Robust and High Quality Video Watermarking with the use of Temporal Redundancy

K. F. Tsang, O. C. Au
Department of Electrical and Electronic Engineering
The Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong

ABSTRACT

In this paper, we present a new video watermarking technique, which utilize the characteristic of temporal redundancy of video sequence to improve the quality and robustness of the watermarked sequence. The proposed watermarking technique can be combined with many existing 2D image-watermarking algorithm to take advantage of their robustness against various attacks. For every watermark bit a_i, the pseudo random sequence $(-1)^{a_i} p$ is added to the mid-band coefficients of a block and the complementary sequence, i.e., $-(-1)^{a_i} p$ is added to the same block in another frame of the insertion pair. To retrieve the embedded watermark bit, the block in the first frame is subtracted by the same block of the second frame, resulting in a watermark with double the magnitude $(-1)^{a_i} p - \left[-(-1)^{a_i} p\right] = 2(-1)^{a_i} p$. Since adjacent video frames are highly correlated and the DCT coefficients are almost the same (especially in non-moving regions), the subtraction of the block pair also cancels out most of the interfering DCT coefficients originated from the host signal, such that the interference to the watermark signal from the host signal is minimized during detection. Experiment shows that the probability of receiving an error bit can be reduced or the picture quality is improved while maintaining the same probability of error.

Keywords: Video watermarking, Robustness, Temporal redundancy

1. INTRODUCTION

Illegal copying of videotape in the 80s' is not a serious problem, since the time of reproduction is long and expensive. However, since the existence of Video Compact Disc (VCD), video content can be reproduced easily and perfectly in the digital domain. Copyright infringement of video content becomes a serious problem nowadays. We can also notice that pirates are now putting full-length movies on the Internet for download. The problem will become even more serious as the speed of the Internet increase and many Internet users get a broadband connection directly to the Internet at their home. As a result, digital video watermarking becomes more and more important and is regarded as a tool for copyrights protection. There are two main concerns on digital watermarking imposed by different group of people. From the copyright owners' point of view, robustness of the embedded watermark is the most important, since the copyrights, copy control, or play control information are stored in the watermark. On the other hand, end users are not concern about the robustness but the perceptual quality of the pictures. These are the two basic requirement of a watermarking system. The problem is difficult due to these two contradicting requirements. Generally speaking, as the robustness of a watermark increases, the perceptual distortion to the pictures tends to increase.

In video watermarking, most of the existing algorithms insert the watermark into each frame independently using watermarking techniques developed for 2D image. In [3,6,8-9], spread spectrum watermark sequence is inserted into each frame independently using 2D image watermarking techniques. In [5], the 3D (2D + temporal) video sequence is line-scanned to form an 1D signal, and the watermark is inserted using spread spectrum communication techniques.

This kind of approach has the advantage that it shares all the advancements, which are well developed for 2D images, in terms of visual quality and robustness to attacks like transformation, scaling, digital compressions, etc., and also it is robust to frame skipping or changing of frame rate (since each frame are watermarked and watermark can be extracted from one single frame).

However, if we apply the same 2D image watermark to each individual frame, the well known watermark attack, the Averaging Attack, can destroy the embedded watermark easily. This attack aims at separating the watermark $U=f(M,W)$ from the watermarked signal. Attacker can figure out U by adding or averaging multiple watermarked images, say, a sequence of frames of a marked video, M_1+U, M_2+U, ..., M_N+U. Since there are usually a large number of frames that tend to be uncorrelated in the long run, the output of the averaging process is simply the watermark U. The attacker can then extract U from the images. Simple solution to this approach is to use 2 or 3 different watermarking *KEY* alternatively for each video frames, or insert the watermark in different location for each frame, where the locations are chosen with a pseudo-random sequence.

In this paper, we present a new video watermarking technique for correlation based watermark detector, which utilize the characteristic of temporal redundancy of video sequence to ***improve the quality*** and ***robustness of the watermarked sequence,*** while using 2D image watermarking algorithm underneath, such that it can be applied with those advance techniques to against different attacks.

In section 2, we give an example on correlation based watermarking technique. We then describe the proposed "add and subtract" video watermarking insertion and extraction technique in Section 3 and 4 respectively. Experimental Results are showed in Section 5. We conclude in Section 6.

2. SPREAD SPECTRUM WATERMARKING

Many of the existing watermarking techniques use correlation based watermark detector. The most classical example is the Cox's method [1]. It simply inserts a pseudo random sequence into the DCT coefficients with large magnitude, and the PN sequence is generated with a key K. This key is then known as the watermark. During the detection, the watermarked media is first subtracted by the original media in DCT domain. The residual is passed through a correlator for *similarity* test with an input test key J. If the correlation between the embedded key and the test key is larger then a threshold, it will be regarded as a watermark being detected (i.e. the test key is the same as the embedded key).

