3D DCT Video Information Hiding

Young Gon Kim*, Jie Yang*,**, Hye-Joo Lee***, Jin Woo Hong***, Moon Ho Lee*

*Institute of Information & Communication ,Chonbuk National University, Korea
**School of Electronics and Information Wuhan University of Technology, China,
***Broadcasting Media Department, Radio and Broadcast Laboratories, Electronics and Telecommunications Research Institute

Abstract: Embedding information into video data is a topic that recently gained increasing attention. 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 cosine transform of video scene, in contrast with previous works on video watermarking where each video frame was marked separately, or where only intra-frame or motion compensation parameters were marked in MPEG compressed videos. The watermark sequence used is encrypted, pseudo-noise signal to the video. The performance of the presented technique is evaluated experimentally .

1. INTRODUCTION

An increasing number of movies and other video document are recorded on digital supports for public as well as for professional applications. 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][3]. When the ownership of the media is in question, the information can be extracted to characterize the owner or distributor.

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 DCT image watermarking scheme described in [4]and [5] to a three-dimensional DCT video watermarking scheme. With this novel approach, the watermark is embedded into the magnitude of the 3D discrete cosine transform of the video data. The ownership and copyright information are encrypted in key, which is adaptively added into the magnitude values of the three dimensional discrete cosine transform domain[6].

In this paper, we propose a novel 3D 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" watermaking 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 DCT watermarking and lists some relevant properties of the 3D Fourier transform. In section 4 the watermark embedding/extraction processes are detailed. Section 5 presents the 3D watermarking experimental results.

2. VIDEO WATERMARKING

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 frequencies to make the watermark robust. Here 3D discrete cosine transformation is utilized.

The Discrete Cosine Transform is a real-valued, separable orthonormal transform whose basis vectors are composed of samples of cosine functions. The 3D DCT analysis is defined as follows:

Let X be a 3D signal of size M by N by T. Let Y be the 3D DCT of X, also of size M by N by T. The elements of Y can be calculated as

$$Y_{u,v,w} = \alpha_m \alpha_n \alpha_t \sum_{t=0}^{T-1} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X_{m,n,t} \cos\frac{\pi(2t+1)u}{2T} \cos\frac{\pi(2m+1)v}{2M} \cos\frac{\pi(2n+1)w}{2N} \quad (1)$$

$$\alpha_m = \begin{cases} 1/\sqrt{T} & , m=0 \\ 2/\sqrt{T} & , otherwise \end{cases} \quad \alpha_n = \begin{cases} 1/\sqrt{M} & , n=0 \\ 2/\sqrt{M} & , otherwise \end{cases} \quad \alpha_t = \begin{cases} 1/\sqrt{N} & , t=0 \\ 2/\sqrt{N} & , otherwise \end{cases}$$

The inverse 3D DCT of $Y_{u,v,w}$ can be calculated as

$$X_{m,n,t} = \alpha_u \alpha_v \alpha_w \sum_{w=0}^{T-1} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} Y_{u,v,w} \cos\frac{\pi(2u+1)m}{2T} \cos\frac{\pi(2v+1)n}{2M} \cos\frac{\pi(2w+1)t}{2N} \quad (2)$$

$$\alpha_u = \begin{cases} 1/\sqrt{T} & , u=0 \\ 2/\sqrt{T} & , otherwise \end{cases} , \alpha_v = \begin{cases} 1/\sqrt{T} & , v=0 \\ 2/\sqrt{T} & , otherwise \end{cases} \quad \alpha_w = \begin{cases} 1/\sqrt{T} & , w=0 \\ 2/\sqrt{T} & , otherwise \end{cases}$$

Since the DCT is separable, this 3D transform is the same as finding the 1 dimensional DCT along each of three dimensions of X. When the 3D-DCT is applied to video, the transformation is also applied to the time dimension. The 3D-DCT will be more efficient than frame based 2D-DCT when there is correlation between frames.

We propose three DCT domain based methods to embed watermarks without incurring noticeable visual artifacts. As mentioned in [4]. We embed the watermark

---

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

information in the middle band $\Re$ (fig.1) to tradeoff between visual quality and attack robustness. First, we define some notations.

