 0.25	0	
RC_HighPass	best	–3.29	0	Selected frequency band in passing area
	worst	–3.59	0	

Table 4: Robustness test results for "Thousand Yard Stare"

Attack name	*ODG of attacked file*	*BER %*	*SYNC error*	*parameters*
No attack	0	0	0	–
MP3	–0.1	0	0/23	≥ 128
AddBrumm	–3.6	1	0/23	1-7500,1-5000
AddDynNoise	– 2.2	0	0/23	1-2
ADDFFTNoise	–1.1	0	0/23	4096,1-8800
Addnoise	–0.5	0	0/23	1-60
AddSinus	–3.3	0	0/23	No Limitation
Amplify	–0.1	0	0/23	80-145
FFT_Invert	–3.7	0	0/23	No limitation
FFT_RealReverse	–3.7	3	1/23	No limitation
FFT_Stat1	–3.9	36	3/23	1024
Invert	–3.7	0	0/23	–
RC_LowPass	–0.2	0	0/23	Selected frequency band in passing area
RC_HighPass	–3.7	0	0/23	Selected frequency band in passing area

In order to reduce computation time and memory usage, the near 4-minute long "Thousand Yard Stare" song was divided into 23 clips of 10 seconds each. Then, the synchronization method described in [10] and the embedding algorithm described in this paper was applied for each clip separately. For this song, 16 synchronization bits, "1 0 1 1 0 0 1 1 1 1 0 0 0 0 1 0" with a quantization factor equal to 0.125, were embedded in the first 80 samples of each clip and then the information watermark was embedded in the remaining samples of the 10-second segment. Finally all these 10-second clips were joined together to generate the marked signal. We have used different scaling factors in the range [0.1, 0.6] for different clips. The payload and transparency results given in Table 2 for this file consider the effect of both the synchronization codes and the information watermarks. Table 4 shows the effect of various attacks on ODG and BER for the marked "Thousand Yard Stare" signal. The whole file is attacked, then it is scanned in time domain to find the synchronization codes and, finally, the secret information of each clip is extracted. The "SYNC error" column shows the detection error of synchronization code after attacks, which shows that the synchronization algorithm [10] is robust against attacks. Figure 3 visualizes the test results. This plot shows how the capacity and perceptual distortion are changed with different tuning parameters. The BER for all test results under the MP3-128 attack on this plot is equal to zero.

Only a few attacks, such as low pass filter ─which only leaves low frequencies unaltered─ with a cut-off frequency less than 6 kHz damage the hidden data. However, the ODG of this attack is extremely low (about –3.5, *i.e.* very annoying). This means that the attack does not only remove the hidden data, but also destroys the perceptual quality of the host signal. On the other hand, if the cut-off frequency is larger than 8 kHz the BER is about zero and the ODG of attack is in the acceptable range.

A very relevant issue in audio watermarking is computation time. As FFT is a fast transform, this method is very useful for real-time applications. Table 5 illustrates the embedding and extracting times and compares them with the computation time of FFT and the Daubechies wavelet transform. The results for the song "Thousand Yard Stare" are the average of all the 10-second clips. It is worth mentioning that these computation times have been obtained with an Intel (R) core (TM) 2 Duo 2.2 GHz CPU and 2 GB of RAM memory. It can be noticed that the extracting time is one order of magnitude smaller than the file playing time. Thus, it is perfectly possible to recover the embedded data in a real-time scenario.

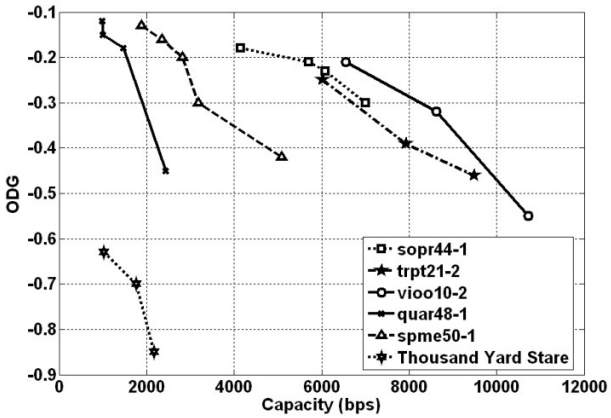


Fig 3. Comparison between payload (bps) and Transparency (ODG) for BER=0 under MP3 attack (bitrate 128 kbps)

Table 5: Computation time

Audio File	*Time (sec)*	*FFT time*	*Db2 time*	*Embedding time (sec)*	*Extracting time (sec)*
spme50_1	18	0.46	0.50	1.78	1.67
vioo10_2	30	0.57	0.47	4.30	4.01
trpt21_2	17.8	0.58	0.48	1.89	1.76
sopr44_1	23.5	0.60	0.53	3.65	3.30
quar48_1	23	0.59	0.48	2.59	2.29
Thousand Yard Stare	23×10	0.17	0.15	0.69	0.59

The method proposed in this paper has been compared with several recent audio watermarking strategies. It must be taken into account that none of the works in the reviewed literature produce capacity of the order of 5 kbps, such as the proposed scheme. All the audio data hiding schemes which produce very high capacity are fragile against signal processing attacks. Because of this, it is not possible to establish a comparison of the proposed scheme with other audio watermarking schemes which are similar to it as capacity is concerned. Hence, we have chosen a few recent and relevant audio watermarking schemes in the literature. In Table 6, we compare the performance of the proposed watermarking algorithm and several recent audio watermarking strategies robust against the MP3 attack. The results are given for SQAM files. [4-6, 9] use SQAM [12] files for evaluating their suggested schemes. All the schemes in this table are robust against MP3 compression with a 128 kbps bitrate. Under this attack, the BER is equal to zero for all the compared schemes.

[7] Evaluates distortion by mean opinion score (MOS), which is a subjective measurement, and achieves transparency between imperceptible and perceptible but not annoying, MOS = 4.7. [4, 5, 9] have a low capacity but are robust against common attacks.

Capacity, robustness and transparency are the three main properties of an audio watermarking scheme. Considering a trade-off between these properties is necessary. *E.g.* [4] proposed a very robust, low capacity and high distortion scheme. However [7] and the proposed scheme introduce high capacity and low distortion technique but they are not as robust as the low-capacity method described in [4].

This comparison shows the superiority in both capacity and imperceptibility of the suggested method with respect to other schemes in the literature. This is particularly relevant, since the proposed scheme is able of embedding much more information and, at the same time, introduces less distortion in the marked file.

Table 6: Comparison of different watermarking algorithms

Algorithm	Capacity (bps)	Imperceptibility in SNR (dB)	Imperceptibility (ODG)
[4]	2	42.8 to 44.4	–1.66 to –1.88
[9]	2.3	Not reported	Not reported
[5]	4.3	29.5	Not reported
[6]	Coded image with 9×35 bits	24.3 to 34.4	–0.36 to –0.72
[7]	689	Not reported	Not reported
proposed	1478 to 8719	35.2 to 44.6	–0.18 to –0.78

In the last few years, very good results in image data hiding have been published. Ni *et al*. [17] proposed a high capacity data hiding with very low distortion. For general test images such as Lena and Baboon they embedded about 5 kbit in the whole image with PSNR above 40 dB, *i.e.* the embedding rate for a 512×512×8 image is 0.0024 bits of information per each image bit. The proposed method in this paper embeds about 5 kbit in a second. It means 5 kbit in 44100×16 bits that equals to 0.0071 per audio bit. If we consider the compression rate of MP3-128 (about 12:1), since this method completely robust against MP3-128, the embedding rate for each bit of audio sample equals **0.085**, that is 35 times more than the information bit rate achieved with the image method of Ni *et al*. Some other image data-hiding schemes have been presented [18] increasing the payload up to 0.02 bits per image bit. Even in this case, the suggested audio scheme presented here achieves more than four times that capacity.

4. Conclusion

In this paper, we describe a high-capacity watermarking algorithm for digital audio which is robust against common audio signal processing. A scaling factor (*s*) and the selected frequency band to embed the hidden information are the two parameters of this method which regulate the capacity, the perceptual distortion and the robustness of the scheme. Furthermore, the suggested scheme is blind, since it does not need the original signal for extracting the hidden bits. The experimental results show that this scheme has a high capacity (about 5 kbps) without significant perceptual distortion and provides robustness against common signal processing attacks such

as noise, filtering or MPEG compression (MP3). Besides, the proposed method achieves a higher embedded bit to host bit rate than recent image data hiding methods. In addition, the CPU time required by the proposed scheme is short enough to use the scheme in real-time applications.

