BLIND MPEG-2 VIDEO WATERMARKING ROBUST AGAINST SCALING

Yulin Wang and Alan Pearmain

Department of Electronic Engineering, Queen Mary, University of London

ABSTRACT

Blind watermarking techniques that need no original information during watermark detection are more desirable than informed ones for practical usage and convenience in watermark extraction. In this paper, a blind MPEG-2 watermarking technique which is robust against arbitrary ratio scaling is described. It can be directly applied in other block-DCT based video compression techniques. The main advantage of the scheme is its simplicity, blindness and adjustable watermark payload by trading-off with imperceptibility and robustness. Since the watermark extraction is based on self numerical references within frames of watermarked video, synchronization between the original and watermarked video is not necessary for detection.

1. INTRODUCTION

Geometric attacks generally not only result in geometric distortions but also incidental numerical loss. The widely cited spread-spectrum watermarking schemes are robust against noise and distortions, since their watermark detector usually employs statistical methods to extract or detect the watermark but synchronization is crucial for success. Many other watermarking schemes involve some cryptographic procedures, such as the use of secret information to pick or skip blocks for watermark detection. Any geometrical attacks plus secret block selection utilized in the embedding process may confuse the watermark detector.

Geometric processing of video, such as scaling, is very common in video applications, while for watermarking video, geometric distortions are generally more difficult to handle than numerical processing, therefore robustness against geometric attacks remains one of the most difficult areas of watermark research. Most suggested approaches fall into one of the following categories: synchronization [1,2,3], autocorrelation [4,5] and invariant watermarks [6,7,8]. One approach to obtain synchronization is exhaustive search, which entails inverting a large number of possible distortions, and testing for a watermark after each one. As the number of possible distortions increases, the computational cost and false positive probability become unacceptable. Synchronization patterns can be embedded in content to simplify the search. However, they introduce two failure modes: failure to correctly detect the registration pattern and failure to detect the watermark after registration. Techniques that rely on feature points that naturally occur in the image to obtain synchronization will spend significant time locating the feature points, both in the embedding and detection stages. Moreover, since not all of the media have the same number of feature points, this kind of technique can only be used for specific known media. The autocorrelation technique is similar to the pattern synchronization approach. It embeds extra data in addition to real watermark information to obtain synchronization for watermark detection, which further distorts the host media or sacrifices watermark payload.

Invariant watermarking embeds the watermark in a geometric-invariant transform, such as a log-polar Fourier transform, eliminating the need to identify and invert the specific geometric distortions, such as rotation, scaling and transformation. But this kind of technique is very fragile to slight geometric distortion, e.g. small-angle rotation or near-one ratio scaling. Moreover the computational cost to obtain the invariant domain from the heterogeneous transform, such as block DCT in MPEG-2, is too high. None of above techniques appear suitable and practical for watermarking video, especially compressed video.

Obtaining a watermark scheme that is robust against geometric attacks is very hard, while watermarking compressed video is more challenging than introducing a technique for images or raw video. Taking MPEG-2 video as an example, there is less redundancy in compressed video, and only the organized VLC codes of block DCT coefficients and macro block motion vectors are available. Any slight geometric processing of MPEG-2 video results in regenerating of the whole video, which in turn totally changes the organization/structure and almost all the fields in the bit stream, including the user fields or extra fields in the headers. This is why conventional block-based watermark techniques are extremely vulnerable to any geometrical attacks if no full knowledge of the location of the embedded watermark is provided for watermark detection.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

Focusing on robustness against video downsize scaling, we propose a novel blind watermarking technique for MPEG-2 video. Scaling video generally means applying the same scaling ratio to every individual frame. If we construct a "shadow" frame of one master frame with the same size, then after scaling the "shadow" frame will still synchronize with its master frame. This is also true for other geometric processing. The "shadow" frame can be generated by extrapolation and estimation by using adjacent previous frame(s) and past frame(s) of the master frame. Selecting a pair of full DCT coefficients in the same location from the master frame and its shadow frame, we can embed one watermark bit using a technique similar to [9]. Both the watermark embedding and detection of our technique are implemented in the frequency domain and only partial MPEG-2 decoding is needed. The embedded watermark is not only extremely robust against such uniform geometric attack as scaling, but also extremely robust against uniform numerical processing.