1. VOB (Volume of blocks) is composed of more adjacent video frame blocks which form a basic unit in our algorithm.
2. $B^r_{x,y,z}$ represents an volume block of $M \times N \times T$ pixels in the non-overlapping partition of the r-th watermark volume block. The block coordinates x, y and z are in the range of $0 \le x < \frac{J}{M}$, $0 \le y < \frac{K}{N}$ and $0 \le z < \frac{L}{T}$,respectively, where J, K and L stand for the video size.
3. $D^r_{x,y,z} = \{d^r_{x,y,z}(p,q,r) \mid 0 \le p < M, 0 \le q < N, 0 \le r < T\}$ represents the $M \times N \times T$ DCT coefficients corresponding to $B^r_{x,y,z}$.

Embedding algorithm:

A VOB consists of only one volume block $B^r_{x,y,z}$ so that the concatenation to form a watermark volume can be more efficient.

Step 1: Construct volumetric images by combing each T image into a stack. Here T is the depth of the constructed stack.

Step 2: The forward 3D DCT of the volumetric images is calculated as the equation (1).

Step 3: Divide each volumetric image into $M \times N \times T$ volumetric blocks. (M and N are 8, T is 6 in our experiments.)

Step 4: For each VOB in the *r*th watermark volume, We compute

$$\mu^r_{x,y,z} = \frac{d^r_{x,y,z}(i,j-1,k) + d^r_{x,y,z}(i,j,k)}{2}, (i,j,k),(i,j-1,k) \in \Re \qquad (3)$$

Step 5: Calculate $d^{r\,'}_{x,y,z}(i,j-1,k)$ and $d^{r\,'}_{x,y,z}(i,j,k)$ by

$$\begin{cases} d^{r\,'}_{x,y,z}(i,j-1,k) = \mu^r_{x,y,z} - (-1)^{w(l)} \cdot q \\ d^{r\,'}_{x,y,z}(i,j-1,k) = \mu^r_{x,y,z} + (-1)^{w(l)} \cdot q \end{cases} \qquad (4)$$

Where q is a constant quantity and $w = \{w(l) \mid l = 1,2,\ldots,ll\}$ is the watermark bit information to be embedded, where ll is the length of the watermark sequence.

Step 6: Calculate the watermarked volumetric block $B^{r\,'}_{x,y,z}$ via inverse 3D DCT transformation of

$D^{r\,'}_{x,y,z}$ in the equation (2).

For security consideration, the embedding orders of coefficients in one VOB can be randomized by a key, which should be provided in watermark retrieval. The watermark sequence used is encrypted, pseudo-noise signal to the video. The noise-like watermark is statistically undetectable to thwart unauthorized removal.

Retrieving algorithm:

Authorized recovery of the hidden information is easily accomplished without the knowledge of the original video: the algorithm is blind. To extract the watermark bit information, we construct watermarked volumetric images by combing each T image into a stack and divide each volumetric image into $M \times N \times T$ volumetric blocks like embedding process. For each watermark bit, calculate e_k by

$$e_k = \sum_{r=1}^{mm} d^{r\,''}_{x,y,z}(i,j-1,k) - d^{r\,''}_{x,y,z}(i,j,k) \qquad (5)$$

If $(e_k\ {}_0)$ then $\widetilde{w}(l) = 1$, else $\widetilde{w}(l) = 0$.

Obviously, our retrieving algorithm takes advantage of multiple embedding of watermark data in more than one watermark volume. We calculate the correlation between the extracted watermark and the original watermark, that was used in embedding phase:

$$R(z) = xcorr(w, \widetilde{w}) \,. \qquad (6)$$

If the absolute value of the correlation is higher than the threshold (*th*=15), the watermark is present, otherwise not.

3. EXPERIMENT RESULTS

Our watermarking algorithm has been tested over different bit rates. We considered three test sequences: "Tennis", "Miss" and "foreman", which represent three different video scenarios. The three 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 DCT 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 DCT involves using the same technique in all three dimensions (horizontal, vertical and temporal). The $M \times N \times T$ DCT cube contains information regarding each of the T image frames. The DCT frame 0 contains much of the information in the DCT cube, while the opposite is true for other DCT frames. Most of the information in DCT frames 1 to T-1 is contained in the area of motion. The most of non-zero DCT coefficient after quantization are concentrated on the frame 0 to 2, so we embed the watermark bits only in DCT frame 0 to 2 in the equation (3) and (4). An original, watermarked frame from the video sequence and watermark detection response are shown in Fig2, Fig.3 and Fig.4.

