THE COMBINATION OF WATERMARKING ALGORITHMS FOR MP4 VIDEO

Thanh Tung Nguyen*, Chan Nam Ngo, Tran Khanh Dang

Faculty of Computer Science and Engineering, Ho Chi Minh City University of Technology, Ho Chi Minh city, Vietnam

*Email: thanhtung@cse.hcmut.edu.vn

Received: 15 July 2014, Accepted for publication: 16 October 2014

ABSTRACT

Nowadays, there has been an explosive growth in the digital technique and Internet usage. Some challenging problems have emerged such as illegal manipulation, distribution of digital video, copyright protection, digitized properties verification, authentication and content integrity. Watermarking is the process of embedding some additional data (copyright information) along with the container, which has proved to be the solution to these problems above. In this paper, we present the MP4 file structures, classify watermarking algorithms and summarize the watermarking characteristics and possible attacks. Then, we propose a solution that combines digital video, audio and MP4 structure algorithms in order to deal with the copyright protection.

Keywords: steganography, MP4 video, video watermarking, audio watermarking, spatial domain watermark, frequency domain watermark.

1. INTRODUCTION

Recently, there has been a tremendous use of Internet and multimedia technologies. They have become our daily life amenities. Meanwhile, a large amount of data is copied, modified and distributed without the owner’s permission. Hence, the owners of multimedia products are concerned about illegal copying of their products. As a result, copyright protection for multimedia data has become an important issue. Many researches and technologies have been proposed to provide methods to solve this issue. An effective approach that has been proposed is to embed the copyright information in multimedia documents is digital watermarking.

In watermark embedding, a watermark signal (text, image or audio, etc.) is constructed and then embedded into an original signal (video) to produce the watermarked signal, which uniquely identifies the ownership. After embedding the watermark, the watermarked signal can resist various attacks. The watermark detector determines whether the watermark is present or not on testing the signal at its input.
Video watermarking research received less attention than image watermarking because of its hard inheritance, however, many algorithms have already been proposed. Some algorithms embed watermark in the spatial domain such as Least Significant Bit (LSB) and Correlation Based Techniques (CBT). In these types of algorithms, the watermark is added by editing the pixel values in each frame. Hence, these methods are not robust to attacks and common signal distortions. On the contrary, frequency domain algorithms are more robust to distortions. These algorithms modify the transform coefficients of the frames of the video sequence. The most commonly used transforms are the Discrete Cosine Transform (DCT), the Discrete Fourier Transform (DFT), and especially the Discrete Wavelet Transform (DWT) because it provides both spatial and frequency domain characteristics.

Moreover, the recent trend is to combine the video watermarking with other algorithms to increase watermarking properties such as robustness and visibility. Audio watermarking is an attractive candidate because audio signals have an unpredictable nature and characteristic redundancy that make them ideal to be used as a cover of hiding secret message. LSB is one of the earliest techniques researched in the information hiding of digital audio. Some other algorithms are LSB encoding, Parity encoding, Spread Spectrum (SS) coding.

Besides many techniques that use the redundant information for embedding a secret message, another approach bases on the structure of the container (image, audio or video). Although it seems that any structure can be used to store secret message, there are not much studies in this approach [1, 2]. In this paper, we review an algorithm that bases on MP4 video structures (the periodically repeating groups of pictures and the interleaving of the video audio streams). The input of this algorithm is watermarked MP4 video that is embedded by video and audio algorithms.

The structure of this paper is as follows. In the subsequent section we present the MP4 video structures that is used to insert copyright information. The various aspects of digital watermarking are explored in Section 3 such as classification, important characteristics and possible attacks on watermark. Section 4 presents the proposed technique that combines digital video, audio and MP4 structure algorithms in order to deal with the copyright protection. Finally, section 5 concludes the paper.

2. MATERIALS AND METHODS

In order to insert the information into MP4 video files, our proposed system uses two MP4 structures: the periodically repeating groups of pictures (GOPs) and the interleaving of the video and video streams [1].

The GOP structure contains several types of frames: I, B, and P-frames. Each frame consists of the slices of macroblocks, and each macroblock is a set of four 8x8 matrices. The MP4 video stream is composed of GOPs. A GOP is a sequence of frames between two succeeding I-frames and each GOP must begin with an I-frame. This frame has the full picture information and does not need other frames in the GOP. It is intra-frame encoded. On the other hand, P-frames and B-frames are outer-frame because they do not involve the full picture information. They only have some parts of the previous or next frames. Each P-frame is generated based on the previous I-frame or P-frame. It is called forward prediction. Besides, B-frame is backward prediction. B-frames are created using the bidirectional interpolation prediction. The forward prediction is the first to determined, then the backward. The encoder
decides the ratio for calculating the average from both of these predictions. Because B-frames bring a large number of errors into the picture, they are then not used for further predictions. *Figure 1* illustrates the GOP in the video stream.