This kind of correlation-based detection has been using in communication system with spread spectrum modulation for a long time. In spread spectrum modulation, a narrow-band signal is modulated to a much-wider-band signal with a noise-like pseudo random signal, which is generated with a *KEY*. The spread spectrum modulation gives the advantages that the detection is self-synchronized. The embedded information can be extracted without the use of original media and it is hard to be detected without the knowledge of the pseudo random sequence, because the modulated signal is just like background noise.

Here is an example of spread spectrum watermark generation [5]. Let $\{a_j\}$ be the binary sequence (+1 or −1) of the information bits that we would like to insert into a frame. This binary sequence is spread with the chip rate r_c to obtain the spread sequence $\{b_i\}$ of length r_c. The purpose of spreading is to add redundancy by embedding one bit of information into r_c pixels (in case of time domain insertion) or frequency bins (in case of frequency domain insertion) of the video frame. The spread sequence $\{b_i\}$ is then modulated with a pseudo noise sequence $\{p_i\}$ to spread the signal in frequency domain. The final watermark w_i is scaled with a scalar α before insert to the j frame of the host media, while α can also be applied adaptively when human visual model is used, i.e.:

$$w_i = \alpha \cdot b_i \cdot p_i \qquad \text{where } j \cdot r_c \leq i < (j+1) \cdot r_c \qquad (1)$$

Common detection algorithm is to correlate the watermark with the watermarked media. If W is chosen at random, the correlation between M and W is very small, as the random positive and negative terms tend to cancel themselves out, especially for a long sequence. However, in computing $W \bullet W$ all of the terms are positive, adding up to a large value comparing to $M \bullet W$. For this reason, the product $M' \bullet W = M \bullet W + W \bullet W$ can estimate $W \bullet W$ accurately. As a result, the embedded watermark can be retrieved without the use of the original media M. If the correlation gives an positive value (and larger then a threshold) the embedded a_j in frame j is 1, or −1 if it is negative.

Common effectiveness measurement of a spread spectrum watermarking system is the Bit Error Rate (BER) which is defined as:

$$BER = \frac{Number\_of\_error\_bit}{Number\_of\_embedded\_bit} \tag{2}$$

There are two main factor affect the BER of a watermarking system. One is the watermark strength, i.e the α value. The other factor is the interference from the host media.

As in communication system, the BER or the probability of error is inversely proportional to the α value. One can decrease the BER by increasing the strength of the watermark. However, increase of the strength of the watermark would induce more distortion to the host media, or the watermark become less imperceptible.

The other way of improving the BER is to reduce the interference from the host media. This can be done by selecting a PN sequence which is less correlated with the host media, or by post-processing (e.g. filtering) the watermarked media before watermark detection.

Based on these two factors, we propose an "add and subtract" watermark insertion and detection technique for correlation based watermarking algorithm, which can reduce the BER by eventually increasing the strength of the watermark and reducing the interference from the host media, while improving the imperceptibility of the embedded watermark.

3. WATERMARK INSERTION

The proposed technique inserts the watermark information into the mid-band of the uncompressed video to provide a better quality and more robust watermarked sequence. Every two video frames form an insertion pairs. We divide each video frame into 8x8 blocks and DCT transform each block. 16 mid-band coefficients are selected for watermark insertion [7](see figure 1). We select the mid-band coefficients for watermark insertion because the human eyes are more sensitive to low frequency distortion and high frequency coefficients can be easily removed by low passed filtering or lossy compression. For every watermark bit $w_i = \alpha \cdot b_i \cdot p_i$, being added to the mid-band coefficients of a block, the complement of the watermark is added to the co-located block in the other frame of the insertion pairs, i.e., $-w_i$ is inserted. Note that we use 16 mid-band coefficients so that r_c is equal to 16.

For a CIF (352X288) format Video, there are 1584 of 8x8 blocks. If we embed one bit of information to each block, we can have a data rate of 1584·(30/2) = 23760bps for a 30fps sequence.

The pseudo code of the insertion is as follow:

```
for each video frame pair
        for each video block pair
                apply DCT to the block pair;
                select the mid-band DCT coefficients;
                add the watemark bit wi = α · bi · pi to the 1st block;
                subtract the watemark bit wi = α · bi · pi from the 2nd block;
                next watemark bit;
                next block pair;
        end
        next frame pair;
end
```

0	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63

Figure 1: The zigzag ordering of DCT coefficients with mid-band coefficients shown in shadowed area.