Table 1,2 and3 show average PSNR performance for each frame in a GOP structure. Note that for digital images, noise with PSNR higher than 30 dB is hardly noticeable in general. It can be seen that the proposed method doesn't causes perceptually artifacts.

We conducted tests of adding Gaussian noise. The attacks of resizing (with re-sampling), quantizatio and re-quantization, etc. can be modeled by signal noise addition. BER are calculated as follows:

$$BER = \frac{bit\ \ errors}{total\ \ embedded\ \ bits} \times 100\% \qquad (7)$$

Here, the total embedded bits refers to the total amount of bits the watermarker software attempts to embed in the sequence. The simulating results are shown in Fig 5.

One of the main requirements on watermarking schemes is robustness against intentional or unintentional attacks attempting to remove or destroy the watermark. It is possible that attackers may try to destroy the embedded watermark by filtering (different filters) and re-quantization with a coarser step size (i.e., a larger quantization factor Q_f). The test results for the attackers are shown in Fig.6, 7, 8, 9 and Table 4.

4. CONCLUSIONS

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] Swanson, M.D.; Bin Zhu; Chau, B.; Tewfik, A.H. Swanson, M.D.; Bin Zhu; Chau, B.; Tewfik, A.H." Multiresolution video watermarking using perceptual models and scene segmentation," Image Processing, 1997. Proceedings., International Conference on , 1997,pp: 558 -561 vol.2

[4] Wen-Nung Lie; Guo-Shiang Lin; Chih-Liang Wu; Ta-Chun Wang." Robust image watermarking on the DCT domain", Circuits and Systems, 2000. Proceedings. ISCAS 2000 Geneva. The 2000 IEEE International Symposium on , Volume: 1 , 2000, 228 -231 vol.1

[5] Yinghui Wu; Xin Guan; Kankanhalli, M.S.; Zhiyong Huang "Robust invisible watermarking of volume data using the 3D DCT ," Computer Graphics International 2001. Proceedings , 2001, pp: 359 -362

[6]Wenwu Zhu; Zixiang Xiong; Ya-Qin Zhang "Multiresolution watermarking for images and video: a unified approach,"Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on , Volume: 1 , 1998,pp: 465 -468 vol.1

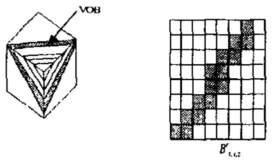


Fig. 1 The embedding scheme of the proposed algorithm

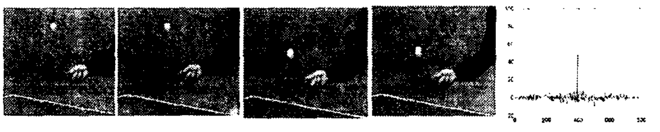

(a) (b) (c) (d) (e)

Fig2. A frame from the Tennis video, (a) and (c)original, (b)and (d) watermarked, (e)detection response


(a) (b) (c) (d) (e)

Fig3. A frame from the Miss video, (a) and (c)original, (b)and (d) watermarked,(e) detection response


(a) (b) (c) (d) (e)

Fig4. A frame from the Foreman video, (a) and (c)original, (b)and (d) watermarked,(e) detection response

Table 1 PSNR in a GOP after watermark embedded in the Foreman video

Frame	Before Watermarking (dB)	After Watermarking (dB)	Change in PSNR (dB)
1	41.6296	32.6699	8.9597
2	41.7306	38.0449	3.6857
3	41.8683	37.5910	4.2773
4	41.9514	37.5499	4.4015
5	41.8974	36.4863	5.4111
6	41.9837	36.5942	5.3895
7	41.8729	37.5532	4.3197
8	41.7106	35.9465	5.7641
9	41.7618	32.3637	9.3981
10	42.2601	38.3627	3.8974
11	42.5086	36.8227	5.6859
12	42.7240	37.1198	5.6042

Table 2 PSNR in a GOP after watermark embedded in the Tennis video

Frame	Before Watermarking (dB)	After Watermarking (dB)	Change in PSNR (dB)
1	43.2401	32.22	11.0201
2	44.1408	37.0996	7.0412
3	44.6098	36.9187	7.6911
4	44.6927	37.3006	7.3921
5	44.7311	36.7536	7.9775
6	44.6802	36.6612	8.019
7	44.2882	36.6673	7.6209
8	43.1695	36.1230	7.0465
9	44.5968	32.0289	12.5679
10	45.0348	36.5798	8.455
11	44.5087	36.5381	7.9706
12	43.8109	36.8850	6.9259