*Figure 1.* The illustration of the GOP structure.

*Figure 2.* The audio and video streams in the MP4 file structure.

The video stream does not constraint the number of used P and B-frames in a GOP. The parameter N defines the number of frames in a GOP and parameter M denotes the distance between a P and an I-frame. Moreover, an MP4 file is a multimedia container for storing various data streams. The audio and video streams are the most commonly interleaved streams in it. The individual data from one type of a stream are called chunks. They may have different sizes and different number of frames. The physical data organization in MP4 files are organized into the so called atoms: moov, mdat, stbl, stts, stss, stsc, stsz, and stco. The audio and video data are mutually interleaved in the mdat atom as *Figure 2*. 
3. THE DIGITAL WATERMARKING

Digital watermarking is a powerful technique that provides copyright protection for videos by hiding appropriate information in the original video to declare rightful ownership. A watermark is a digital code permanently embedded into the digital carrier (i.e. text, audio or video). In this section, firstly, we classify the watermarking approaches. Then we summarize the important characteristics of digital watermarking. Finally, we revise a variety of common attacks on the watermark.

3.1. The watermarking classification

Watermarking algorithms can be classified based on several criteria [2], [3]. The first classification that a person thinks about digital watermarking is based on visibility: **visible and invisible** watermarks. For watermark detection and extraction actions, we have **blind and non-blind technique**. The blind techniques do not require the original data for these actions. Based on ability of watermark to resist attack, we have **fragile and semi-fragile** watermarking. But the two main categories are based on the method of hiding watermark information in the host carrier. The two categories are **Spatial Domain Watermarking (SDW)**, and **Transform (Frequency) Domain watermarking (TDW)**. In SDW techniques, embedding and detection actions are performed on spatial pixels values using luminance, chrominance, color space or overall video frame. However, a common digital signal processing operations can damage or remove the SDW watermark. Hence, SDW techniques are not robust. In contrast, TDW techniques alter overall original values of the host carrier. These techniques are more robust and transparent compared to SDW techniques, making it very difficult to remove or destroy the watermark.

3.2. The watermarking characteristics

There are many **properties** that are commonly used to determine the quality of a watermarking technique [3, 11, 13]. To be effective, watermark should achieve the properties such as:

- **Robustness** describes if the watermark should be impossible to remove even if the detecting or embedding algorithms are public. Of course, any watermark can be removed with sufficient knowledge of particular embedding technique. In integrity protection application, watermark should not be robust against intentional modification attacks. When the watermarking detection cannot determine the embedded information, the testing data is no longer integrity. However, in copyright protection application, watermark always remains in the host carrier, even if it is subjected to intentional and unintentional attacks. Hence, based on the requirements of the application, we decide to apply watermarking algorithms with appropriate robust level.

- **Security** is to ensure that the embedded information cannot be retrieved by malicious user. According to the Kerckhoff’s principle [4], malicious user may know all about the embedding algorithm, the detector and at least one watermarked data, except the private key. Hence, security depends on the size of this key.

- **Capacity** is the amount of information that can be embedded in a watermark as well as the possibility of embedding multiple watermarks in one document.

- **Unambiguous** denotes that the retrieved watermark should uniquely identify the
Thanh Tung Nguyen, Chan Nam Ngo, Tran Khanh Dang

Copyright owner of the host carrier.

- **Transparency** is based on the properties of the human visual or auditory system. An unobtrusive watermark causes no affects, imperceptibility or quality loss. The watermark should be perceptually invisible and the quality of its container should not be destroyed by its presence.

- **Complexity** describes the effort and time we need for encoding and decoding watermark from the carrier. In the real time applications, this parameter is essential and should be fairly fast and low computational complexity.

- **Blind watermarking** describes if the extraction needs the original data or not.

- **The verification procedure** describes if the technique includes a private verification using cryptography algorithms.

The optimization of these characteristics cannot be done at the same time. We cannot require to embed a large message at the same time large robustness. We arrive at a reasonable compromise.

### 3.3. The watermarking attacks

In digital watermarking, there are various classifications of watermarking attacks [2]. This section presents some possible attacks on watermarks.

- **Common attacks** are very simple actions such as cropping, compression, rotation and zooming. However, these attacks has been successful with image and video. Moreover, cropping and compression also affect watermarked audio.

