Robust 3D Wavelet Watermarking for Video

Jie Yang[*,**], Young Gon Kim[**], Moon Ho Lee[**]

[*] School of Electronics and Information Wuhan University of Technology, China,
[**] Institute of Information & Communication ,Chonbuk National University, Korea

Abstract: This paper proposes a new approach for digital watermarking and secure copyright protection of video, the principal aim being to discourage illicit copying and distribution of copyrighted material. The method presented here is based on the three dimensional discrete wavelet transform of video scene.The watermark is a copyright information encoded in the form of a spread spectrum signal to ensure the system security and is embedded in the 3D DWT magnitude of video chunks. The performance of the presented technique is evaluated experimentally.

1. INTRODUCTION

The development of digital video is more recent than that of other Medias because of the large bandwidth required. Electronic components however continue growing more powerful, while their cost decrease rapidly. The efficient access and distribution provided by digital media have led to major concerns regarding the protection of digital intellectual property. Creators and distributors of audio, image and video are hesitant to provide access to their intellectual property given the problems associated with digital copyright enforcement. Digital watermarks have been proposed to address this issue by embedding owner or distribution information directly into the digital media. The information is embedded by making small modifications to the samples in the digital data [1][2]. When the ownership of the media is in question, the information can be extracted to characterize the owner or distributor.

Image and video coding methods using wavelet transforms have been successful in providing high rates of compression while maintaining good image quality. They have generated much interest in the scientific community as alternative methods to DCT based compression schemes, such as MPEG standards. The DCT is usually implemented on a block-by-block basis. This approach leads to "blocking artifacts" at high compression ratios. In such a situation, the edges of the data blocks become visible and the appearance of the picture frames becomes noticeably degraded.The wavelet transform, widely used for image compression can be extended naturally to video sequence. Most video compression algorithms rely on 2D based schemes employing motion compensation and estimation techniques. On the other hand, efficient 3D algorithms exist which are able to capture temporal redundancies in a more natural way for 3D wavelet/subband coding, without motion compensation. Normally, 3D WT algorithms are based on a group of frames (GOF), similar to the group of pictures (GOP) used in the MPEG standards.. Potential applications for the architecture include high definition television (HDTV) and medical data compression, such as magnetic resonance imaging (MRI), etc.

Video watermarking introduces some issues not present in image watermarking. Due to large amounts of data and inherent redundancy between frames, video signals are highly susceptible to pirate attacks, including frame averaging, frame dropping, and statistical analysis, etc. The high correlation between successive frames of a video sequence makes it possible to achieve high coding efficiency in a video coding system by reducing the temporal redundancy. The basic approach adopted here is to mark the uncompressed video sequence. In the method proposed here, in contrast to the former ones, the video is considered as a three-dimensional signal with two dimensions in space and one dimension in time. The basic idea is to extend the two dimensional robust DWT image watermarking scheme to a three-dimensional DWT video watermarking scheme. With this novel approach, the watermark is embedded into the magnitude of the 3D DWT of the video data. The ownership and copyright information are encrypted in key, which is added into the magnitude values of the three dimensional DWT domain in the form of a spread spectrum signal.

In this paper, we propose a novel 3D DWT watermarking algorithm for volume data in video which is invisible and robust. "Invisible" means that the 2D rendered image of the watermarked volume is perceptually indistinguishable from that of the original volume. "Robust" watermarking implies that the watermark is resist to most intentional or unintentional attacks. This paper is organized as follows. Section 2 overviews requirements of, and techniques used for image and video watermarking. Section 3 introduces the new concept of 3D DWT watermarking and lists some relevant properties of the 3D DWT transform. In section 4 the watermark embedding/extraction processes are detailed. Section 5 presents the 3D watermarking experimental results.

2. THE PROPOSED WAVELET BASED WATERMARKING METHOD

The method is based on the 3D wavelet transform (WT) and multiresolution representation (MRR) for video signals, which includes the watermark embedding and extraction processes. Watermarking a volume data in video is essentially the process of altering the voxel values in a manner to ensure that a viewer of its volume-rendered image does not notice any perceptual change between the original volume rendering and the watermarked volume rendering. We hide the watermark sequence into multiple temporal and spatial domains to make the watermark robust.

a. The three Dimensional Wavelet Transform

The 3D DWT may be interpreted as a decomposition of the signal at different levels of resolution via multiresolution analysis [3] and can be implemented recursively by using a pair of discrete filters via an analysis/synthesis scheme as given in Fig.1. In such an approach, a signal at a resolution 0 named a_0 is decomposed into two low resolution signals, an approximation signal a_{-1} and a detail signal d_{-1}, using a low-pass filter (*g*) and a high-pass filter (*h*) respectively, followed by a 2 to 1 decimation. At each resolution j, the decomposition could be repeated on the approximation signal a_j to get a_{j-1} and d_{j-1}. Finally the wavelet decomposition of the original signal a_0 is given by the

---

* This work partial supported by KISTEP-Chonbukdo and ETRI.

approximation signal a_k at the lowest resolution k and all the detail signals d_j.