2. THE PROPOSED SCHEME

Although a scaling attack will totally destroy the bit stream structure of MPEG-2 video, including the number of 8x8 blocks, each full frame still exists if we considered it as one unit. The spatial scaling has roughly equivalent effect in the full DCT domain, shown in Fig. 1, truncating the upper-left corner of original frame to fill in the full DCT band of the downsized frame, or DC, AC_1 and other low frequency components remain unchanged, although there exists a slight difference between the different scaling. Since filtering and averaging may be involved depending on different scaling techniques, the full DCT values in the scaled frame may not be exactly the same as the counterpart of the original frame, but this does not affect our proposed technique. By exploiting this property, we can embed a watermark that is robust against scaling.

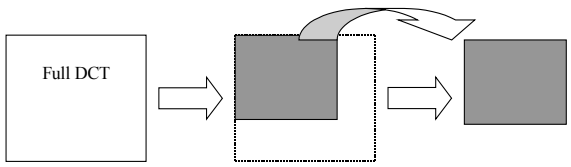


Fig. 1. Spatial downsizing is equivalent to cropping the upper left part of the original full DCT

Treating the full frame of the MPEG-2 video as a watermark unit, we need to obtain the full DCTs directly from the block DCTs in order to save computation cost. As we know, DCT is a kind of linear transform, and full

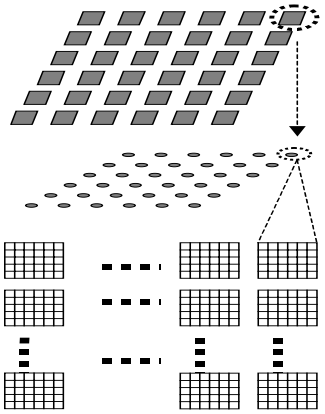


Fig. 2. Full DCT can be considered as two nest-loop DCTs, individual block DCT and "decimation" DCT

$$DCT_{full} = \sqrt{\frac{1}{LM}}.A_1 \begin{pmatrix} C_{0,0} & C_{0,1} & \dots & C_{0,M-1} \\ C_{1,0} & C_{1,1} & \dots & C_{1,M-1} \\ \dots & \dots & \dots & \dots \\ C_{L-1,0} & C_{L-1,1} & \dots & C_{L-1,M-1} \end{pmatrix} .A_2^T \qquad (1)$$

where,

$$A_1 = \begin{cases} \sqrt{\frac{1}{2}} A(u,i) & , u = 0, (i \bmod N) \neq 0 \\ \sqrt{2}\, A(u,i) & , u \neq 0, (i \bmod N) = 0 \\ A(u,i) & , otherwise \end{cases}$$

$$A_2 = \begin{cases} \sqrt{\frac{1}{2}} A(v,j) & , v = 0, (j \bmod N) \neq 0 \\ \sqrt{2}\, A(v,j) & , v \neq 0, (j \bmod N) = 0 \\ A(v,j) & , otherwise \end{cases}$$

$$A(u,i) = \cos\left(\frac{(2i+1)u\pi}{2LN}\right), \quad \text{therein} \quad u,i = 0,1,2,...LN$$

$$A(v,j) = \cos\left(\frac{(2j+1)v\pi}{2MN}\right), \quad \text{therein} \quad v,j = 0,1,2,...MN$$

DCT can be considered as containing nested two-level transforms, illustrated in Fig. 2. In the first level, we treat each 8x8 block as one "point", while in the second level each block is further DCT calculated using real points --- pixels. In [10], the authors present a fast algorithm to implement the calculation of sub-band DCT from full DCT, as in equation (1). The frame size is LN x MN. The size of matrix A_1, A_2 and DCT_{full} are LN x LN, MN x MN, LN x MN respectively. Each $C_{i,j}$ is a N x N block. For watermarking MPEG-2 video, we select N=8. For a specific size of video, A_1 and A_2 are constant matrixes for every frame, so we can further reduce the computation cost by a look-up table when converting between full DCT and block DCT.

2.1 Watermark embedding

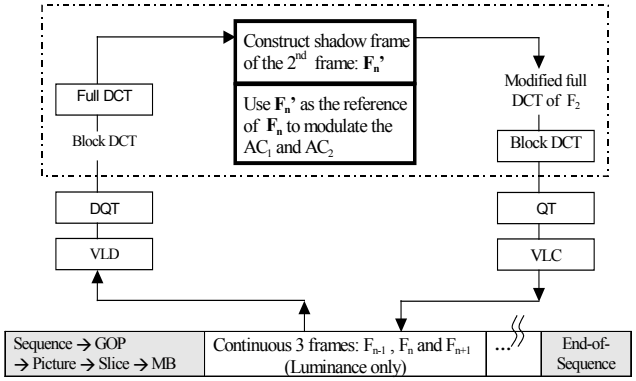


Fig. 3. Hiding watermark bits in MPEG-2 bit stream domain by grouping each 3 adjacent frames