4. WATERMARK EXTRACTION

When retrieving the embedded watermark bit, the block in the first frame of the insertion pair is subtracted by the co-located block of its pair, resulting a ***double*** in magnitude of the watermark components. Since adjacent video frames are highly correlated and the DCT coefficients are almost the same (especially in the background non-moving regions of a video), the subtraction of the insertion pair also ***cancels out*** most the DCT coefficients originated from the host signal, such that the interference from the host signal is minimized during the correlation. (see Eq. 3) This is the main contribution of this work, since the bit error rate (BER) can be reduce by the facts that the magnitude of the watermarks are eventually doubled and there is less interference from the host signal; or the picture quality is improved while maintaining the same BER, according to communication theory.

$$[(M_i+w_i) - (M_{i+1} - w_i)] \bullet w_i = 2|w_i|^2 \qquad \text{assuming that } M_i \neq M_{i+1} \qquad (3)$$

The decoding of the watermark bits is the same as other spread spectrum watermarking. After the DCT block is subtracted from its pair DCT block, the extracted mid-band coefficients are correlated with the same pseudo noise sequence (i.e. generated with the same key). The embedded information bit a_j is 1 if the correlation is positive and –1 if it is negative.

The pseudo code of the extraction is as follow:

```
for each video frame pair
        for each video block pair
                apply DCT to the block pair;
                select the mid-band DCT coefficients;
                subtract 1st block by the 2nd block;
                correlate the residual with the same p_i;
                return the extracted watermark bit i as 1 if correlation > 0 or –1 otherwise;
                next block pair;
        end
        next frame pair;
end
```

5. SIMULATION RESULTS

In order to test the proposed "add and subtract" watermarking technique, we implement the algorithm in Matlab. We also implement a simple "add only" watermarking technique for comparison.

"Add only" watermarking

The "add only" watermarking embeds a binary message sequence into a video frame by first dividing a frame into 8·8 blocks. The blocks are then DCT and the spread spectrum encoded watermark sequence, $w_i = \alpha \cdot b_i \cdot p_i$, is added to the 16 mid band DCT coefficients. (see Figure 1) After processing every block, the video frame is inverse transformed back to pixel domain and proceeds to next video frame. This is basically the same as image watermarking that is now applied to every single frame. The watermark detection is simply by correlating the blocks with the same pseudo random sequence p_i. The embedded bit of the block is 1 if the correlation is larger than zero or –1 if it is smaller than zero.

Figure 2a shows the original 90th frame (resolution is 352·288, in 8-bit grayscale) of the video "salesman" and figure 2b shows the watermarked version of the frame with the watermark strength α equal to 5 applied. The binary message of [1 1 1 –1 –1] is embedded to the blocks repeatedly starting from the top left corner. The watermarked frame has a PSNR value of 41.1945dB. The embedded watermark message is extracted from the watermarked frame and the BER is calculated, which is 0.1067. The error is from the quantization noise and interference from the host media.

The proposed "add and subtract" watermarking

Frame 90 and 91 of the sequence "Salesman" are watermarked with the proposed "add and subtract" method in the experiment. The same message sequence [1 1 1 –1 –1] is inserted to the blocks repeatedly throughout this insertion pair frames. Half of the watermark strength as in the "add only" watermarking is used in this case (i.e. α equal to 2.5). Figure 3a and 3b show the watermarked frame 90 and 91 respectively. As expected, since the α value used here is smaller than the one in the normal "add only" method, the PSNR of the insertion pair frames is higher. The PSNR of the watermark frame 90 and 91 are 46.1817dB and 46.2303dB, which indicate a 5dB improvement over the normal method. The BER of this case is zero, i.e. the embedded bit can be extracted completely without an error.

Lossy compression test

In order to test the robustness of the watermarking techniques under lossy compression, the watermarked video frames are compressed individually using JPEG. (see figure 4) The PSNR and BER of the compressed watermarked frames are calculated and summarized in Table 1. We can see from the results that the PSNR and BER values are improved by using the proposed technique.

