A SCHEME FOR JOINT WATERMARKING AND COMPRESSION OF VIDEO

Rakesh Dugad and Narendra Ahuja

Department of Electrical and Computer Engineering
Beckman Institute, University of Illinois, Urbana, IL 61801.
dugad@vision.ai.uiuc.edu

ABSTRACT

We present a scheme for jointly watermarking and compressing digital video. The amount of watermark added is *adapted* to the expected degradation of the watermark due to compression. This results in a more robust watermark. This is achieved without any appreciable decrease in the quality of the decoded video compared to the case when the watermark is not adaptive. Results are presented for the flower garden sequence.

1. INTRODUCTION

Immense proliferation of digital media has created an immediate need for methods for their copyright protection. Much effort has been put over the last half a decade towards schemes for watermarking of digital media.

Though digital video is a sequence of frames the problem of watermarking digital video has an additional set of unique requirements compared to watermarking digital images/frames individually. For example computational complexity is an important issue since the amount of data is huge and many applications have real-time constraints. While a given amount of watermark may not be visible in a still frame, the same amount of watermark when added to all the frames may create visual artifacts like pulsating of pixel values when playing the video. The problem is further aggravated by the aggressive compression (like the B-frames in MPEG like schemes which are predicated from both past and future frames) employed for storing or transmitting digital video. Surviving such compression requires adding more watermark which would create visually disturbing artifacts. The computational complexity requirement prohibits using sophisticated HVS models to get around these artifacts. The huge amount of data also makes video watermarking schemes more susceptible to collusion attacks. Other kinds of attacks include frame dropping or reordering or changing the compression format.

Digital video watermarking schemes can be divided into three broad categories: those that operate on uncompressed data [1, 2, 3], those that operate directly on the compressed bit stream [4, 5] and those in which the processes of watermarking and compression are combined together [6]. The scheme in [1] works in the uncompressed domain. It segments the given video sequence into scenes. Then a 3-D wavelet transform of the each segment is performed. Explicit model of HVS is used to provide spatial and temporal masking. The watermark is added to each temporal component of the 3-D wavelet transform. The watermark that is added to the high frequency temporal components is localized and changes rapidly with time whereas that added to low frequency temporal components is spread over the entire sequence and changes slowly over time. This makes the scheme robust to such attacks as frame dropping or averaging and the watermark can be retrieved from a single video frame. However the scheme has high computational complexity.

The scheme in [4] operates on MPEG-2 compressed bit stream. The watermark is pseudo-random sequence of +1,-1 of the same size as the given video sequence. 8 × 8 block DCT of the watermark is performed. Each non-zero coefficient corresponding to each of the VLC codewords in the compression bit stream is watermarked by adding the corresponding watermark coefficient. An additional constraint is that the length of the VLC code word after adding the watermark should not be more than its length before adding the watermark. This makes sure that the bit rate of the compressed stream is not increased. A drift compensation scheme is proposed to make sure that the encoder and decoder do not get out of sync due to addition of the watermark.

The scheme in [6] embeds the watermark in the GOP structure of the compressed bit stream. Since most encoders use standard GOP structures like IPBBPBB... the watermark is embedded by choosing irregular GOP structures that are unlikely to be normally used. However this scheme could be subverted by decompression and recompressing using a different GOP structure. Moreover it can conflict with the rate control algorithm of the encoder.

Our scheme falls in the category of combining the processes of watermarking and compression. We focus on MPEG-

The support of the Office of Naval Research under grant N00014-96-1-0502 is greatfully acknowledged.

0-7803-6297-7/00/$10.00 © 2000 IEEE

2 kind of compression schemes. The idea is to monitor the amount of compression and the resultant degradation of the added watermark. The amount of watermark added is then adapted to make the watermark more robust against the accompanying degradation due to compression. For example the amount of watermark added is increased when there is greater degradation and decreased when the degradation is less. This is achieved with negligible loss in the quality of the *decoded* video sequence (compared to the case when the watermark is not adaptive) and without appreciably increasing the complexity of the coder. Degradation (or attack on the watermark due to compression) is measured in terms of the MSE of the decoded frame w.r.t. the *watermarked* original. The degradation of the most recently coded frame of the same type as the current frame is used to determine the amount of watermark that should be added when coding the current frame. Since the B frames are coded most aggressively they suffer most degradation and hence more watermark gets added to the B frames. Since the B frames are not used as reference frames the added watermark is not propagated. Also it is the quality of the *decoded* video frames that is of concern and we observed experimentally that the decoded video quality is effected very marginally since much of the added watermark is erased in the process of coding the B-frames. For the P frames the increase in the added watermark is lesser since their degradation is much less. We do not increase the amount of watermark added to the I frames. Note that the amount of watermark added is adaptive to the bit rate and coding complexity of the video sequence since these parameters decide the degradation of the decoded frames.