Embedding is performed by partially decoding the MPEG-2 bit stream to the block DCT level to obtain block DCT coefficients of luminance, as shown in Fig. 3. After obtaining the block DCT of each frame, we can compose the full DCT directly from these block DCTs by using equation (1). Now, we consider the whole video as lots of groups, each containing 3 adjacent frames F_{n-1}, F_n and F_{n+1}. For each group, we construct a shadow frame F_n' of the middle frame F_n, similar to constructing a "tween frame" between F_{n-1} and F_{n+1}. Generally there exists linear relationship, especially low-frequency components, among the continuous adjacent frames, we generate F_n' by averaging F_{n-1} and F_{n+1}.

Using the selected full DCT coefficient AC_i in F_n' as reference, differentially modulate the same AC_i in F_n with equation (2) to embed one watermark bit. (DC components are not selected due to its sensitiveness to human eye). Since the difference between F_n frame and F_n' is very small, a small modulation depth Δ can provide reasonable robustness.

$$\begin{aligned} &\text{Set } AC_i \geq AC_i' + \Delta \quad \text{to embed bit '1'} \\ &\text{Set } AC_i \leq AC_i' - \Delta \quad \text{to embed bit '0'} \end{aligned} \tag{2}$$

Δ acts implicitly as the decision threshold for watermark bit detection. By adjusting its value we can make a trade-off between watermark imperceptibility and robustness against concomitant volumetric loss under geometric attacks. Δ is chosen as 10% *$|AC_i|$ in our experiment.

After the above steps, we get the full DCT of the watermarked frame Fn, and then we decompose the full DCT into 8x8 block DCTs by rearranging the equation (1), and writing back to the MPEG-2 bit stream, as shown in Fig. 3.

2.2. Extraction of watermark bit

After attacks the new regenerated video is still MPEG-2 format video (we only consider this kind of situation in this paper, which is very common), the frame of the new video is still multiples of 8x8 blocks. The watermark bit extraction from the watermarked MPEG-2 video is straightforward, only the composition of full DCT from block DCTs is needed, which is exactly the same as the first part of the watermark embedding work. After obtaining the full DCT coefficients of F_n' and F_n, we use equation (3) to extract each watermark bit. The Δ does not obviously appear in the expression for watermark bit extraction, but its value will indeed affect the percentage of the uncertainty of watermark bits.

$$\begin{aligned} &\text{If } AC_i > AC_i', \text{ then the extracted bit is '1'} \\ &\text{If } AC_i < AC_i', \text{ then the extracted bit is '0'} \\ &\text{If } AC_i = AC_i', \text{ uncertain} \end{aligned} \tag{3}$$

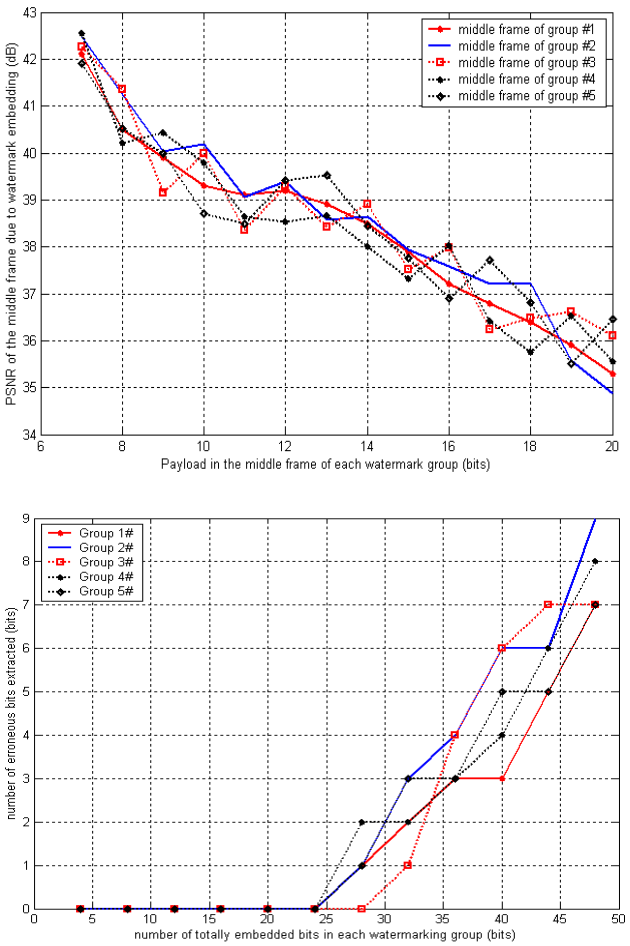


Fig. 4. PSNR vs. payload of each group, and payload vs. extraction error rate under no extra attacks (the first 5 watermark groups, or the first 15 frames)