The reconstructed signal $\hat{a}_{j+1}$ is calculated by a 1 to 2 sampling of a_j and d_j and by a filtering with the low-pass ($\hat{g}$) and the high-pass ($\hat{h}$) interpolating filters. In the Z transform domain, the previous operations result in:

$$\hat{A}_{j+1}(z) = \frac{1}{2}\{[g(z)\hat{g}(z) + h(z)\hat{h}(z)]A_{j+1}(z) + [g(-z)\hat{g}(z) + h(-z)\hat{h}(z)]A_{j+1}(-z)\} \quad (1)$$

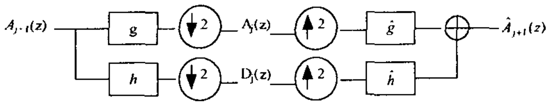

Fig.1 Wavelet Transform

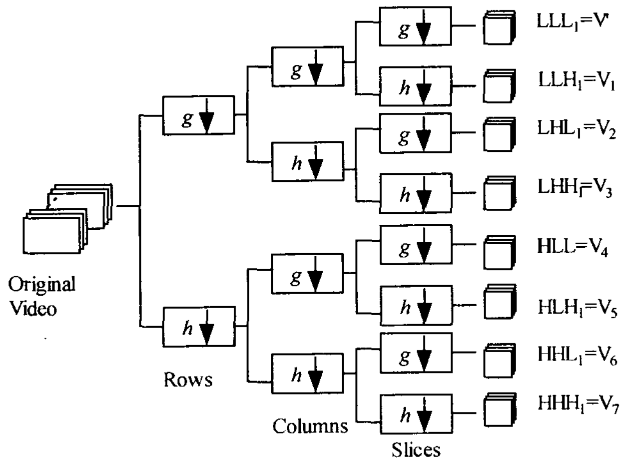

Fig.2 3D Wavelet Transform

In this equation, the first term contains the frequency response of the decomposition reconstruction system and the second one corresponds to the aliasing term. By imposing conditions on filters g, $\hat{g}$,h and $\hat{h}$, perfect reconstruction is achieved, then $\hat{A}_{j+1}(z) = A_{j+1}(z)$ and the decomposition / reconstruction is free of aliasing and free of module distortion. Many solutions have been proposed to achieve the perfect reconstruction. In the same way, 3D wavelets can be constructed as separable product of 1D wavelets by successively applying a 1D analyzing wavelet in three spatial directions (x,y,z). Fig.2 shows a separable 3D decomposition of a volume: after being applied on the rows and on the columns, the analysis/synthesis filters followed by a 2 to 1 decimation are applied along the third dimension. At the end of the decomposition, 8 subvolumes of lower resolution are obtained: the approximation subvolume from resolution '-1' named V_1 to V_7 subvolumes of details. The separable 3D wavelets will provide an equal decorrelation of the original volume voxels in the three directions.

b. Watermark encoding

Spread spectrum communication schemes transmit a narrow-band signal via a wide-band channel by frequency spreading [4]. For watermarking, ideas from spread spectrum communications are highly applicable: a narrow-band signal (the watermark) has to be transmitted via a wide-band channel with interference (the image or video signal). Specifically, the idea of direct sequence spread spectrum communication can be adopted for watermarking of video in a similar fashion , as explained in the following. Let us denote

$$m_k \in \{0,1\} \qquad k=1,2,\ldots,N_b \qquad (2)$$

a sequence of watermark bits that has to be embedded into the video stream. This discrete signal is spread by a large factor *cr*, called the chip-rate, to obtain the spread sequence

$$b_{i,k}=m_k, \quad i=1,2,\ldots,cr, \quad k=1,2,\ldots,N_b \qquad (3)$$

The purpose of spreading is to add redundancy by embedding one bit of information into *cr* pixels of the video signal. The spread sequence b_i is amplified with a locally adjustable amplitude factor $\alpha_k \geq 0$ and is then modulated by a binary pseudo-noise sequence

$$p_{i,k}, \quad p_{i,k} \in \{-1,1\}, k=1,2,\ldots,N_b \qquad (4)$$

which serves for frequency spreading . The modulated signal, i.e. the spread spectrum watermark

$$w_{i,k} = a_i \cdot b_{i,k} \cdot p_{i,k}, \quad i=1,2,\ldots,cr, k=1,2,\ldots,N_b \qquad (5)$$

is added to the 3D WT coefficients of video chunks, i.e. V={$V_{i,j}$ }and j=1,2,...7, yielding the watermarked video signal for the k'th information bit

$$V_{i,j} = U_{i,j} + w_{i,k}, \quad i=1,2,\ldots,cr, j=1,2,\ldots,7, k=1,2,\ldots,N \qquad (6)$$