References

[1] N. Lie, L. C. Chang, "Multiple Watermarks for Stereo Audio Signals Using Phase-Modulation Techniques", *IEEE Trans. Signal Processing*, Vol. 53, No. 2, pp. 806–815, Feb. 2005.
[2] H. J. Kim, Y. H. Choi, " A novel echo hiding scheme with backward and forward kernels", *IEEE Trans. Circuit and Systems*, Vol. 13, pp. 885-889, Aug. 2003.
[3] S. Esmaili, S. Krishnan, K. Raahemifar, "A novel spread spectrum audio watermarking scheme based on time - frequency characteristics". *IEEE Conf. Electrical and Computer Engineering*, Vol. 3, pp. 1963-1966, May 2003.
[4] S. Xiang, H.J. Kim, J. Huang, "Audio watermarking robust against time-scale modification and MP3 compression," *Signal Processing*, Vol.88 n.10, pp.2372-2387, October, 2008.
[5] M. Mansour and A. Tewfik, "Data embedding in audio using time-scale modification," *IEEE Trans. Speech Audio Process.*, Vol. 13, no. 3, pp. 432–440, 2005.
[6] Y.Q. Lin, W.H. Abdulla, "Multiple Scrambling and Adaptive Synchronization for Audio Watermarking", *IWDW*, LNCS 3304, Springer-Verlag, pp. 456-469, 2007.
[7] J. J. Garcia-Hernandez, M. Nakano-Miyatake and H. Perez-Meana, "Data hiding in audio signal using Rational Dither Modulation", *IEICE Electron. Express*, Vol. 5, No. 7, pp.217-222, 2008.
[8] D. Megías, J. Herrera-Joancomartí, and J. Minguillón. "Total disclosure of the embedding and detection algorithms for a secure digital watermarking scheme for audio". *Proceedings of the Seventh International Conference on Information and Communication Security*, pp. 427-440, Beijing, China, December 2005.
[9] W. Li, X. Xue, "Content based localized robust audio watermarking robust against time scale modification" *IEEE Trans. Multimedia*, Vol. 8, No. 1, pp. 60-69, Feb. 2006.
[10] X.-Y. Wang and H. Zhao. "A novel synchronization invariant audio watermarking scheme based on DWT and DCT". *IEEE Trans. on Signal Processing*, Vol. 54(12), pp. 4835–4840, December 2006.
[11] Y. Lin and W. Abdulla. "A secure and robust audio watermarking scheme using multiple scrambling and adaptive synchronization". In Proceedings of the *6th International Conference on Information, Communications & Signal Processing*, pp. 1–5, 2007.
[12] SQAM Sound Quality Assessment Material, http://andrew.csie.ncyu.edu.tw/html/mpeg4/sound.media.mit.edu/mpeg4/audio/sqam/index.html
[13]No, Really, "Rust". http://www.jamendo.com/en/album/7365.
[14] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerens, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, and B. Feiten, "PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality," *Journal of the AES*, vol. 48(1/2), pp. 3–29, 2000.
[15] OPTICOM OPERA software site. http://www.opticom.de/products/opera.html
[16] Stirmark Benchmark for Audio. http://wwwiti.cs.uni-magdeburg.de/~alang/smba.php.
[17] Ni Zhicheng, Y.Q. Shi, N.Ansari, and W. Su, "Reversible data hiding," *IEEE Trans. on Circuits and Systems for Video technology*, 16(3):354–362, March 2006.
[18] D. M. Thodi and J. J. Rodriguez, Expansion embedding techniques for reversible watermarking, *IEEE Trans. on Image Processing*, Vol. 16, no. 3, pp.721–730, 2007.

2.7 DWT–based high capacity audio watermarking

81. **M. Fallahpour,** David Megías, "DWT–based high capacity audio watermarking" *IEICE Trans. Fundamentals*, Vol.E93-A, No.01, pp.-, Jan. 2010. (Impact factor 0.43)

IEICE TRANS. FUNDAMENTALS , VOL. E85-A/B/C/D, No. xx JULY

LETTER

DWT–based high capacity audio watermarking

M. Fallahpour,[†] *student member* and D. Megias[†]

Summary This letter suggests a novel high capacity robust audio watermarking algorithm by using the high frequency band of the wavelet decomposition, for which the human auditory system (HAS) is not very sensitive to alteration. The main idea is to divide the high frequency band into frames and then, for embedding, the wavelet samples are changed based on the average of the relevant frame. The experimental results show that the method has very high capacity (about 5.5 kbps), without significant perceptual distortion (ODG in [–1 ,0] and SNR about 33 dB) and provides robustness against common audio signal processing such as add noise, filtering, echo and MPEG compression (MP3).

Keywords: ***Audio watermarking, wavelet transform.***

1. Introduction

Protecting data from unauthorized copying and distribution in an imperceptible manner, based on the properties of the human auditory system (HAS), is the aim of digital audio watermarking.

A wide work has been carried out in understanding the characteristics of the HAS and applying this knowledge to audio compression and audio watermarking. Based on the HAS, the human hearing sensitivity in higher frequencies is lower than in middle frequencies. It is thus clear that, by embedding data in the high frequency band, the distortion will be mostly inaudible and, hence, more transparency can be achieved.

In fact, audio watermarking schemes take advantage of the properties of the HAS and different transforms, resulting in various techniques such as embedding algorithms based on low-bit coding, echo, patchwork [1], rational dither modulation [2], Fourier transform [3], quantization [4, 5, 7] and the wavelet transform [6,8].

Among the existing transforms, the wavelet transform has many advantages in audio signal processing. Its inherent frequency multi-resolution and logarithmic decomposition of frequency bands resembles the human perception of frequencies, since it provides the decomposition to mimic the critical band structure of the HAS.

In the proposed scheme, the last high frequency band of the third level wavelet decomposition (DDD), where the HAS is not very sensitive to alteration, is used for embedding. This band of wavelet samples is divided into frames and then, the average of the absolute values of each frame's samples is computed. After that, in the embedding process, all wavelet samples are scanned and if each sample satisfies a given condition then the corresponding secret bit is embedded into it. The corresponding secret bit is embedded into a single wavelet sample and the next secret bit is embedded into the next suitable sample. The idea of dividing the wavelet samples into frames and calculate the average of samples in each frame is used to discover suitable wavelet samples for embedding and propose an appropriate value for the embedded samples. The samples selected for embedding are changed based on the absolute values of each frame's samples. If the corresponding secret bit is "0", the suitable sample is changed to $-m_i$ and, if it is "1", the corresponding sample is changed it to $+m_i$, where m_i is the average of the i-th frame.

The experimental results show that high capacity, remarkable transparency and robustness against most of common attacks are achieved.

The rest of the letter is organised as follows. In Section 2, the proposed method is presented. In Section 3, the experimental results are shown. Finally, Section 4 summarizes the most relevant conclusions of this research.

2. Proposed scheme

The embedding and extracting processes of the proposed scheme are described in this section.

2.1 Embedding

The embedding steps are described below.

1. Compute the third level wavelet transform of the original signal.
2. Divide the DDD samples into frames of a given length and, based on the average of the absolute values of each frame's samples, compute the average m_i for each frame.

$$m_i = \frac{1}{s} \sum_{j=(i-1)s+1}^{is} |c_j| \qquad (1)$$

   Where $\{c_j\}$ are the wavelet coefficients of the high-frequency sub-band (DDD), s is the frame size and m_i is the average of the i-th frame.
3. The marked wavelet coefficients $\{c_j'\}$ are achieved by using equation (2).

$$c_j' = \begin{cases} m_i & |c_j/m_i| < k,\ w_l = 1 \\ -m_i & |c_j/m_i| < k,\ w_l = 0 \\ c_j & |c_j/m_i| \geq k \end{cases} \qquad (2)$$

Manuscript received January xx, 20xx.
Manuscript revised March xx, 20xx.
[†] The author is with NTT , Musashino-shi, 180-8585
[††] The author is with IEICE Office, Minato-ku, Tokyo, 105-0011 Japan.

Where $i = \lfloor j/s \rfloor + 1$, since each frame has a particular average (m_i), w_l is the l-th bit of the secret stream, k is the embedding interval ($k > 2$) and $\lfloor \cdot \rfloor$ denotes the floor function. *I.e.* if c_j in $[-km_i, km_i]$ then, depending on the secret bit, it is changed to $-m_i$ or $+m_i$. Each secret bit is embedded in a single suitable coefficient and thus, after embedding the bit, the index l is incremented and the next secret bit is embedded in the next suitable coefficient. It is worth pointing out that each secret bit is embedded into each appropriate wavelet sample, not into a frame, thus the embedding capacity is depends on the number of suitable wavelet samples and not on the number of frames.

4. Finally, the inverse DWT is applied to the modified wavelet coefficients to get the marked audio signal.

The modified area of DWT coefficients for each frame is $[-km_i, km_i]$ which is determined by the absolute mean value of each frame and the embedding interval k. By increasing k, the interval is extended and the number of modified coefficients which satisfy that $|c_j/m_i| <$ is increased, thus capacity and distortion become greater. To adjust robustness and transparency, a scaling factor α, which defines the strength of watermark ($0.5 < \alpha < k$), is used. In fact, in equation 2, instead of changing c_j' to m_i, it can be changed to αm_i.

2.2 Extracting

In the receiver, m_i' is calculated by Equation (1) for the marked samples and an interval is defined such that, if c_j' is in the interval, a secret bit can be extracted. The secret bit stream is retrieved by using equation (3)

$$w_l' = \begin{cases} 1 & 0 \le c_j'/m_i' \le ((k+\alpha)/2) \\ 0 & -((k+\alpha)/2) \le c_j'/m_i' < 0 \end{cases} \quad (3)$$

Where c_j' is the sample of the high frequency band of the third level wavelet decomposition (DDD) of the marked signal, α is the strength of watermark and w_l' is the l-th bit of the extracted secret stream. *E.g.* if $k = 2$ and $\alpha = 1$ then if c_j' in $[0, 1.5m_i']$ the secret bit is "1", and if it is in $[-1.5m_i'.0)$, then the secret bit is "0".