- **Brute force attacks** try to damage all aspects of the embedded watermark that maybe use to contain secret information. These type of attacks do not make any effort to identify or remove the watermark. For video example, an attack attempts to change whole pictures in the video stream, signal in the audio stream as well as the video file structure. A brute force attack can include some common attacks.

- **Confused detection attacks** attempt to insert some fake watermarks into a watermarked object. The detector confuses which are the first, authoritative watermarks. A certification authority can timestamp the secret information in order to prove his ownership.

- **Disabled detection attacks** applies some distortive transformation over the host carrier in order to degrade its watermarks. An effective distortive attack makes all watermarks of a watermarked object become undetectable. However, their watermarks still remain in the host carrier. Improving watermark detectors, we can identify them.

- **Removal attacks** try to analyze the testing data in order to detect the presence, location of the watermark and then remove it from its container. A removal attack is successful where all watermarks are removed and unwatermarked object still remains enough original data to be of value.

### 4. DESIGN OF THE PROPOSED SYSTEM

This section explains a new approach (*Figure 3*) that combines MP4 structure and video/audio algorithms. Firstly, the secret information (text, image) is serialized to binary stream. In order to increase security, we encrypt this secret binary stream. The output of the
The combination of watermarking algorithms for MP4 video

Cryptography process is the watermark stream. Secondly, the original MP4 video file is exploited by the Structure Pre-process that we will detail in next section. The output of this process is the modified MP4 video that is used in watermarking processes as well as the watermarking extraction (non-blind). Feature extraction process receives the embedded video using structure algorithms in order to extract video/audio features. Finally, the video/audio watermarking processes use these features and the watermark stream to create the watermarked MP4 video.

Next, we review some algorithms which are used in the structure watermarking process, video watermarking process and audio watermarking process.

![Figure 3. Inserts secret information into MP4 video.](image)

4.1. The MP4 structure algorithms

Based on the MP4 structures, Jókay [1] proposed two watermarking techniques. The number or the order of P and B-frames in the GOP structure are not specified. It does not fix the number of frames in the individual chunks of the video and audio streams. Hence, we can modify these number as well as their order in the GOP structure. The disadvantage of Jókay [1] technique is that it is impossible to embed secret information into a fixed GOP structure video or static frame rate without making any suspicion. In order to overcome this disadvantage, firstly, we pre-process the original MP4 file structure to determine whether it is possible to make its structured watermark become doubtful. If possible, structure pre-process modifies the original file. The output file of this process may be used in another embedding process or to detect the watermark.

In this technique, we insert a single bit of secret information using a change of the number of P/B frames in a GOP in order to avoid the picture degradation. A change is simply done by inserting or deleting a P/B frame at the end of a GOP. We would do the same for the chunks structure by alter the number of frames in one chunk. If the number of frames in some chunk of
the video/audio streams is changed, it is necessary to upgrade the relevant related control information in the file.

The process of embedding the information is the same for both file structures. Each GOP or chunk we insert only one bit. We call L is the GOP’s length or the chunk we want to modify, call V is the value of secret bit. Hence, the value of 0 bit is zero, 1 bit is one. If the sum of L and V is even, no change is done. Otherwise, the last frame in the GOP is removed and we double the first frame following the next I-frame to keep the synchronization. This doubling action may cause an error in the video. This error can be fixed by inserting an empty B/P-frame. This frame means that it does not modify its previous frame in the GOP. In the decoding process, we call L’ is the length of the GOP or chunk of the watermarked video. If L’ is even, the secret bit is a 0 bit, otherwise it is a 1 bit.

4.2. The audio watermarking algorithms

Audio watermarking is a potential candidate because audio signals have an unpredictable nature and characteristic redundancy that make them ideal to be used as a cover for hiding secret message. In this section, we review the two common audio algorithms [5], [12] below:

- **Least Significant Bit (LSB)** encoding is one of the earliest researches and is the simplest way to embed information in a digital audio file and other digital types. The LSB of the target audio stream is replaced with the binary of secret information. The capacity of this technique is large. It means that we can insert a large amount of information that does not affect the audible sounds. However, LSB is not robust, it cannot protect the secret information from simple modifications such as compression and format conversion.

- **Parity encoding** method breaks an audio stream into separate regions that each region has a parity bit. The length of region decides number of LSBs. If the parity bit of a host region matches the secret bit, no change is done. Unless, this technique decides a LSB of this region need to be flipped such as the last bit of it. This method is based on LSB encoding. Therefore, it is not robust. However, we can choose a large region in order to have more LSB choices. Because of high customization capacity, the watermark is more unobtrusive.