We present two schemes based on this concept. The first scheme adds the watermark to the 8 × 8 block DCT of the motion compensated frame difference (residual) for the P and B frames and the 8 × 8 block DCT of the I frames. This scheme is computationally very fast. The only additional computation required to adaptively add the watermark is computing the MSE of the decoded frames (note that the encoder already needs to create decoded frames to use as reference for motion compensation) w.r.t. the watermarked original. This computation can be carried out directly in the DCT domain (due to Parseval's Theorem). Since the DCT coefficients of the watermarked original and the coded frames are readily available in the encoder no additional IDCT have to be performed. Hence this scheme is computationally very fast.

The second scheme adds the watermark to the individual frames and not to the residual. This scheme however uses the wavelet transform for watermarking and hence is computationally more expensive.

The main ingredients of both the schemes have been taken from our previous work presented in [7].

2. THE FIRST SCHEME

Consider an MPEG-2 like compression scheme. Let V_{kn} denote the kth transform coefficient of the nth frame. For P and B frames this would be the 8 × 8 block-DCT coefficient of the residual after motion compensation. When watermarking the nth frame a *frame-sized* watermark x_{kn} with uniform distribution having zero mean and unit variance is generated. The watermark is i.i.d. over space and time. Depending on the contents of the video sequence arguments can be made against or in favor of generating independent watermarks to survive collusion attacks. We do not address this problem in this paper. For the purpose of this paper we generate the watermarks for different frames independently although the scheme is equally applicable otherwise.

Let T_1 and T_2 be two fixed thresholds. Let α be a fixed parameter controlling the minimum amount of watermark added and let γ be a variable parameter which increases the amount of watermark added according to the degradation of the decoded frame. The watermarked coefficient V'_{kn} is computed as:

$$V'_{kn} = V_{kn} + \alpha\gamma|V_{kn}|x_{kn} \qquad (1)$$

for all (k, n) for which $|V_{kn}| > T_1$. The coefficients at other locations and also the DC coefficient are left unchanged. The motivation for such a scheme was given in [7]. γ is chosen as:

$$\begin{aligned} \gamma &= 1.0 \text{ for I frms} && (2)\\ &= \text{MEDIAN}(1.0, 0.15 * rms_P, 1.25)\text{for P frms} && (3)\\ &= \text{MEDIAN}(1.0, 0.15 * rms_B, 1.50)\text{for B frms} && (4) \end{aligned}$$

where rms_i refers to the root mean square error of the most recently coded frame of type i w.r.t. the *watermarked* original. Its a measure of the compression attack suffered by a frame of that type. As mentioned before this computation can be done directly with the DCT coefficients available to the encoder. γ varies from frame to frame but is same for all coefficients of a given frame. The specific values shown in Eqs.(2)-(4) were determined experimentally. Since the B frames are not used as reference frames and since they are subject to aggressive compression (rms_B are expected to be higher than rms_P) γ is allowed to have higher values for B pictures compared to the P pictures.

Watermark detection consists of computing the correlation coefficient between decoded DCT coefficients whose absolute value is greater than $T_2 > T_1$ in the suspect video frame and the given watermark. This is compared to a threshold to judge the presence of the watermark. If the video frames do not start from the beginning synchronization between the given video frames and the watermark sequence will be required. Note that detection is done directly in the DCT domain and there is no need for full decoding.