Note that, in the sender/coder, the embedding intervals and the embedded values are obtained in terms of the average of the samples in each frame. Thus, in the receiver/decoder, we need calculate the average of each frame to extract the secret bits. Since the DWT samples in the interval $[-km_i, km_i,]$ are changed to αm_i or $-\alpha m_i$, it is clear that the average of the absolute values is equal to αm_i in the receiver. If the signal is distorted by attacks, the absolute mean of the coefficients m_i' will be slightly altered. However, the experimental results show that this change does not affect the extraction process since an interval, not a constant number, is used for extracting. *E.g.* under the MP3-128 compression attack, the variation is about 5% which is acceptable for extraction. The suggested algorithm is blind, since the original signal values are not required in the receiver

In a real application, the cover signal would be divided into several blocks of a few seconds and it is essential that the detector can determine the position (the beginning sample) of each of these blocks. One of the most practical solutions to solve this problem is to use synchronization marks such that the detector can determine the beginning of each block. *E.g.* [8] can be used with the method described here in order to produce a practical self-synchronizing solution.

To increase security, pseudo-random number generators (PRNG) can be used to change the secret bit stream to a stream which makes more difficult for an attacker to extract secret information from. For example, the embedded bitstream can be constructed as the XOR sum of the real watermark and a pseudo-random bit stream. The seed of the PRNG would be required as a secret key both at the embedder and the detector.

3. Experimental results

To show the performance of the proposed scheme and to consider the applicability of the scheme in a real scenario, five songs (RIFF-WAVE files) included in the album *Rust* by No, Really [9] with genre *electric folk* have been selected. All audio clips are sampled at 44.1 kHz with 16 bits per sample and two channels. The three-level wavelet decomposition is implemented with the 8-coefficient Daubechies wavelet (db8) filter. The experiments have been performed for each channel of the audio signals separately. We provide imperceptibility results both as SNR and ODG where ODG = 0 means no degradation and ODG = –4 means a very annoying distortion. The SNR is provided only for comparison with other works, but ODG is a more appropriate measurement of audio distortions, since it is assumed to provide an accurate model of the subjective difference grade (SDG) results which may be obtained by a group of human listeners. The SNR results are computed using the whole (original and marked) files, whereas the ODG results are provided using the advanced ITU-R BS.1387 standard [10] as implemented in the Opera software [11] (the average of measurements taken in frames of 1024 samples).

Table 1 shows the perceptual distortion and the payload obtained for the five songs with tuning settings which lead to BER equal to zero (or near zero) under the attacks described in Table 2. The values of the parameters are k = 6, α = 3 and the frame size is equal to 10. In fact, by selecting k = 6, almost all wavelet samples in the DDD area are used for embedding. We have used several random bits for embedding leading to different transparency results which are shown in the ODG column.

Table 1: Results of 5 mono signals (robust against table 2 attacks)

Audio File	*Time (m:sec)*	*SNR (dB)*	*ODG of marked*	*payload (bps)*
Beginning of the End	3:16	33 to 38.1	–0.3 to –0.7	5502
Citizen, go back to sleep	1:57	29.8 to 33.3	–0.5 to –0.7	5499
Go	1:51	30 to 34.1	–0.5 to –0.8	5504
Thousand Yard Stare	3:57	36.1 to 39	–0.1 to –0.6	5501
Rust	2:33	27.2 to 32.1	–0.4 to –0.7	5499
average	**2:43**	**33**	**–0.5**	**5501**

Note that all the results have an ODG between 0 (not perceptible) and –1 (not annoying), the average SNR is 33 dB and capacity is around 5.5 kbps in all the experiments. The proposed method is thus able to provide large capacity whilst keeping imperceptibility in the admitted range (–1 to 0).

Table 2 illustrates the effect of various attacks provided in the Stirmark Benchmark for Audio v1.0 [12] on ODG and BER for the five audio signals of Table 1. For these results, the embedding method has been applied, then the SMBA software has been used to attack the marked files and, finally, the detection method has been performed for the attacked files. The ODG in table 2 is calculated between the marked and the attacked-marked files. The parameters of the attacks are defined as detailed on the SMBA web site [12] for the proposed scheme. Other schemes may use different parameters. For example, for the AddBrumm attack, 1 to 6 k shows the strength and 1to 7 k shows the frequency. This row illustrates that any value in the range 1to 6 k for the strength and 1 to 7 k for the frequency could be used without any change in BER. For the RC_LowPass attack, the parameter defines the cut-off frequency in the range [2 kHz, 22 kHz], and the BER is in the range [0, 4%] for all tested frequencies and not only for the default cut-off frequency in [12] (15 kHz). In fact, this table shows the range (the worst and best values) of ODG and BER for the five test signals.

This scheme uses the high frequency band of the wavelet coefficients for embedding. Hence, it may seem that it would be fragile against attacks which manipulate the high frequency bands. In Table 3, The MP3 and RC low-pass filter attacks are analyzed in depth with different types

Table 2: Robustness test results for five selected files and comparison with schemes in this literature

Attack name	*parameters*	**ODG of attacked file**	***BER %***					
			proposed	*[1]*	*[2]*	*[3]*	*[6]*	*[7]*
AddBrumm	1 to 6k, 1 to 7k	–3.3 to –3.7	0 to1	–	0	0 to 1	–	–
AddDynNoise	1 to 2	–2 to –2.3	2 to 7	–	2	0 to 8	–	–
ADDFFTNoise	2048,400	–0.3 to –0.1	0 to 2	–	1	1 to 2	–	–
Addnoise	1 to 20	–0.8 to –0.4	0 to 4	2	1	0 to 1	–	0
AddSinus	1 to 5k , 1 to 7k	–3.1 to –2.5	0	–	0	0	–	–
Amplify	10 to 200	–0.2 to –0.1	0 to 1	–	0	0	–	–
BassBoost	1 to 50,1 to 50	–3.8 to –3.2	0 to 2	–	–	0	–	–
Echo	1 to 10	–3 to –2.3	0 to 3	1.2	63	0 to 1	–	6
FFT_HLPassQuick	2048,1 to 10k,18k to 22k	–3.6 to –3.3	0 to 2	–	5	1 to 4	–	–
FFT_Invert	1024	–3.8 to –3.1	0	–	2	1 to 2	–	–
FFT_RealReverse	2 , 2048	–3.5 to –3	11 to 24	–	–	–	–	–
FFT_Stat1	2 , 2048	–3.6 to –2.9	14 to 23	–	1	–	–	–
invert	-	–3.3 to –2.8	0	–	–	0	–	–
Resampling	44/22/44	–2.2 to –1.8	38 to 47	1	0	5	0	0
LSBZero	-	– 0.1 to 0.0	0	–	0	0	–	0
MP3	128	– 0.2 to 0.0	0 to 3	0.3	–	0 to 5	0	–
Noise_Max	1 to 2,1 to 14k,1 to 500	– 0.3 to –0.1	0 to 1	–	–	0 to 1	–	–
Pitchscale	1.1	–3.7 to –3.3	32 to 61	–	–	0 to 1	–	–
RC_HighPass	1k to 22k	–3.7 to –0.1	0 to 1	–	–	0 to 1	–	–
RC_LowPass	2k to 22k	–3.2 to –0.2	0 to 4	2	0	0	0	3
Smoth	–	–3.7 to –3.3	15 to 31	–	–	–	–	–
Stat1	–	–2.3 to –1.4	21 to 44	–	8	–	–	–
TimeStretch	1.05	–3.8 to –3.4	44 to 65	–	–	–	–	–
quantization	16 to 12	–0.5 to –0.2	2 to 4	0.5	–	–	0	0

Table 3. Robustness results for variety audio types under MP3 and RC Lowpass

MP3 rate	256	160	128	96	64
BER	0 to 1	0 to 4	0 to 9	7 to 17	12 to 27
ODG of attacked file	–0.1 to 0.0	–0.2 to 0.0	–0.3 to –0.1	–0.6 to –0.3	–0.8 to –0.5
Cut of frequency of RC_lowpass filter(kHz)	20	15	10	5	2
BER	0 to1	0 to 1	0 to 2	1 to 9	4 to 18
ODG of attacked file	–0.2 to –0.0	–0.4 to 0.0	–0.6 to –0.2	–1.7 to –0.7	–3.6 to –2.9

of audio clips. This table shows that the BER is increased by decreasing the MP3 rate also by decreasing cut-off frequency of the low-pass filter. As mentioned above, in all watermarking schemes based on application properties of a technique should be chosen. For instance in the proposed scheme based on the specific application, the embedding interval and scale factor may be changed. *E.g.* if $k = 5$ and $\alpha = 2$ for the clip "Beginning of the End", ODG = –0.22 and BER under MP3-128 is 0.12 but for $k = 5$ and $\alpha = 3$, ODG = –0.38 and BER = 0.07. These examples show the necessity of considering a trade-off between capacity, transparency and robustness in all audio watermarking schemes, included this technique. Furthermore, by repeating secret bits and using error correction codes in all watermarking schemes, robustness is increased at the price of reducing capacity. For example, under the MP3-96 attack, if we repeat the secret bit three times, then BER will be decreased by about 50% (since two or three bit errors would be required to change a secret bit) while capacity would be decreased to 33%.

A few attacks such as Pitchscale and TimeStretch in Table 2 and RC Lowpass filter with cut-off frequency less than 2 KHz in Table 3 remove the hidden data (BER > 15%). Note, however, that the ODG of these attacks are extremely low (about –3.5, *i.e.* very annoying). This means that the attack does not only removes the hidden data, but also destroys the perceptual quality of the host signal. Hence, the suggested scheme provides a convenient trade-off between transparency and robustness for very high capacity (as shown in Table 4).