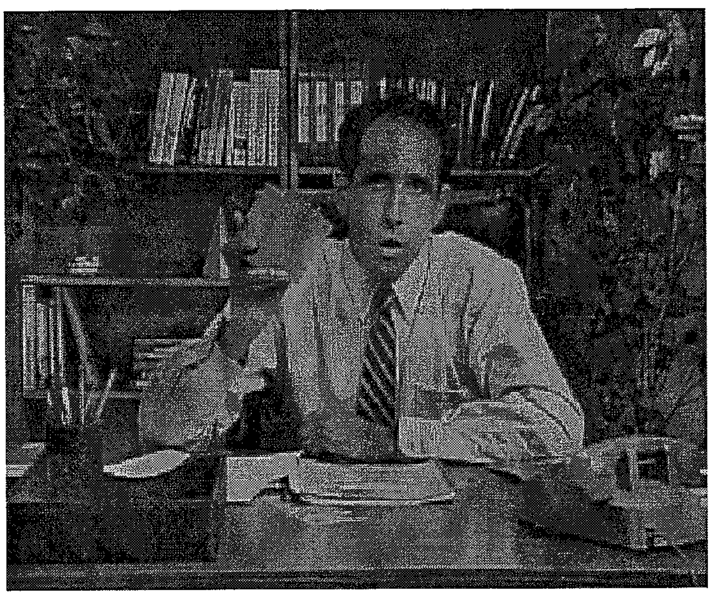

Figure 2a: The original frame 90 of the video sequence "Salesman"

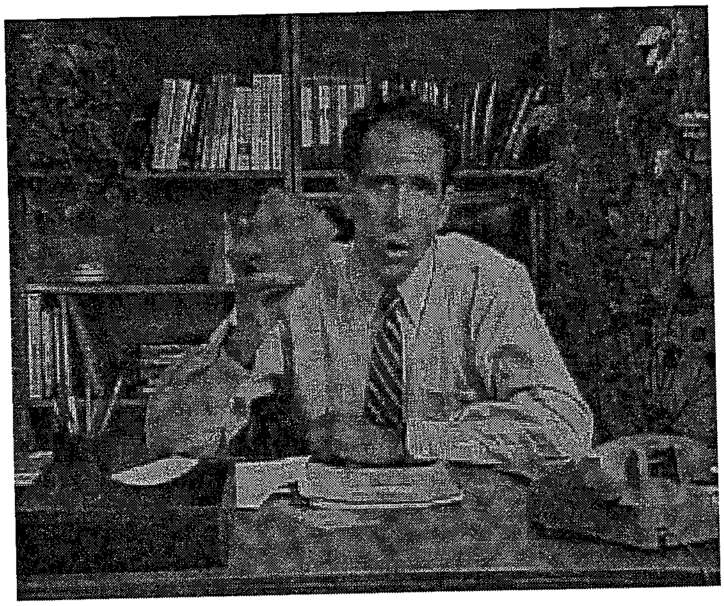

Figure 2b: The watermarked frame 90" using normal "add only" watermarking method.