Figure (a) shows the self and cross correlation coefficients for 60 decoded frames of the flower garden sequence. Here $\alpha = 0.15, T_1 = 20$ and $T_2 = 25$. We see that by using an adaptive scheme as proposed the self (auto) correlations have increased considerably (especially for the B frames) without an increase in the cross-correlations. This is really important for surviving any further attacks on the compressed video like recompression. Figure (b) shows the PSNR of the corresponding decoded frames with respect to the *unwatermarked* (no watermark) original frames. We see that the PSNR for decoded frames using the adaptive watermarking scheme is only marginally lower compared to the non-adaptive scheme. Moreover subjective tests have shown that there is no appreciable visual difference in the quality of decoded frames with and without addition of the watermark. This happens because in (1) the watermark added is modulated according to the strength of the DCT coefficient. Figure (c) shows the actual γ values used for each frame.

3. THE SECOND SCHEME

Here the watermark is added to the individual frames themselves and not to the residual frames. Also we use a three level DWT (with Daubechies 8-tap filter) of the frames to add the watermark. The low-low subband is not watermarked. It was found experimentally that much more watermark can be added in the wavelet transform domain without causing perceptual degradation compared to the DCT domain. Hence using the wavelet transform is a computationally fast alternative to using computationally demanding HVS model for adding the watermark in the DCT domain. Other details of this scheme are same as the first scheme. Note that this scheme requires a forward and inverse DWT for watermark casting and a forward DWT for watermark detection. However this scheme is expected to be more robust to attacks like recompressing the video with a different GOP structure in which case the residual frames would be different from those in the originally watermarked stream.

Figure (d) shows the cross and self correlations with this scheme. Here $\alpha = 0.10, T_1 = 20$ and $T_2 = 25$. The self correlation values are now much lower compared to those in Figure (a) since the watermark was added in the frames and not in the residual which actually gets coded. However the cross-correlation values are also much lower because there are a much larger number of coefficients satisfying the constraint $|V_{kn}| > T_2$ during detection of the watermark (since this constraint is applied to the frames and not the residual). It is clear from Figure (d) that the separation between the self and cross correlations is much more for the adaptive scheme than the non-adaptive ("same") scheme.

4. REFERENCES

[1] M. Swanson, B. Zhu, and A. H. Tewfik, "Multiresolution scene based video watermarking using perceptual models," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 540–550, May 1998.

[2] V. Darmstaedter, J. F. Delaigle, D. Nicholson, and B. Macq, "A block based watermarking technique for MPEG-2 signals: Optimization and validation on real digital tv distribution links," in *ECMAST*, Berlin, Germany, May 1998.

[3] J. Dittman, M. Stabenau, and R. Steinmetz, "Robust MPEG video watermarking technologies," in *Proc. ACM Multimedia*, Bristol, UK, Sept. 1998.

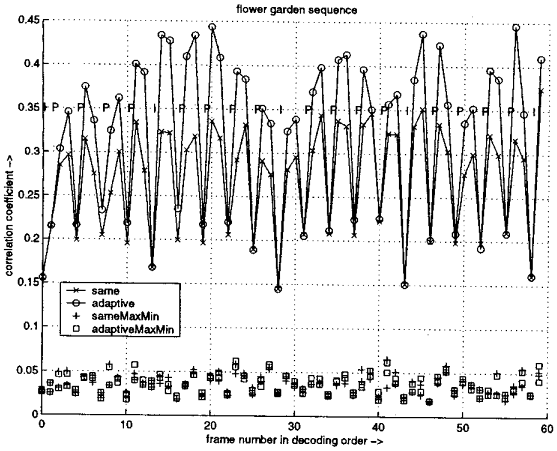
[4] F. Hartung and B. Girod, "Digital watermarking of raw and compressed video," in *Proc. SPIE Digital Compression Technologies and Systems for Video Commun.*, Oct. 1996, vol. 2952, pp. 205–213.

[5] G. C. Langelaar, R. L. Lagendijk, and J. Biemond, "Real time labeling methods for MPEG compressed video," in *Proc. 18th Symp. Information Theory in the Benelux*, Veldhoven, The Netherlands, May 1997.