Table 4: The comparison of different watermarking algorithms

Algorithm	***Audio File***	***SNR (dB)***	***ODG of marked***	***Payload (bps)***
[1]	Song	25	–	43
[2]	Song	–	–	689
[3]	Song	30.5	–0.6	2996
[6]	Song	30	–	172
[7]	Classical music	25	–	176
proposed	Song	33	–0.5	5501

In Table 4, we compare the performance of recent audio watermarking strategies which are robust against the common attacks with the proposed method. [2] measures distortion by mean opinion score (MOS), which is a subjective measurement, and achieves transparency between imperceptible and perceptible but not annoying, (MOS = 4.7). [7] has a low capacity but is robust against most common attacks. It is worth pointing out that the suggested scheme outperforms the capacity and transparency results of the reviewed methods. In fact, the capacity results are much higher than those of the literature, and only comparable with those of [3]. Note, also, that [3] does not report robustness results for several of the attacks considered in Table 2.

Although the schemes in the literature use different audio signals and attack parameters, we try to summarize abilities of each algorithm in capacity of embedding secret information and transparency in Table 4 and robustness against attacks in Table 2. The comparison shows that the compared schemes are robust against common attacks and, also, the transparency is in an acceptable range. However, the capacity of most of the chosen schemes is about a few hundred bits per second, whereas the suggested scheme provides 5.5 kbps. Furthermore this comparison proves that the capacity of the proposed scheme is very remarkable whilst keeping the transparency and BER (against attacks) in an acceptable range.

Using frames of wavelet samples results in increasing robustness against several attacks, since the average of the samples is more invariant against attacks than the value of each individual sample. Thus, by increasing the frame size, better robustness can be achieved. However, increasing the frame size implies that the same value would be used for more samples, decreasing the accuracy and transparency (audio quality) of the marked signal. In our experiments, a frame size equal to 10 has provided remarkable transparency and acceptable robustness. Other applications may require different values for the frame size.

4. Conclusion

Using the high frequency band of the wavelet decomposition where human auditory system (HAS) is not very sensitive to alteration leads to a high-capacity watermarking algorithm for digital audio which is robust against common audio signal processing. The suggested scheme divides the high frequency band (DDD) of the wavelet transform into frames and uses the frames' average which are the same in the sender and receiver, resulting in a blind scheme. The experimental results show that this scheme has very high capacity (about 5.5 kbps) without significant perceptual distortion and provides robustness against common signal processing attacks such as noise, echo, filtering or MPEG compression (MP3). A comparison of the suggested method with recent results in the literature also shows that the suggested scheme outperforms other works as transparency and capacity are concerned, whilst providing robustness against common signal processing attacks.

References

[1] H. Kang, K. Yamaguchi, B. Kurkoski, K. Yamaguchi, and K. Kobayashi, "Full-Index-Embedding Patchwork Algorithm for Audio Watermarking", *IEICE TRANS. on Information and Systems*, E91-D(11):2731-2734, 2008

[2] J. J. Garcia-Hernandez, M. Nakano-Miyatake and H. Perez-Meana, "Data hiding in audio signal using Rational Dither Modulation", *IEICE Electron. Express*, Vol. 5, No. 7, pp.217-222, 2008.

[3] M. Fallahpour; D. Megías, "High capacity audio watermarking using FFT amplitude interpolation" *IEICE Electron. Express*, Vol. 6, No. 14, pp. 1057-1063, 2009.
[4] B. Chen and G. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.
[5] Z. Xu, K. Wang, X.h. Qiao, "Digital Audio Watermarking Algorithm Based On Quantizing Coefficients," *IEEE Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing* 0-7695-2745-0/06 , 2006.
[6] M. Pooyan, A. Delforouzi, "Adaptive and robust audio watermarking in wavelet domain" *Third International Conference on International Information Hiding and Multimedia Signal Processing* , V2 Pages 287-290, 2007
[7] M. A. Akhaee, M. J. Saberian, S. Feizi, F. Marvasti. "Robust Audio Data Hiding Using Correlated Quantization With Histogram-Based Detector" *IEEE TRANS. ON Multimedia*,V11, P 1-9, 2009.
[8] X.-Y. Wang and H. Zhao. "A novel synchronization invariant audio watermarking scheme based on DWT and DCT". *IEEE Trans. on Signal Processing*, 54(12):4835–4840, 2006.
[9] http://www.jamendo.com/en/album/7365
[10] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerens, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, and B. Feiten, "PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality," *Journal of the AES*, vol. 48(1/2), pp. 3–29, 2000.
[11] OPTICOM OPERA software site. http://www.opticom.de/products/opera.html.
[12] Stirmark Benchmark for Audio. http://wwwiti.cs.uni-magdeburg.de/~alang/smba.php.

CHAPTER 3

CONCLUSIONS AND FUTURE RESEARCH

3.1 Conclusions

Robust and high capacity digital audio watermarking algorithms and high capacity image data hiding methods are studied in this thesis. The main results of this work are the development of novel audio watermarking and image data hiding algorithms with state-of-the-art performance, high capacity and transparency for image data hiding and robustness, high capacity and low distortion for audio watermarking. Based on the media, the research results can be divided into image and audio parts.

3.1.1 Image contents

In the image data hiding area, capacity and PSNR are the two main properties and robustness is not relevant. It must be taken into account that capacity and PSNR are two parameters which change against each other, *i.e.* when capacity is increased, the PSNR is decreased and vice versa. Using the histogram of an image is a new idea in image data hiding [63]. However, the total capacity achieved by using the maximum point of the histogram of the original image to embed secret information is not very large. The aim of this research was to obtain a narrow histogram after a modification step (such as prediction or tilling), resulting in an increased capacity. Moreover, reversibility is the main challenge in this kind of methods such as medical applications, since it is required that not only the hidden message but also the original (unmarked) image can be extracted at the detector side.

The peak point of the histogram of the image is used for embedding information. The value of this peak identifies the capacity payload. Based on the idea of using the histogram to embed secret information, prediction and tilling techniques are used to achieve good capacity and transparency. Parts 1, 2 and 3 summarise the results of the published papers [68, 92, 69] contributed by the author of this thesis.

Part 1

"Reversible Data Hiding Based On H.264/AVC Intra Prediction" [68] presents a novel high-capacity reversible data hiding algorithm, called shifted intra prediction error (SIPE), which is based on shifting the differences between the cover image pixels and their predictions. Large capacity of embedded data (15-120 kbits for a 512 × 512 greyscale image), PSNR above 48 dB, applicability to almost all types of images, simplicity and short execution time are the key features of this algorithm. The SIPE method is applicable for error resilient solutions in the H.264 advanced video coding. Therefore, the SIPE method has several advantages with respect to the methods reviewed in this thesis, in which the suggested algorithms are considered among the best methods in lossless data hiding.

Part 2

"Reversible image data hiding based on gradient adjusted prediction" [92] illustrates a new lossless data hiding method for digital images using an image prediction technique. In the proposed method, the idea of shifting the prediction error is used. The use of gradient-adjusted prediction (GAP), which is one of the best casual predictors, results in excellent results. Prediction based on neighbouring pixels, not neighbouring blocks, and the use of seven neighbour pixels lead to a narrower histogram compared to the histogram achieved using the H264/AVC intra prediction error. Against H264/AVC, the GAP generates real numbers, not integer numbers, which leads to the use of intervals for embedding. This method is able to embed a huge amount of data (15-140 kbits for a 512 × 512 × 8 greyscale image) whilst the PSNR of the marked image versus the original image is very high.

A comparison in terms of capacity between the prediction-based schemes described in [68, 92] with [63] shows an increase in capacity of about 500-1000%, ,where [63] proposed the best results up to 2006.

Part 3

In "High capacity, reversible data hiding in medical images" [69], a data-hiding scheme based on the shifted histogram is shown to be better if it is applied to image tiles rather than whole the image. This not only improves the data-hiding capacity, but also increases the quality of the marked image. This is mainly due to the fact that the sum of the peaks of the individual pixel intensity histograms is greater than the single peak of

the histogram of the whole image Besides, the individual histograms are much narrower and sharper than the histogram of the whole image, creating more possibility for zeros, as well as allowing for greater distances between the peak points of the tiles. This way, while the region-of-interest data can be free from disturbances, the embedded data can also be hidden according to the perceptual characteristics of the human visual system, and the number of minimum (zero) points in each image tile is lower than that of the whole image. This results in an improved marked image quality while maintaining high data embedding capacity. Compared with the data hiding method applied to the whole image [63], this scheme can result in a 30%-200% capacity improvement with still better image quality, depending on the medical image content. Finally, the use of the individual histograms (for different tiles) makes it possible to distribute the embedded bits among the image.

3.1.2 Audio contents

In the audio watermarking field, capacity payload, transparency, robustness and computational efficiency are the main objectives. One of main challenges in audio processing is to estimate the sensitivity of the human ear to audio distortion. To overcome this challenge, the use of a psychoacoustic model is quite useful to design watermarking algorithms, as shown in the designed FFT-based schemes [78, 80]. The human ear is sensitive to frequency and, for this reason, psychoacoustic models are based on frequency domain data. In addition, to achieve robustness, the frequency domain is also a better choice compared to the time domain. In this thesis, the frequency domain, and more precisely, the Fast Fourier Transform (FFT) and the Digital Wavelet Transform (DWT) have been chosen since they are more efficient than other transforms and provide more robustness than the methods designed in the time domain.

Different attacks produce various changes in audio. One of the most important attacks in audio is compression. To defeat MPEG compression (MP3), comparing the original with a compressed/decompressed signal to find a safe area for embedding is a convenient possibility, and this idea is exploited in the proposed algorithm [80]. In addition, the use of different techniques like differences, predictions, interpolation and spatial transforms in FFT resulted in high capacity methods which lead to publish three different papers [78- 80]. The fourth paper takes advantage of the DWT to design a high capacity, transparent and robust watermarking scheme [81]. Parts 1, 2, 3 and 4

summarise the results of the published papers [78-81] contributed by the author of this thesis.