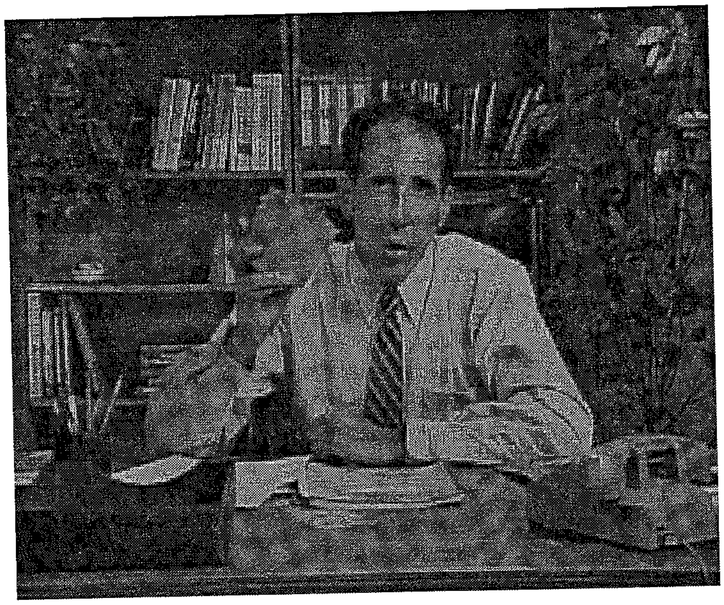

Figure 3a: Watermarked frame 90 of "Salesman" using the proposed method

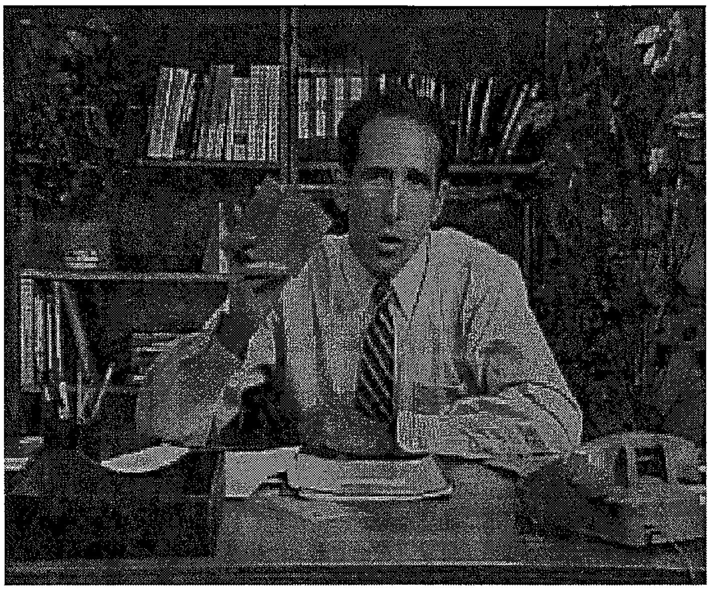

Figure 3b: Watermarked frame 91 of "Salesman" using the proposed method

Figure 4a: Watermarked frame 90 using the proposed method and JPEG compressed with quality factor 99

Figure 4b: Watermarked frame 91 using the proposed method and JPEG compressed with quality factor 99

Summary of the experimental results			
	Watermarked frame	Watermarking Technique	
		Ordinary "Add only"	Proposed "Add and Subtract"
PSNR (dB)	No compression	41.1945	46.1817 / 46.2303*
	JPEG (factor=50)	33.7408	34.0545 / 34.0529*
	JPEG (factor=99)	23.0468	23.0493 / 23.0609*
BER	No compression	0.1067	0
	JPEG (factor=50)	0.2847	0.2045
	JPEG (factor=99)	0.4198	0.3984

* note: frame 90 / frame 91

Table 1: Summary of the experimental results.

6. CONCULSIONS

In this paper, we present a *"add and subtract"* watermark insertion and extraction technique for correlation based watermarking algorithm, with the use of the important characteristic of video sequence: Temporal Redundancy. The watermark information is modulated to a spread spectrum signal, which can improve the detection accuracy and robustness due to its statistical advantages. We add and subtract the spread spectrum watermark in the mid-band DCT coefficients of the insertion pair frames. For watermark extraction, once we subtract the first frame by its pair frame, the remaining residuals are mostly the watermark signal and it has double of the magnitude inserted originally. This important feature, which utilizes the temporal redundancy characteristic, can improve the detection accuracy significantly. On the other hand, we can insert a smaller watermark signal to reduce the distortion to the video while keeping a level of detection accuracy. Experiment shows that our technique can improve the watermark robustness to lossy compression too. The proposed "add and subtract" video watermarking techniques can be used with many other correlation based 2D image watermark insertion and extraction algorithms, rather than the novel one that we used for experimental propose, so that it can be robust to other attacks like rotation, transformation, scaling, etc.

REFERENCES

[1] J. Cox, J. Killian, F.T. Leighton and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia", *IEEE Trans. on Image Processing*, Vol. 6, No. 12, pp.1673-87, Dec 1997.

[2] I.J. Cox, and M.G. Linnartz, "Some General Methods for Tampering with Watermarks", *IEEE Journal on Selected Areas in Communications, Vol.* 16, No. 4, pp. 587-593, May 1998.

[3] J. Dittmann, M. Stabenau, R. Steinmetz, "Robust MPEG Video Watermarking Technologies", *Proc. of ACM Multimedia 98*, pp.71-80, 1998.

[4] F. Hartung, J.K. Su, and B. Girod, "Spread Spectrum Watermarking: Malicious Attacks and Counterattacks", *Proc. of SPIE Conference on Security and Watermarking of Multimedia Contents*, Vol. 3657, pp. 147-158, Jan 1999.

[5] F. Hartung, B. Girod, "Watermarking of Uncompressed and Compressed Video", *Signal Processing*, Vol. 66, pp. 283-301, 1998.

[6] F. Hartung, B. Girod, "Digital Watermarking of MPEG-2 Coded Video in the Bitstream Domain", *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing,* Vol.4, pp.2621-4, 1997.

[7] C.T. Hsu, J.L. Wu, "DCT-Based Watermarking for Video", *IEEE Trans. on Consumer Electronics*, Vol. 44. No. 1, pp.206-216, Feb. 1998.

[8] Inoue H, Miyazaki A, Araki T, Katsura T, "A Digital Watermark Method using the Wavelet Transform for Video Data", *Proc. of IEEE Int. Sym. on Circuits and Systems*, Vol.4, pp.247-50, May 1999.

[9] M.D. Swanson, B. Zhu, B. Chau, A.H. Tewfik, "Object-based Transparent Video Watermarking", *Proc. of IEEE First Workshop on Multimedia Signal Processing*, pp.369-374, 1997.