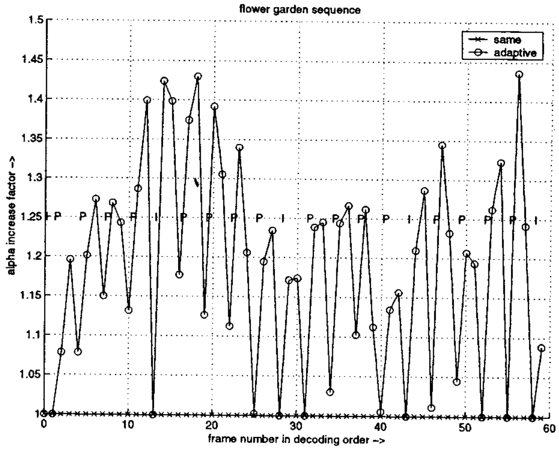
[6] J.-P. Linnartz, "MPEG PTY marking," Available at: http://diva.eecs.berkeley.edu/~linnartz/pty.html, 1998.

[7] Rakesh Dugad, Krishna Ratakonda, and Narendra Ahuja, "A new wavelet-based scheme for watermarking images," in *IEEE International Conference on Image Processing*, Chicago, Oct. 1998, vol. 2, pp. 419–423.

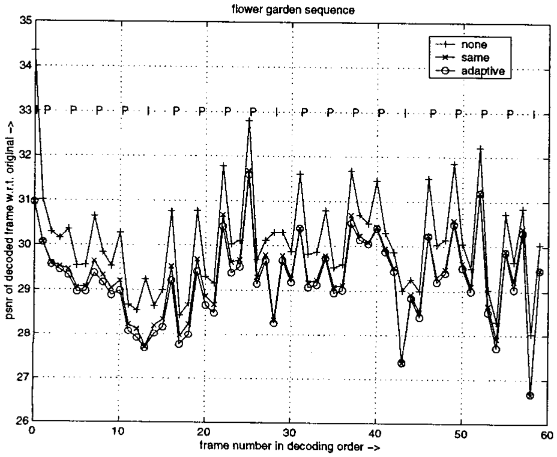
All results are shown for the 352 × 240 flower garden sequence coded at 3 Mb/s using the Software Simulation Group's MPEG-2 encoder (a) 'same' refers to the case when $\gamma = 1.0$ for I, P and B. 'adaptive' refers to the case when γ is given by Eqs.(2)-(4). MaxMin refer to the absolute values of the maximum and minimum cross correlations taken over a set of 100 watermarks independent from those added to the sequence. (b) 'none' refers to the case when no watermark is added. (c) γ values (d) same as (a) but for second scheme.


flower garden sequence
correlation coefficient ->
frame number in decoding order ->
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
10
20
30
40
50
60
I P P P P I P P P P I P P P P I P P P P I
same
adaptive
sameMaxMin
adaptiveMaxMin

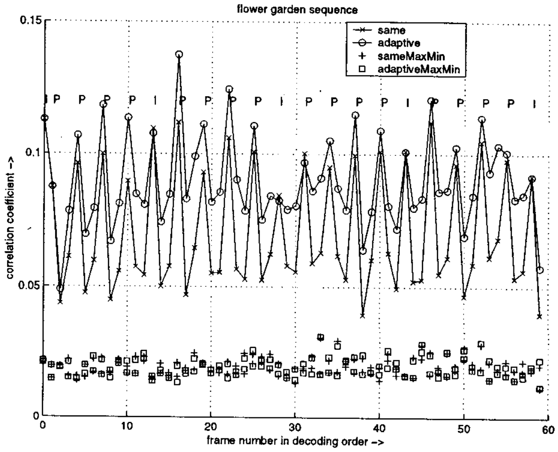
(a)


flower garden sequence
alpha increase factor ->
frame number in decoding order ->
1.5
1.45
1.4
1.35
1.3
1.25
1.2
1.15
1.1
1.05
1
0
10
20
30
40
50
60
I P P P P I P P P P I P P P P I P P P P I
same
adaptive

(c)


flower garden sequence
psnr of decoded frame w.r.t. original ->
frame number in decoding order ->
35
34
33
32
31
30
29
28
27
26
0
10
20
30
40
50
60
I P P P P I P P P P I P P P P I P P P P I
none
same
adaptive

(b)


flower garden sequence
correlation coefficient ->
frame number in decoding order ->
0.15
0.1
0.05
0
0
10
20
30
40
50
60
I P P P P I P P P P I P P P P I P P P P I
same
adaptive
sameMaxMin
adaptiveMaxMin

(d)