Part 1

In "High capacity method for real-time audio data hiding using the FFT transform" [78], the algorithm was implemented taking special care for the efficient usage of the two restricted resources of computer systems: memory space and CPU time. This method offers to the industrial user the capability of watermark embedding and detection in time immediately comparable to the real playing time of the original audio file, while the end user/audience does not find any artifacts or delays hearing the marked audio file. The high capacity data hiding algorithm for digital audio described in this contribution is appropriate for real-time applications. A scaling coefficient (q) and the selected frequency band to embed the hidden information into it are the two main parameters of this method. The experimental results show that using different values for q and the frequency band lead to different capacity, perceptual distortion (ODG) and bit error rates (BER) of detection after MP3 compression. Furthermore, the suggested scheme is blind, since it does not need the original signal for extracting the hidden bits. The experimental results show that this scheme has a very good capacity (5 kbps or above), without significant perceptual distortion and provides robustness against MPEG compression (MP3).

Part 2

In "Robust high-capacity audio watermarking based on FFT amplitude modification" [80] a high-capacity watermarking algorithm for digital audio, which is robust against common audio signal processing, is proposed. This scheme provides further robustness against attacks and additional security measures compared to [78]. The main properties of this scheme are summarised below:

1. The frequency band and the scaling parameter s are chosen using an automatic process.
2. The frequency band for embedding the secret bits is chosen based on the difference between the original audio and an MP3-compressed version of the audio signal. The bits are embedded into a frequency interval for which this difference is below some specified threshold.

3. To increase security, the tuning parameters are embedded into audio frequency samples using a Pseudo-Random Number Generator (PRNG) to scramble both the values and the position of the parameters. The seed of the PRNG is required as a secret key both at the sender and the receiver.
4. A time-domain synchronisation technique is used to divide the audio signal into blocks of a given length for large files.

This method provides excellent transparency (ODG from –1 to 0), very high capacity (from 1.5 to 8.5 kbps) and robustness against different signal processing attacks.

Part 3

In "High capacity audio watermarking using FFT amplitude interpolation" [79], the aim of the proposed method is to develop a high bit-rate audio watermarking technique with robustness against common attacks and good transparency. This algorithm is based on the difference between the original and the interpolated amplitudes of the FFT samples as obtained using the spline interpolation. The ratio between the interpolated error to the interpolated sample and the selected frequency band are the two parameters of this method, which can be selected adaptively to regulate the capacity, the perceptual distortion and the robustness of the scheme. Furthermore, the suggested scheme is blind. The experimental results show that this scheme has quite a high capacity (about 3 kbps) without significant perceptual distortion (ODG about –0.5) and provides robustness against common signal processing attacks such as echo, noise, filtering, re-sampling, and MPEG compression (MP3). Besides, the CPU time required by the scheme is short enough (about 20% of the playing time) to use it in real-time applications.

Part 4

The letter "DWT–based high capacity audio watermarking" [81] suggests a novel high capacity robust audio watermarking algorithm by using the high frequency band of the wavelet decomposition, for which the human auditory system (HAS) is not very sensitive to alterations. The main idea is to divide the high frequency band into frames and then, for embedding, the wavelet samples are changed depending on the average of the corresponding frame. The experimental results show that the method has very high capacity (about 5.5 kbps), without significant perceptual distortion (ODG in [–1 ,0] and

SNR about 33 dB) and provides robustness against common audio signal processing such as add noise, filtering, echo and MPEG compression (MP3).

Table 1 and Table 2 compare capacity (bps), transparency (ODG and SNR), extracting time and robustness against attacks (BER) of the proposed schemes and a recent paper published in IEEE Trans. Multimedia. Furthermore, the comparison part of theses papers [79-81] show that the schemes overcome the capacity of the existing audio watermarking schemes, whilst keeping transparency in the high quality area (ODG in [–1, 0]) and robustness against the attacks described in the audio watermarking literature.

Table1. Comparison between the proposed schemes and a recent scheme published in IEEE Trans. Multimedia [94]

Algorithm	***Audio File***	***SNR of marked (dB)***	***ODG of marked***	***Payload (bps)***	***Extracting time/ duration of file***
Real time scheme [78]	SQAM files [95]	–	–2 to 0	2.5 k to 8.5 k	0.1 to 0.2
FFT scheme [80]	Songs [96]	35 to 44	–1 to 0	1.5 k to 8.5 k	0.1 to 0.2
Interpolation scheme [79]	Songs [96]	30.5	–1 to 0	3 k	0.1 to 0.3
Wavelet scheme [81]	Song s[96]	33	–1 to 0	5.5 k	0.1 to 0.3
IEEE Trans. Multimedia [94]	Classical music	25	–	176	–

Table2. Comparison of robustness against different attacks between the proposed schemes and a recent scheme published in IEEE Trans. Multimedia

Attack name	parameters	ODG of Attacked file	BER %				
			Real time scheme [78]	FFT scheme [80]	Interpol-ation scheme [79]	Wavelet scheme [81]	IEEE Trans. Multim-edia [94]
AddBrumm	1 – 6k, 1 – 7k	–2 to –4	–	0–1	0 – 1	0 –1	–
AddDynNoise	1 – 2	–2 to –3	–	0–1	0 – 8	2 – 7	–
ADDFFTNoise	2048,400	–1 to –2	–	0	1 – 2	0 – 2	–
Addnoise	1 – 20	0 to –1	–	0	0 – 1	0 – 4	0
AddSinus	1 – 5k , 1 – 7k	–3 to –4	–	0	0	0	–
Amplify	80 – 145	0 to –1	–	0	0	0 – 1	–
BassBoost	1 – 50,1 – 50	–3 to –4	–	–	0	0 – 2	–
Echo	1 – 10	–2 to –3	–	–	0 – 1	0 – 3	6
FFT_HLPassQuick	2048,1–10k,18–22k	–3 to –4	–	–	1 – 4	0 – 2	–
FFT_Invert	1024	–3 to –4	–	0	1 – 2	0	–
FFT_RealReverse	2 , 2048	–3 to –4	–	0–3	–	11 – 24	–
FFT_Stat1	2 , 2048	–3 to –4	–	0-36	–	14 – 23	–
invert	-	–3 to –4	–	0	0	0	–
Resampling	44/22/44	–1 to –2	–	–	5	38 – 47	0
LSBZero	-	–0.1 to 0.0	–	–	0	0	0
MP3	128	0 to –1	0 – 3	0	0 – 5	0 – 3	–
Noise_Max	1–2,1– 14k,1 – 500	0 to –1	–	–	0 – 1	0 – 1	–
Pitchscale	1.1	–3 to –4	–	–	0 – 1	32 – 61	–
RC_HighPass	Passing area	–3 to –4	–	0	0 – 1	0 – 1	–
RC_LowPass	Passing area	0 to –1	–	0	0	0 – 4	3
Smoth	–	–3 to –4	–		–	15 – 31	–
Stat1	–	–1 to –2	–		–	21 – 44	–
TimeStretch	1.05	–3 to – 4	–		–	44 – 65	–
quantization	16 – 12	0 to –1	–		–	2 – 4	0

3.2 Possible directions for future research

3.2.1 Image data hiding

To achieve a narrower histogram is the main goal of the histogram shifting approach. Note that, with an ideal narrow histogram, capacity will be about 1 bpp with excellent transparency (PSNR = 48.2dB), which is perfect for all applications. Furthermore, the suggested approach is reversible. Thus, the most convenient way to improve the results published in this thesis is to design a transform which makes it possible to achieve even a narrower histogram which will lead to design a better scheme for image data hiding as capacity is concerned.

3.2.2 Audio watermarking

Among the existing transforms, the wavelet transform has several advantages in audio signal processing. Its inherent frequency multi-resolution and logarithmic decomposition of the frequency bands resemble the human perception of frequencies, since it provides the decomposition to mimic the critical band structure of the HAS. The wavelet transform is suitable for robust applications and makes it possible to design schemes by using different wavelet frequency bands. The results already obtained in this thesis forecast the possibility of obtaining even better watermarking schemes, as the properties of capacity, transparency and robustness are considered, using the discrete wavelet transform.

Another direction in the audio field is the implementation of a prototype of a complete application, such as a broadcast monitoring system, which can integrate an efficient audio watermarking system implemented in real time, a synchronisation method and digital broadcasting transmitter and receiver devices to demonstrate the applicability of the suggested schemes in a real scenario.