U_i denotes the initial value of 3D WT coefficient. Due to the noisy nature of the pseudo-noise signal $p_{i,k}$, the watermark signal $w_{i,k}$ is also a noise-like signal and thus difficult to detect, locate, and manipulate. For simplicity, a binary pseudo-noise sequences is assumed in (4). There exist infinitely many such sequences. Such sequences can for example be generated by feed-back shift registers producing m-sequences, any other random number generator, or by chaotic physical processes. Since the pseudo-noise signal is the secret key for embedding and retrieval of the watermark, and for security reason, sequences should be used that are not easy to guess.

c. Watermark Embedding Process

The watermark embedding process consists of the following three steps;

(1)Wavelet Transform of the Original video signals

The original video signals A_0 is decomposed into the MRR by applying the wavelet transform at several times. In the MRR, A_0 is decomposed both into the set of wavelet coefficient vectors $LLH_1, LHL_1, LHH_1, \cdots, HHH_1$ at every resolution level, and into the coarsest approximation LLL_l.

(2) Modification of the Wavelet Coefficient Vectors by the

watermark.

It is usually true that the human eyes are not sensitive to the small geometric changes in bumpy areas[4]. So in order to make the embedded watermark perceptually invisible, only the wavelet coefficient vectors $LLH_1, LHL_1, LHH_1, \cdots, HHH_1$ (i.e. $V_1, V_2, \ldots, V_7$) should be selected to be modified in the watermarking.

(3)Inverse Wavelet Transform for the modified Wavelet Coefficient Vectors

Inverse wavelet transform is taken for the set of modified wavelet coefficient vectors $LLH_1, LHL_1, LHH_1, \cdots, HHH_1$ and the unchanged coarsest approximation LLL_1. The watermarked video data $\hat{A}_0$ obtained from the result of the inverse transform. The watermarked video $\hat{A}_0$ will be distributed instead of the original video data.

d.Watermark Extraction Process

The watermark extraction process consists of following steps;

(1) Wavelet Transform of the Watermarked video Data

The received (watermarked) video data $\hat{A}_0$ will be decomposed respectively into the MRR $L\hat{L}H_1, L\hat{H}L_1, L\hat{H}H_1, \cdots, H\hat{H}H_1 (i.e. \hat{V}_1, \hat{V}_2, \ldots, \hat{V}_7)$ and $L\hat{L}L_1$ ($\hat{V}'$)by applying the wavelet transform.

(2) Calculations of the correlation coefficients

Authorized recovery of the hidden information is easily accomplished, even without knowledge of the original, unwatermarked signal, by means of a correlation receiver. Prior to the correlation step, the input signal, i.e. the watermarked video sequence, is attacked, yielding a distorted watermarked video signal which 3D WT coefficients $\hat{V}_{i,j} = U_{i,j} + w_{i,k} + n_t$. The correlation coefficient $Z_{k,j}$ for the k'th information bit is defined as

$$z_k = \mathrm{corr}(\hat{V}_{i,j}, p_{i,k}) \tag{7}$$

where

$$\mathrm{corr(A,B)} = \tilde{A} \cdot \tilde{B} / \sqrt{(\tilde{A} \cdot \tilde{A})(\tilde{B} \cdot \tilde{B})},$$

$$\tilde{A} = (A - mean(A)), \tilde{B} = (B - mean(B)) \quad \text{and}$$

$$\tilde{A} \cdot \tilde{B} = \sum_{i=1}^{N} \tilde{A}(i)\tilde{B}(i).$$

The k'th information bit is decided as follows;

$$m_k = \begin{cases} 1 & if \quad z_{k,j} > \tau_j \\ 0 & if \quad z_{k,j} < \tau_j \end{cases} \tag{8}$$

τ_j is the adaptive threshold which depends the mean values of $Z_{k,j}$. Here, $\tau_j = \beta \cdot mean(z_{k,j})$, $\beta = 2\sim2.5$,so that the embedded information can be retrieved losslessly. (8) means that the transmitted bit was a +1 if the correlation between the video signal with embedded watermark containing the current bit and the pseudo-noise signal is positive. If the correlation is negative, the transmitted bit was a -1.