3. EXPERIMENTAL RESULTS

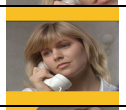
Geometric attacks		Error rate (%)
Type	parameters	
Aspect ratio conversion	4:3->16:9	0
	16:9->4:3	0
Arbitrary-ratio scaling	2.0:1	0
	2.5:1	0
	3.0:1	0
	4.0:1	0
	4.2:1	0
Arbitrary-angle rotation	0.1	0
	0.2	0
	1.5	7.5
	10	15.6
	90	20.6
Cropping upper-left 200x200 area		26.6
Padding boundary: Top=60, bottom=60		19.7
Skewing with 55 degrees		18.9
Non-geometric attacks		
Further decreasing bit rate	6Mbps->4Mbps	0
	6Mbps->3Mbps	0
	6Mbps->2Mbps	0
Format conversion	mpeg2<-> avi	0

Table. 1 Average error rate of watermark extraction in each watermarking group, payload =15bits/group

We used the CIF video SUSIE ON THE PHONE to test our technique, including perceptibility evaluated with PSNR, watermark payload and robustness against common video attacks. Selected experimental results are listed in Fig. 4 and table 1.

From figure 4, we can see that the more bits we embed in each group, the lower the PSNR. If the payload exceeds a threshold for the number of embedded bits, some of the bits will be extracted incorrectly, even without extra attacks. From table 1, we can see that this technique is extremely robust against scaling, but less robust to other geometric attacks.

4. CONCLUSION

A novel technique for watermarking MPEG-2 video based on "shadow-frame" generation is proposed. This is extremely robust against decreasing bit-rate by downsize scaling with an arbitrary ratio, as well as video format conversion and, which are common unintentional processing operations. Compared with other video watermarking techniques, the technique is computationally efficient. Only partial decoding of the MPEG-2 video plus the conversion between full DCT and block DCT is needed. The detection is straightforward and blind, which is important in real video applications.

5. REFERENCES

[1] Chih-Wei Tang and Hsueh-Ming Hang, "A feature-based robust digital image watermarking scheme, " *IEEE Transactions on Signal Processing*, vol. 51, pp 950 –959, April 2003

[2] Yafei Shao, Li Zhang, Guowei Wu and Xinggang Lin, "A novel frequency domain watermarking algorithm with resistance to geometric distortions and copy attack," *Proceedings of the 2003 International Symposium on Circuits and Systems, 2003*. vol. 2, pp II-940 -II-943, May 2003

[3] R. Lancini, F. Mapelli and S. Tubaro, "A robust video watermarking technique in the spatial domain, " *4th EURASIP-IEEE Region 8 International Symposium on Video/Image Processing and Multimedia Communications* pp 251 -256, June 2002

[4] Ping Dong and N.P. Galatsanos, "Geometric robust watermarking through watermark pattern shaping," *Proceedings of 2003 International Conference on Image Processing*, vol.1, pp 493 –496, 2003

[5] C.V. Serdean, M.A. Ambroze, M. Tomlinson, and J.G. Wade,"DWT-based high-capacity blind video watermarking, invariant to geometrical attacks," *IEE Proceedings of Vision, Image and Signal Processing*, vol.150, pp 51 –58, 2003

[6] Dong Zheng and Jiying Zhao, "Apply phase information in RST image watermarking," *IEEE International Conference on Consumer Electronics*, pp. 218 -219, 2003

[7] C.Y. Lin, M. Wu, J.A. Bloom, I.J. Cox, M.L. Miller and Y.M. Lui, "Rotation, scale, and translation resilient watermarking for images, " *IEEE Transactions on Image Processing*, vol. 10, pp. 767 –782, 2001

[8] M. Pawlak and Yongqing Xin, "Robust image watermarking: an invariant domain approach, " *IEEE Canadian Conference on Electrical and Computer Engineering, 2002*. vol. 2, pp. 885 - 888, 2002

[9] W. Nender, D. Gruhl, N. Morimoto and A. Lu, "Techniques for data hiding, "*IBM Systems Journal*, vol35, pp.313-336, 1996

[10] Jianmin Jiang and Guocan Feng, "The spatial relationship of DCT coefficients between a block and its sub-blocks," *IEEE Transactions on Signal Processing*, vol. 50 pp. 1160 –1169, 2002