REFERENCES

1. H. Yu, D. Kundur, and C.-Y. Lin, “Spies, thieves, and lies: The battle for multimedia in the digital era,” *IEEE Multimedia*, vol. 8, no. 3, pp. 8–12, Jul.-Sep. 2001.
2. I. Cox , M Miller & J. Bloom “Digital Watermarking”. *Morgan Kaufmann Publishers*, San Franciso, CA, 2003.
3. M. Wu & B. Liu “Multimedia Data Hiding”. *Springer Verlag*, New York, NY, 2003.
4. D. Kundur “Watermarking with diversity: Insights and implications”. *IEEE Multimedia* 8(4): p 46–52, 2001.
5. J. A. Bloom, I. J. Cox, T. Kalker, J.-P. Linnartz, M. L. Miller, and B. Traw, “Copy protection for DVD video,” *Proceedings of the IEEE*, this issue, vol. 87, no. 7, pp. 1267-1276, July 1999.
6. J. J. Eggers and B. Girod, “Informed Watermarking”. Boston, MA: Kluwer, 2002.
7. W. Bender, D. Gruhl, and N. Morimoto, “Techniques for data hiding,” *Proc. SPIE*, vol. 2420, pp. 40-48, Feb. 1995.
8. J. Cox and M. L. Miller, “The first 50 years of electronic watermarking,” *EURASIP J. Appl. Signal Process.*, vol. 2002, no. 2, pp. 126-132, Feb. 2002.
9. F. Hartung and M. Kutter, “Multimedia watermarking techniques,” *Proc. IEEE*, vol. 87, pp. 1079-1107, July 1999.
10. M. Swanson, B. Zhu, and A. Tewfik, “Current state of the art, challenges and future directions for audio watermarking,” in *Proc. IEEE ICMCS* Florence, Italy, vol. 1, pp. 19-24,1999.
11. M. Arnold, “Audio watermarking: Features, applications and algorithms,” in Proc. *IEEE Int. Conf. Multimedia and Expo*. ,vol. 2, pp. 1013–1016, New York, 2000.
12. F. C. Mintzer, L. E. Boyle, et al., “Toward Online, Worldwide Access to Vatican Library Materials,” *IBM Journal of Research and Development*, Vol. 40, No. 2, , pp 146-147, March 1986.
13. J. Johnston, “Estimation of perceptual entropy using noise masking criteria”. In: Proc. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, New York, NY, p 2524–2527, 1998.
14. U. Rajendra Acharya, Anand Deepthi, P. Subbanna Bhat and U.C. Niranjan, “Compact storage of medical image with patient information”, *IEEE Trans. Inform. Technol. Biomed.* (4), pp. 320–323, 2001.
15. G. Coatrieux, B. Sankur, and H. Maˆıtre, “Strict integrity control of biomedical images,” in *Security and Watermarking of Multimedia Contents III*, vol. 4314 of *SPIE Proceedings*, San Jose, Calif, USA, January 2001.
16. A. Wakatani, .: "Digital watermarking for ROI medical images by using compressed signature image". *In Proc. 35th Hawaii International Conference on System Sciences*, pp.2043-2048, 2002.

17. F. Y. Shih, Y-.Ta Wu, "Robust watermarking and compression for medical images based on genetic algorithms," Journal of Information Sciences, Elsevier, vol. 175, n°3, pp.200-216, 2005.

18. N. F. Johnson, Z. Duric, and S. Jajodia. “Information Hiding: Steganography and Watermarking – Attacks and Countermeasures”. *Kluwer Academic Publishers*, 2000.

19. A. Lang, J. Dittmann, R. Spring and C. Vielhauer, Audio watermark attacks: from single to profile attacks, In: City University of New York (Veranst.): Multimedia and Security, MM & Sec’05 (Workshop New York, NY, USA August 1-2 2005), New York, NY, ACM, pp. 39–50, ISBN 1-59593-032-9, 2005

20. M. D. Swanson, B. Zhu, A. H. Tewfik, and L. Boney, “Robust audio watermarking using perceptual masking,” *Elsevier Signal Processing, Special Issue on Copyright Protection and Access Control*, vol. 66, no. 3, pp. 337-355, 1998.

21. S. Craver, J. Stern, Lessons learned from SDMI, in: *Proceedings of the IEEE Multimedia Signal Processing MMSP’01 workshop*, IEEE, Cannes, France, 2001.

22. D. Pan, “A tutorial on MPEG/audio compression,” *IEEE Mult. Med*., pp. 60–74, Summer 1995.

23. M. Wu, W. Trappe, Z. J. Wang, and K. J. R. Liu, “Review paper: Collusion resistant fingerprinting for multimedia,” *IEEE Signal Process. Mag.*, vol. 21, no. 2, pp. 15–27, Feb. 2004.

24. W. Trappe, M. Wu, Z.J. Wang, and K.J.R. Liu, “Anti-collusion fingerprinting for multimedia,” *IEEE Trans. Signal Processing*, vol. 51, pp.1069–1087, Apr. 2003.

25. Wu et al., 2003 Wu, C., Zhou, J., Bian, Z., Rong, G., 2003. “Robust Crease Detection in Fingerprint Images”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Madison, Wisconsin, June Vol. II, pp. 505–512.

26. H. Zhao, M. Wu, Z.J. Wang, and K.J.R. Liu, “Nonlinear collusion attacks on independent fingerprints for multimedia,” in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP’03)*, pp. 664–667, Hong Kong, Apr. 2003.

27. Z.J. Wang, M. Wu, H. Zhao, W. Trappe, and K.J.R. Liu, “Resistance of orthogonal Gaussian fingerprints to collusion attacks,” in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP’03)*, pp. 724–727, Hong Kong, Apr. 2003.

28. D. Kirovski, H. Malvar, and Y. Yacobi, “A dual watermarking and fingerprinting system,” Microsoft Res. Lab., Redmond, WA, Tech. Rep. MSR-TR-2001-57, 2001.

29. D. Boneh, and J. Shaw, “Collusion-secure fingerprinting for digital data”. *IEEE Transactions on Information Theory* 44(9): p 1897–1905, 1995.

30. Y. Yacobi, “Improved boneh-shaw content fingerprinting”. *In: Proc. Cryptographer’s Track at RSA Conference*, San Francisco, CA, p 378–391,2001.

31. Termont P, De Stycker L, Vandewege J, Op de Beeck M, Haitsma J, Kalker T, Maes M & Depovere G, “How to achieve robustness against scaling in a real-time digital watermarking system for broadcast monitoring”. In: Proc. *IEEE International Conference on Image Processing*, Vancouver, BC, p 407–410, 2000.

32. P. Termont, L. De Strycker, J. Vandewege, J. Haitsma, T. Kalker, M. Maes, G. Depovere, A. Langell, C Alm, and P. Norman, “Performance measurements of a real-time digital watermarking system for broadcast monitoring”. *In: Proc. IEEE International Conference on Multimedia Computing and Systems,* Florence, Italy, p 220–224,1999.

33. G. Depovere, T. Kalker, J. Haitsma, M. Maes, L. De Strycker, P. Termont, J. Vandewege, A. Langell, C. Alm, P. Normann, G. O'Reilly, B. Howes, H. Vaanholt, R. Hintzen, P. Donnely, A. Hudson, The VIVA project: digital watermarking for broadcast monitoring, in: *Proceedings of the IEEE International Conference on Image Processing*, Vol. 2, pp. 202–205, 1999.

34. T. Kalker, and J. Haitsma, “Efficient detection of a spatial spread-spectrum watermark in mpeg video streams”. In: *Proc. IEEE International Conference on Image Processing*, Vancouver, BC, p 407–410, 2000

35. E. Zwicker, and h. Fastl, “Psychoacoustics”. *Springer Verlag*, Berlin, Germany, 1999.

36. P. Noll, “Wideband speech and audio coding”. *IEEE Communications Magazine* 31(11): p 34–44,1993.

37. Fridrich, M. Goljan, and R. Du, “Distortion-free data embedding”. *Lecture Notes in Computer Science* 2173: p 27–41,2002.

38. Y. Lee, & L. Chen, “High capacity image steganographic model”. *IEE Proceedings Vision Image Signal Processing* 147(3): p 288–294, 2000.

39. J. Fridrich, M. Goljan, and R. Du, “ Lossless data embedding - new paradigm in digital watermarking”. *Applied Signal Processing*: p 185–196, 2002.

40. T. Cedric, R. Adi, I. Mcloughlin "Data concealment in audio using a nonlinear frequency distribution of PRBS coded data and frequency-domain LSB insertion", *Proc. IEEE Region 10 International Conference on Electrical and Electronic Technology*, Kuala Lumpur, Malaysia, pp. 275-278, September 2000.

41. R. J. Ruiz and J. R. Deller. “Digital watermarking of speech signals for the national gallery of the spoken word”. In ICASSP, Turkey, 2000

42. R. Lancini, Mapelli F., and S. “Tubaro. Embedding indexing information in audio signal using watermarking technique”. In *IEEE Region and international symposium on video/image processing and multimedia communications (VIPromCom),* pages 257{261, Zadar, Croatia, June 2002.

43. B. Ko, R. Nishimura and Y. Suzuki, “Time-spread echo method for digital audio watermarking using pn sequences”. In: *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, FL, p 2001–2004,2002.

44. S. Foo S, T. Yeo and D. Huang, “An adaptive audio watermarking system” In: Proc. *IEEE Region 10 International Conference on Electrical and Electronic Technology*, Phuket Island- Langkawi Island, Singapore, p 509–513,2001.

45. P. Bassia, I. Pitas and N. Nikolaidis, “Robust audio watermarking in the time domain”. *IEEE Transactions on Multimedia* 3(2): p 232–241, 2001.

46. I. Cox, J. Kilian, F. Leighton and T. Shamoon, “Secure spread spectrum watermarking for multimedia”. *IEEE Transactions on Image Processing* 6(12): p 1673–1687, 1997.

47. D. Kirovski and H. Malvar, “Spread-spectrum watermarking of audio signals”. *IEEE Transactions on Signal Processing* 51(4): p 1020–1033, 2003.

48. M. Swanson, B. Zhu, A.Tewfik and L. Boney, “Robust audio watermarking using perceptual masking” *IEEE Transactions on Signal Processing* 66(3): p 337–355, 1998.