Table 3 PSNR in a GOP after watermark embedded in the Miss video

Frame	Before Watermarking (dB)	After Watermarking (dB)	Change in PSNR (dB)
1	46.1216	33.1612	12.9604
2	47.3277	42.0139	5.3138
3	47.4600	42.3600	5.1000
4	47.4288	43.9609	3.4679
5	47.3875	42.5785	4.8090
6	47.4254	44.3384	3.0870
7	47.3833	44.1965	3.1868
8	46.4444	43.4246	3.0198
9	46.7961	33.2396	13.5565
10	47.6081	42.1826	5.4255
11	47.5987	42.5811	5.0176
12	47.4490	44.2380	3.2110

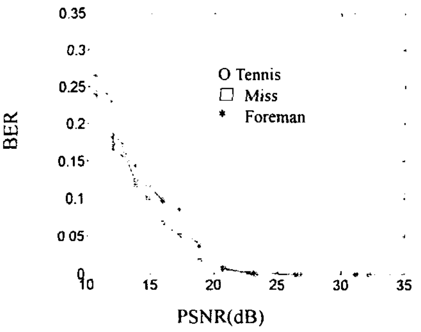
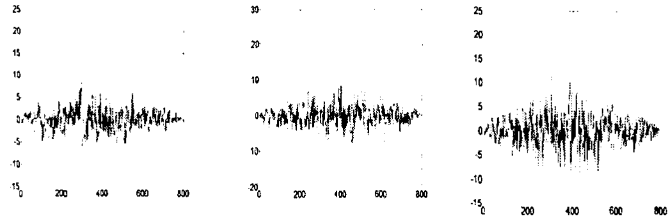


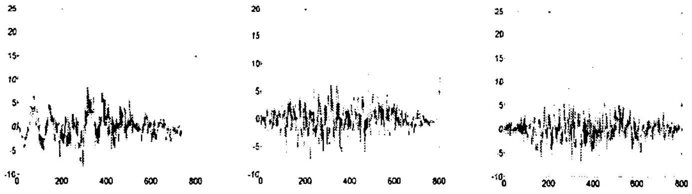
Fig.5 BER vs PSNR for the three video



(a) watermarks (b)watermarks (c)watermarks

Fig. 6 Watermark detection response after the filtering process for Tennis video

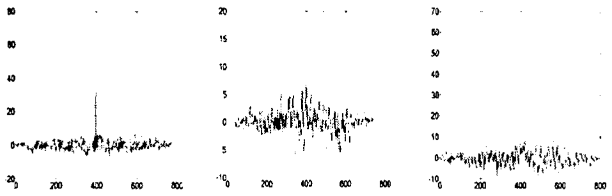
(a)Low pass filter, (b)Median filter (c)Wiener filter



(a) watermarks (b)watermarks (c)watermarks

Fig.7 Watermark detection response after the filtering process for Miss video

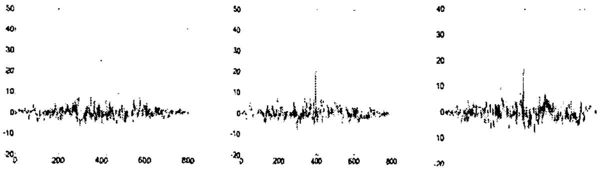
(a)Low pass filter, (b)Median filter (c)Wiener filter



(a) watermarks (b)watermarks (c)watermarks

Fig. 8 Watermark detection response after the filtering process for Foreman video

(a) Low pass filter, (b)Median filter (c)Wiener filter



(a) (b) (c)

Fig. 9 Watermark detection response with Q_f=32.

Tennis, (b)Miss, (c)Foreman

Table 4 The experiment results after the filtering process

Object	Tennis		Miss		Foreman	
	PSNR (dB)	BER	PSNR (dB)	BER	PSNR (dB)	BER
Lowpass filter	26.331	0.3788	32.8488	0.3903	25.4251	0.282
Midian filter	28.713	0.3840	38.3572	0.4697	31.8224	0.330
Winener filter	32.863	0.3586	42.3422	0.3838	32.9394	0.295
Q_f=16	33.314	0	35.5448	0	32.9773	0
Q_f=18	36.053	0.0985	34.6425	0.3384	32.3253	0.146
Q_f=32	28.949	0.1717	29.9607	0.3611	28.5139	0.181