4.3. The video watermarking algorithms

Many video watermarking algorithms have been proposed. They are classified according to their methods of hiding secret information in the host container. The two categories are Spatial Domain Watermarking (SDW), and Transform (Frequency) Domain watermarking (TDW) [Error! Reference source not found.].

- **Spatial domain watermarks (SDW)**

  The SDW techniques insert the watermark into the host object by modifying the spatial pixel values of this host. The main advantages of these methods are that they are simple, hence, they have very low computational cost. Hence, they are widely used in video watermarking, especially real-time applications. Besides, they absolutely focus on spatial synchronization that makes them vulnerable to de-synchronization attacks. They do not consider the temporal axis. A common digital signal processing operations can damage
The combination of watermarking algorithms for MP4 video

or remove the SDW watermark. This techniques are not robust. Therefore, a good watermark must combine spatial approaches and others such as Transform Domain watermarking.

**Least Significant Bit (LSB)** modification is the most frequently used technique. In this technique, the LSB of each pixel is used to insert a bit of secret information. In video watermarking, we must focus on the multiple frame collusion because LSB does not consider temporal axis.

**Correlation based techniques** use a pseudo-random noise generator algorithm such as White Gaussian Noise signal as the spreading signal [6]. Firstly, the secret information is convert into matrix \(s(x, y)\). Secondly, we apply a pseudo-random noise generator algorithm to produce watermark \(W(x, y)\). Then, watermark \(W(x, y)\) is embedded into the original content \(O(x, y)\) according to the equation: 

\[ O_w(x, y) = O(x, y) + kW(x, y) \]

\(O_w\) is the watermarked content and \(k\) is a gain factor. As we increase the value of \(k\), it will increase the robustness of the watermark but expense the quality of watermarked contents. To extract the watermark, we use this pseudo-random algorithm again and compute the correlation between the noise pattern and possibly watermarked video. If the correlation exceeds a certain threshold \(T\), the watermark is detected as well as a single bit is set. This method can be extended to a multiple-bit watermark by dividing the frame up into blocks.

- **Transform domain watermarks (TDW)**

The TDW techniques modify the transform coefficients of the frames of the video sequence in order to insert the watermark into the carrier. These techniques are more robust and transparent compared to SDW techniques, making it very difficult to remove the watermark.

**Discrete Cosine Transform** (DCT) is an important method for video watermarking [7]. The DCT can break an image or video frame into different frequency bands in order to make it easier to embed watermarking information into the middle frequency bands [11]. These bands are chosen because they do not affect the visual important parts and are high immunity from compression and noise attacks. These techniques transform an image from the spatial domain to the frequency domain. Moreover, the most of video compression standards are based on DCT. Video is natural frequency. Based on the temporal masking property of sound, the DCT coefficients of the host video are selected and divided into groups. Then the binary watermark are embedded by modifying each group.

**Discrete Wavelet Transform** (DWT) decomposes a video frame into sub images, 3 details and 1 approximation [8]. The details sub images are horizontal (HL), vertical (LH) and diagonal (HH) detail components. The approximation sub image is lower resolution approximation image (LL).

![DWT sub-bands](image)

*(Figure 4. DWT sub-bands in (a) level 1, (b) level 2.)*
L/H represents low/high frequency and subscript behind them displays the number of layers of transforms. We embed the watermark in the sub-bands LL and HH from level 2 of the wavelet transform of the frame. Embedding the watermarks in high frequency sub-bands make them more imperceptible, however, embedding in low frequency makes them more robust. The primary advantage of this technique is its compatibility with the Human Visual System (HVS). It allows us to embed information in the regions that the HVS is less sensitive. Therefore, watermarks in these regions increase the robustness without making any suspicion. Besides, DWT technique can combine with other transform such as Principal Component Analysis [9], [10] in order to achieve high robustness technique.

5. CONCLUDING REMARKS

In this paper, we revise two MP4 file structures as well as the digital watermarking. The classification, characteristics of digital watermarking and the various common attacks on the watermark are presented. There are many digital watermarking techniques but no one is a perfect. Each technique has its own advantages and disadvantages. In order to take advantages and reduce disadvantages, we proposed a new technique that combines existing techniques such as video, audio techniques and other techniques using MP4 video structures. For example, combining the powerful mathematical video transform Discrete Wavelet Transform (DWT), Parity Encoding technique on audio stream and MP4 GOPs structure technique. The output watermarked MP4 video has enhanced security and robustness and is very difficult to attack which makes our technique successful in the field of video watermarking.

REFERENCES