49. S. Saito , T. Furukawa & K. Konishi, ” A digital watermarking for audio data using band division based on qmf bank”. In: Proc. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, FL, p 3473–3476, 2002.

50. S. Lee & Y. Ho, ”Digital audio watermarking in the spectrum domain”. *Electronics Letters* 46(3): p 744–750, 2000.

51. S. Cheng, H. Yu & Z. Xiong, ”Enhanced spread spectrum watermarking of mpeg-2 aac”. In: Proc. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, FL, p 3728–3731, 2002

52. David Megías, Jordi Herrera-Joancomartí, Julià Minguillón: “Robust Frequency Domain Audio Watermarking: A Tuning Analysis”. *IWDW*: 244-258, *2004.*

53. David Megías, Jordi Herrera-Joancomartí, Julià Minguillón: “A Robust Frequency Domain Audio Watermarking Scheme for Monophonic and Stereophonic PCM Formats”. *EUROMICRO* 2004: 449-452.

54. David Megías, Jordi Herrera-Joancomartí, Julià Minguillón: “Total Disclosure of the Embedding and Detection Algorithms for a Secure Digital Watermarking Scheme for Audio”. *ICICS* 2005: 427-440.

55. David Megías, Jordi Herrera-Joancomartí, Julià Minguillón: “An audio watermarking scheme robust against stereo attacks”. *MM&Sec* 2004: 206-213.

56. C. Xu, J. Wu & Q. Sun “A robust digital audio watermarking technique”. In: Proc. International Symposium on Signal Processing and its Applications, Brisbane, Australia, p 95–98, 1999.

57. Xu C, Wu J & Sun Q “Digital audio watermarking and its application in a multimedia database”. In: Proc. *International Symposium on Signal Processing and its Applications*, Brisbane, Australia, p 91–94, 1999.

58. I.J Cox, J. Kilian, F.T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia" in *IEEE Transactions on Image Processing*, vol. 6, no. 12, Dec.1997, pp:1673 -1687

59. D, Kundur., D. Hatzinakos, "Digital Watermarking using Multiresolution Wavelet Decomposition", Proc. *IEEE Int. Conf. On Acoustics, Speech and Signal Processing*, Seattle, Washington, vol. 5, pp. 2969-2972, May 1998.

60. S. Pereira, JJK ORuanaidh, F. Deguillaume, G. Csurka, & T. Pun, “Template Based Recovery of Fourier-Based Watermarks using Log-polar and Log-log Maps”, *in Proc. IEEE Int. Conf. Multimedia Computing and Systems*, vol. 1, pp. 870--874. Florence, Italy, 1999.

61. David Megías, Jordi Herrera-Joancomartí, Julià Minguillón: “A Robust Audio Watermarking Scheme Based on MPEG 1 Layer 3 Compression”. *Communications and Multimedia Security*: 226-238, 2003.

62. J. Tian, “Reversible data embedding using a difference expansion,” *IEEE Trans. Circuits Syst. Video Technol*., vol.13, no.8, pp.890–896, Aug. 2003.

63. Ni Zhicheng, Y.Q. Shi, N.Ansari, and W. Su, “Reversible data hiding,” *IEEE Trans. on Circuits and Systems for Video technology*, 16(3):354–362, March 2006.

64. M.Fallahpour, M.H. Sedaaghi, "High capacity lossless data hiding based on histogram modification," *IEICE Transactions on Electronics Express* Vol. 4, No. 7 pp.205-210, April 2007. (impact=0.48)

65. C.C Lin, N.L Hsueh, “Hiding Data Reversibly in an Image via Increasing Differences between Two Neighboring Pixels,” *IEICE TRANS. INF. & SYST.*, Vol.E90–D, NO.12 pp 2053-2059, December 2007.

66. M.Fallahpour, M.H. Sedaaghi, "Lossless adaptive data hiding” *WSEAS Trans. on Signal Processing*, 2(6): pp. 909–913, June 2006.

67. M.Fallahpour, M.H. Sedaaghi, "Lossless data hiding for military images" Accepted *in WSEAS Trans. on Circuit and System*, 2006.

68. **M.Fallahpour;** D. MEGÍAS. "Reversible Data Hiding Based On H.264/AVC Intra Prediction. ". *International Workshop on Digital Watermarking (IWDW 2008). Lecture Notes in Computer Science*, 5450, pp. 52–60, 2009.

69. **M. Fallahpour**, D. Megias, M. Ghanbari, “High capacity, reversible data hiding in medical images” *IEEE International Conference on Image Processing , ICIP*2009, in press.

70. **M. Fallahpour**; D. Megias, M. Ghanbari, “High capacity, reversible data hiding in medical images” Submitted revised version to *IET Image Processing*, August 2009.

71. R Geiger, Y Yokotani, G Schuller – Audio data hiding with high data rates based on INTMDCT *Acoustics, Speech and Signal Processing, 2006*. ICASSP 2006

72. S. Esmaili, S. Krishnan, K. Raahemifar, “A novel spread spectrum audio watermarking scheme based on time - frequency characteristics”. IEEE Conf. Electrical and Computer Engineering, Vol. 3, pp. 1963-1966, May 2003.

73. S.J.Lee, S.H.Jung, “A survey of Watermarking Techniques Applied to Multimedia,” in Proc. IEEE. Int. Symposium on Industrial Electronics, vol. I, pp. 272-277, June 2001.

74. N. Sriyingyong, K. Attakitmongcol, “Wavelet-Based Audio Watermarking Using Adaptive Tabu Search,” IEEE Int. symp. Wireless Pervasive Computing, pp. 1-5, Jan. 2006.

75. S. Wu, J. Huang, D. Huang, Y. Q. Shi, “Efficiently Self-Synchronized Audio Watermarking for Assured Audio Data Transmission,” *IEEE Trans. Broadcasting*, Vol. 51, No. 1, pp. 69-76, Mar. 2005.

76. W. Li, X. Xue, P. Lu, “Localized Audio Watermarking Technique Robust Against Time-Scale Modification,” *IEEE Trans. Multimedia*, Vol. 8, No. 1, pp. 60-69, Feb. 2006.

77. J. Dittmann, D. Megías, A. Lang, J. Herrera-Joancomartí: "Theoretical Framework for a Practical Evaluation and Comparison of Audio Watermarking Schemes in the Triangle of Robustness, Transparency and Capacity". *Transaction on Data Hiding and Multimedia Security*: 1-40 (2006)

78. **M. Fallahpour,** D. Megías, "High capacity method for real-time audio data hiding using the FFT transform" *Advances in Information Security and Its Application* Third International Conference, ISA 2009, Springer, Seoul, Korea, June 25-27, 2009.

79. **M. Fallahpour** and D.Megías, "High capacity audio watermarking using FFT amplitude interpolation", *IEICE Electron. Express*, Vol. 6, No. 14, pp.1057-1063, (2009). (Impact factor =0.48).

80. **M. Fallahpour,** D. Megías, "Robust high-capacity audio watermarking based on FFT amplitude modification" *IEICE Transactions on Information and Systems*, Vol.E93-D, No.01, pp.-, Jan. 2010, in press (Impact factor 0.36 in 2008).

81. **M. Fallahpour,** D. Megías, "DWT–based high capacity audio watermarking" *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol.E93-A, No.01, pp.-, Jan. 2010, in press (Impact factor 0.43 in 2008).

82. R. Fujimoto, M. Iwaki, T. Kiryu, "A Method of High Bit-Rate Data Hiding in Music Using Spline Interpolation". *IIH-MSP* , 11-14, 2006

83. C J Weinstein, "Programs for Digital Signal Processing", IEEE Press, 1979.

84. C. Podilchuk and W. Zeng, "Image-adaptive watermarking using visual models," *IEEE J. Select. Areas Commun*., vol. 16, pp. 525-539, May 1998.

85. N. Terzija, W. Geisselhardt, "Robust Digital Image Watermarking Method Based On Discrete Fourier Transform", InProceedings of the *5th IASTED International Conference on Signal and Image Processing*, Honolulu, Hawaii, August 2003.

86. B.J. Falkowski, L. Lim, "Image watermarking using Hadamard transform", *Electron. Lett.* 36 (3), pp. 211–213, 2000.

87. http://www.cs.wm.edu/~srgian/tier-conf-final2007.html

88. A. Deshpande, K. M. M. Prabhu, "A substitution-by-interpolation algorithm for watermarking audio". *Signal Processing, Elsevier* 89(2): 218-225, 2009

89. **M. Fallahpour**; D. Megías, Mohammad Ghanbari, "Subjectively adapted high capacity lossless image data hiding based on prediction errors", submitted in *Multimedia Tools and Applications, Springer*, August 2009 (Impact factor 0.46).

90. **M. Fallahpour,** D. Megías, "High capacity audio watermarking using the high frequency band of the wavelet domain" submitted in *Multimedia Tools and Applications, Springer*, August 2009, (Impact factor 0.46).

91. D. Megías, J. Serra, **M. Fallahpour,** "A novel blind audio watermarking system with enhanced transparency and self-synchronisation" submitted in *Signal processing*, July 2009, (Impact factor 1.2).

92. **M. Fallahpour**, "Reversible image data hiding based on gradient adjusted prediction", *IEICE Electron. Express*, Vol. 5, No. 20, pp.870-876, (2008). (Impact factor = 0.48 in 2008).

93. I. Cox, M. Miller, J. Bloom, J. Fridrich, T. Kalker, Digital Watermarking and Steganography, Morgan Kauffman, 2007, ISBN: 978-0-12-372585-1.