If the wrong pseudo-noise sequence is used, or if it is not in synchronization with the pseudo-noise sequence as used for embedding, the scheme does not work, and the recovered bits are random. Thus, the watermark decoder has to know the pseudo-noise sequence and its possible shift. If the pseudo-noise sequence is known, but its shift is unknown, synchronization can found by means of a sliding correlator: all possible shifts are experimentally applied, and the right shift is found, if the modified correlation sum is significantly larger than for all other shifts. However, finding the correlation is cumbersome and complex, especially for pseudo-noise sequences with a very large cycle.

3. EXPERIMENT RESULTS

Our watermarking algorithm has been tested over different video. We considered two test sequences: "Tennis"and "foreman", which represent two different video scenarios. The two sequences, in CIF format (frame size: 352 x 288 pixels, progressive scan, 4:2:0 subsampling format) are used in the experiment.

The three dimensional WT utilizes the high degree of temporal correlation between successive frames in a video sequence. In contrast to motion vector implementations of inter-frame compression, performing a 3D WT involves using the same technique in all three dimensions (horizontal, vertical and temporal). The $M \times N \times T$ WT cube contains information regarding each of the T image frames. Table 1 shows average PSNR performance for each frame in a GOP structure. Note that for digital images, noise with PSNR higher than 40 dB is hardly noticeable in general. It can be seen that the proposed method doesn't causes perceptually artifacts. Fig.3 and Fig.4 are the results of the original video and watermarked video for two video signal by the proposed method respectively.

BER are calculated as follows:

$$BER = \frac{bit \ errors}{total \ embedded \ bits} \times 100\% \tag{9}$$

The advantage of the 3D approach compared to the frame-by-frame approach is first that it is more robust to averaging attacks since the mark is spread into a volume taking into account the temporal dimension of the video. In addition, the 3D approach offers a larger bandwidth to hide data.A new oblivious approach has been presented for video watermarking which, in contrast to existing methods, considers the video as a three-dimensional signal with two dimensions in space and one dimension in time, and embeds the watermark in the 3D DCT dimensional chunks of video scene. The experiments show that the proposed method is robust to common attackers. The presented method is oblivious, i.e. it does not need any information from the original video during the watermark extraction. We demonstrate the robustness of the watermarking procedure to several video distortions.

REFERENCES

[1]Chiou Ting Hsu and Ja Ling Wu, "Hidden Digital Watermarks in Images," IEEE Trans. On Image Processing, Vol.8, No.1,pp.58-68, January 1999

[2]F.Hartung and B.Girod. Watermarking of uncompressed and compressed video. Signal Processing, 66:283-301, 1998

[3] G.; Thiran, J.-P.,"Lossy to lossless object-based coding of 3-D MRI data Menegaz, "Image Processing, IEEE Transactions on , Volume: 11 Issue: 9 , Sept. 2002 ,pp: 1053 -1061

[4] Lin Luo; Yunnan Wu; Jin Li; Ya-Qin Zhang ",3-D wavelet compression and progressive inverse wavelet synthesis rendering of concentric mosaic",Image Processing, IEEE Transactions on , Volume: 11 Issue: 7 , July 2002 pp: 802 -816

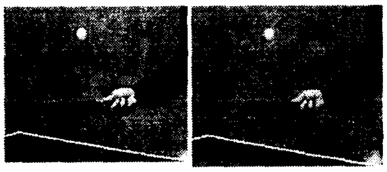

Fig3. A frame from the Tennis video, (a) original, (b) watermarked

Table 1 Example 1(tennis video) of estimated and measured bit error rates for embedded watermarks

Chip rate (*cr*)	Amplification α	PN variance σ_z^2	PSNR	Measured BER
5000	3	1	35.4214	0
6050	3	1	34.5849	0
1800	2.5	1	41.4386	0.0014
3200	2.5	1	38.9442	1.0000e-004
4050	2.5	1	37.9270	0
5000	2.5	1	37.0051	0
6050	2.5	1	37.0051	0
200	2	1	52.8350	0.0420
450	2	1	49.3544	0.0287
800	2.	1	46.8856	0.0157
3200	2	1	40.8824	7.0000e-004
4050	2	1	39.8652	1.0000e-004
5000	2	1	38.9433	0
6050	2	1	38.1067	0
3200	1.5	1	43.3812	0.0024
4050	1.5	1	42.3640	0.0012
5000	1.5	1	41.4420	3.0000e-004
6050	1.5	1	40.6055	2.0000e-004
7000	1.5	1	39.8609	4.0000e-005

Fig4. A frame from the Tennis video, (a) original, (b) watermark