94. M. A. Akhaee, M. J. Saberian, S. Feizi, F. Marvasti, "Robust Audio Data Hiding Using Correlated Quantization With Histogram-Based Detector" IEEE TRANS. ON Multimedia,V11, P 1-9, 2009.

95. SQAM Sound Quality Assessment Material, http://sound.media.mit.edu/mpeg4/audio/sqam/

96. No, Really, "Rust". http://www.jamendo.com/en/album/7365.

APPENDIX A

PROGRAMME COMMITTEES AND ACCEPTANCE LETTERS

A.1 Technical Committee of IWDW 2008

Technical Committee of 7th International Workshop on Digital Watermarking

M. Barni (Univ. of Siena, Italy)
J. Bloom (Thomson, USA)
C. C. Chang (Feng-Chia U., Taiwan)
J. Dittmann (U. Magdeburg, Germany)
J.-Luc Dugelay (Eurecom, France)
M. Goljan (SUNY Binghamton, USA)
B.Jeon (SKKU, Korea)
T. Kalker (HP, USA)
M. Kankanhalli (NUS, Singapore)
A. Ker (Oxford University, UK)
Alex Kot (NTU, Singapore)
C. C. Jay Kuo (USC, USA)
I. Lagendijk (Delft U. tech., Netherland)
H.-Kyu Lee (KAIST, Korea)
C. T. Li (U. Warwick, UK)
S. Lian (France Telecom R&D, China)
Z. Lu (Sun Yat-sen Univ., China)
B. Macq (UCL, Belgium)
K. Martin (Royal Holloway U. London, UK)
N. Memon (Polytechnic U., USA)
K. Mihcak (Bogazici U., Turkey)
M. Miller (NEC, USA)
Z. Ni (WorldGate Communications, USA)
J. Ni (Sun Yat-sen Univ., China)
H. Noda (Kyushu Inst. of Tech., Japan)
J.-S. Pan (NKUAS, Taiwan)
F. Perez-Gonzalez (U. Vigo, Spain)
I. Pitas (U. Thessaloniki, Greece)
A. Piva (U. Florence, Italy)
Y.-M. Ro (ICU, Korea)
A.-R. Sadeghi (U. Bochum, Germany)
H. G. Schaathun (U. Surrey, UK)
K. Sakurai (Kyushu U., Japan)
Q. Sun (Inst. Infocomm Research, Singapore)
H. Treharne (U. Surrey, UK)
S. Voloshynovskiy (U. Geneva, Switzerland)
S. Wang (U. Shanghai, China)
M. Wu (U. Maryland, USA)
S. Xiang (Sun Yat-sen U., China)
G. Xuan (Tongji U., China)
H. Zhang (Beijing U. Tech., China)
D. Zou (Thomson, USA)

A.2 Technical Committee of ISA 2009

Program Committee The 3rd International Conference on Information Security and Assurance

Alessandro Piva (University of Florence, Italy)
Binod Vaidya (Chosun Univerrsity, Korea)
Bo Zhu (Concordia University, Canada)
Boniface Hicks (Penn State University, USA)
Byoungcheon Lee (Joongbu University, Korea)
Chin-Chen Chang (Feng Chia University, Taiwan)
Chunming Rong (University of Bergen, Norway)
Claudio Ardagna (University of Milan, Italy)
Dawu Gu (Shanghai Jiao Tong University, China)
Dharma P. Agrawal (University of Cincinnati, USA)
Dieter Gollmann (Hamburg University of Technology, Germany)
Dorothy Denning (Georgetown University, USA)
Duncan S. Wong (City University of Hong Kong, China)
Edward Jung (Rutgers University, USA)
Francesca Saglietti (University of Erlangen-Nuremberg, Germany)
Gail-Joon Ahn (UNC Charlotte USA)
George Ghinea (Brunel University, UK)
Golden G. Richard III (University of New Orleans, USA)
Guojun Wang (Central South University, China)
Hee-Jung Lee (Kangnam University, Korea)
Isaac Agudo (University of Malaga, Spain)
Ioannis G. Askoxylakis (FORTH-ICS, Greece)
Jaechul Sung (University of Seoul, Korea)
Jan deMeer (SmartSpaceLab, Germany)
Jeng-Shyang Pan (National Kaohsiung University, Taiwan)
Jie Li (University of Tsukuba, Japan)
Jongsung Kim (Korea Universtiy, Korea)
Jianying Zhou (Institute for Infocomm Research, Singapore)
Julio Cesar Hernandez-Castro (Carlos III University of Madrid, Spain)
Jung-Taek Seo (NSRI à The Attached Institute of ETRI, Korea)
Kevin Butler (Penn State University, USA)
Konstantinos Markantonakis (Royal Holloway University of London, UK)
Kouichi Sakurai (Kyushu University, Japan)
Kui Ren (Illinois Institute of Technology, USA)
Lei Hu (The Chinese Academy of Sciences, China)
Liwen He (BT, UK)
Martin Loeb (University of Maryland, USA)
Michael Tunstall (University of Bristol, UK)
Michael W. Sobolewski (Texas Tech University, USA)
Min-Shiang Hwang (National Chung Hsing University, USA)
Nancy Mead (Carnegie Mellon University, USA)
Ning Zhang (University of Manchester, UK)
Pierre-François Bonnefoi (University of Limoges, France)
Pierre Dusart (University of Limoges, France)
Raphael Phan (Loughborough University, UK)

A.3 Acceptance letter *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*

Acceptance letter of “DWT-based high capacity audio watermarking”

[Notification of Acceptance]

October 02, 2009.

Paper Number: 2009EAL2154
Paper/Letter: Letter
Issue: EA
Regular/Special: Regular
Title of Special Section:
Title: DWT-based high capacity audio watermarking
Date of Evaluation: October 02, 2009.

==========
!! ATTENTION !!

Click the following URL right now to let us know that you receive this e-mail.
(This URL is available for only one time.)
https://review.ieice.org/receiptmail_e.aspx?lid=2009EAL2154&code=0910021302041

Dr. Mehdi Fallahpour
telecommunication
Universitat Oberta de Catalunya

Dear Dr. Mehdi Fallahpour,
It is my pleasure to inform you that your above manuscript is accepted
for publication in the IEICE Transactions.

The followings are the future steps to publication.

The followings are the future steps.

1. A notification of "Information on Volume and Number" will reach you
as soon as the publication schedule is fixed.
2. The galley proof will be sent to you via Internet around 6 weeks
before the date of publication (every 1st of month).
3. Please make corrections (ONLY TYPOS), if any, according to "comments
to the author(s)" ON THE GALLEY PROOF to be sent later.
4. An invoice and reprints will be sent to you in the middle of the
publication month. You may pay for it by credit card or bank transwer
after receiving the invoice.

All titles on the IEICE Transactions are published electronically at

1) IEICE Transactions Online
<http://www.ieice.org/eng/trans_online/index.html>
2) J-STAGE <http://www.jstage.jst.go.jp/browse/transfun>
Please be advised that the copyrights for all articles in the Transactions published by the IEICE, regardless of the publication medium, are automatically transferred to the IEICE.

Congratulations, and thank you for submitting the results of your research to the IEICE Transactions.

Sincerely,

Masaki Kawamura
Associate Editor of the IEICE Transactions

Yamaguchi University

Tel: 083-933-5701
Fax: 083-933-5701
Email: kawamura@sci.yamaguchi-u.ac.jp

A.4 Acceptance letter *IEICE Transactions on Information and Systems*

Acceptance letter of "Robust high-capacity audio watermarking based on FFT amplitude modification"

[Notification of Acceptance]

October 06, 2009.

Paper Number: 2009EDP7108
Paper/Letter: Paper
Issue: ED
Regular/Special: Regular
Title of Special Section:
Title: Robust high-capacity audio watermarking based on FFT amplitude modification
Date of Evaluation: October 05, 2009.

==========
!! ATTENTION !!

Click the following URL right now to let us know that you receive this e-mail.
(This URL is available for only one time.)
https://review.ieice.org/receiptmail_e.aspx?lid=2009EDP7108&code=0910060848121

Dr. Mehdi Fallahpour
telecommunication
Universitat Oberta de Catalunya

Dear Dr. Mehdi Fallahpour,
It is my pleasure to inform you that your above manuscript is accepted
for publication in the IEICE Transactions.

The followings are the future steps to publication.

The followings are the future steps.

1. A notification of "Information on Volume and Number" will reach you
as soon as the publication schedule is fixed.
2. The galley proof will be sent to you via Internet around 6 weeks
before the date of publication (every 1st of month).
3. Please make corrections (ONLY TYPOS), if any, according to "comments
to the author(s)" ON THE GALLEY PROOF to be sent later.
4. An invoice and reprints will be sent to you in the middle of the
publication month. You may pay for it by credit card or bank transfwer
after receiving the invoice.

All titles on the IEICE Transactions are published electronically at
1) IEICE Transactions Online
<http://www.ieice.org/eng/trans_online/index.html>
2) J-STAGE <http://www.jstage.jst.go.jp/browse/transfun>
Please be advised that the copyrights for all articles in the Transactions published by the IEICE, regardless of the publication medium, are automatically transferred to the IEICE.

Congratulations, and thank you for submitting the results of your research to the IEICE Transactions.

Sincerely,

Katsunari Yoshioka
Associate Editor of the IEICE Transactions
Interdisciplinary Research Center
Yokohama National University
79-7 Tokiwadai,Hodogaya-ku, Yokohama, 240-8501, Japan
Tel: 045-339-3690
Fax:
Email: yoshioka@ynu.ac.jp