A hybrid watermarking scheme for H.264/AVC video

Gang Qiu, Pina Marziliano, Anthony T.S. Ho
School of Electrical & Electronic Engineering
Nanyang Technological University
Nanyang Avenue, Singapore 639798.
{gqiu2002@pmail, epina@, etsho@}ntu.edu.sg

Dajun He, and Qibin Sun
Institute for Infocomm Research (I^{2}R)
21 Heng Mui Keng Terrace
Singapore 119613.
{djhe, qibin}@i2r.a-star.edu.sg

Abstract

A novel H.264/AVC watermarking method is proposed in this paper. By embedding the robust watermark into DCT domain and the fragile watermark into motion vectors respectively, the proposed method can jointly achieve both copyright protection and authentication. Our scheme outperforms other video watermarking schemes on higher watermarking capacity especially in lower compression bitrates. Furthermore, being well aligned with Lagrangian optimization for mode choice featured in H.264/AVC, the proposed scheme only introduces small distortions into the video content. Experimental results also demonstrate that the proposed solution is very computationally efficient during watermark extraction.

1. Introduction

The booming of digital video applications and the proliferation of powerful duplication / manipulation tools have created a urgent need to protect the owner's copyright and to prevent the content from tampering. Digital watermarking has been proposed recently to address these needs [1]. By embedding imperceptible signature into the video data, the watermarking not only can be used to identify the rightful owner, authorized distributors and intended recipients, but also to authenticate video contents.

In existing real-time video watermarking schemes, *DCT domain watermarking*[2] and *motion vector (MV) watermarking*[3][4] have drawn researchers' attention due to their low computation complexity and the compatibility with the video coding standards[1]. Nevertheless, each technique has its own advantages and disadvantages. DCT domain watermarking inserts watermark by modifying DCT coefficients. By adapting the embedding strength, it can be robust against some attacks such as bit-rate conversion transcoding and filtering. But it is restrained by its low payload in low bit-rate applications because much fewer nonzero DCT coefficients are available for embedding. Exploiting the MVs to carry the watermark has been proposed in [3][4] for compressed video. Since even the video compressed at very low bit-rate has many nonzero MVs, the watermarking capacity can be much higher than that in the DCT domain. But the watermark hidden in MVs is fragile to any manipulation such as filtering.

By combining the advantages of the above two methods, a novel watermarking approach for joint video copyright protection and authentication is presented in this paper. The copyright owner's signature is embedded into DCT coefficients for robust protection; the fragile watermark, which is able to detect malicious attacks, is inserted by the MV watermarking method. The fundamental idea of combining the robust watermark and the fragile watermark comes from [6], in which the embedding is performed by a modified cocktail watermarking method for image applications. The same idea is also utilized in [5] to construct a *semi-fragile* MPEG video authentication system.

The motivation for this paper is to provide more payload and strike a good rate-distortion balance for the watermarking of low bit-rate videos. While the state-of-the-art H.264/AVC video coding standard [7] is able to achieve a significant improvement in rate-distortion efficiency [9], the highly complex temporal and spatial prediction mechanisms involved make the watermarking more challenging. Therefore, it is our interest to develop and evaluate our watermarking scheme for the H.264/AVC standard.

2. The Proposed Scheme

2.1. Overview

In the proposed method, a watermark is inserted into H.264/AVC video stream during encoding process while the detection is performed during decoding. Fig. 1 shows the watermark embedding scheme. The watermark comprises the robust watermark W_R and the fragile watermark W_F.

Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)
1051-4651/04 $ 20.00 IEEE

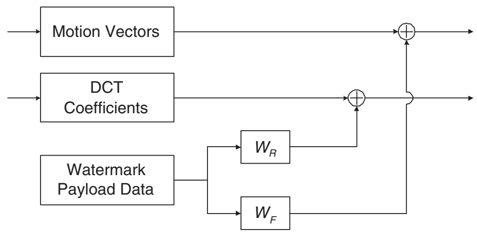


Figure 1. Watermark generation and insertion into both the DCT coefficients and the motion vectors. W_R denotes the robust watermark and W_F denotes the fragile watermark.

This hybrid approach is different from traditional methods in that the W_F watermark is embedded in MVs rather than in DCT domain. In general, the watermarking in DCT domain will degrade the video quality greatly, especially for low bit-rate applications. In [10], Alatter et al. notices that watermarking MPEG-4 video compressed at bit-rate 128kb/s will produce objectionable video quality degradation. With MV watermarking method, our method provides more payload without degrading video quality greatly. In addition, by adding some frame feature information in W_R [5], our scheme can be extended for semi-fragile authentication. While the method in [5] can only detect GOP-level attacks since both W_R and W_F are embedded into the intra-coded frames (or *I-frames*), our scheme is able to detect frame-level dropping or adding because the predictive-coded frames (or *P-frames*) also contain watermarks. Consequently, the proposed watermarking scheme is so flexible that it can achieve joint copyright protection and fragile/semi-fragile authentication at the same time.

To ease the description, this paper describes the proposed scheme in the context of joint copyright protection and fragile authentication. Thus, W_R is the copyright owner's signature and W_F is used for fragile authentication. They are embedded into quantized DCT coefficients of I-frames and MVs of P-frames, respectively.

2.2. Watermarking DCT coefficients

Two basic intra-prediction modes INTRA-4×4 and INTRA-16×16 are used in H.264/AVC, comprising 4×4 and 16×16 block-wise *directional spatial predictions* [9], in which the pixel value of the current block is predicted by the edge pixels of the adjacent blocks. Then, the prediction error is transformed primarily by a new 4×4 integer DCT instead of the float 8×8 DCT, which is widely used in existing standards. While the smaller block-size is justified by the advance of prediction capabilities by using above mentioned prediction modes [9], it makes the embedded watermark more sensitive to attacks or transcoding [8].

We insert watermark bits by altering the quantized AC coefficients of luminance blocks within I-frames. In order to survive the re-compression, two major obstacles are considered as follows.

First of all, the watermark signal $M(u, v)$ must be strong enough to survive the quantization, so that

$$|M_q(u, v)| = |\text{quant}[M(u, v), QP]| \geqslant 1 \quad (1)$$

where the $\text{quant}[\cdot]$ denotes the quantization operation, QP denotes the quantization parameter and (u, v) denotes a position in a 4×4 block $\boldsymbol{B}_k$. Obviously $M(u, v)$ should be even greater if the watermark is required to survive the re-quantization during transcoding.

Furthermore, since the change of the *prediction direction* (or *mode*) during transcoding may alter the value of DCT coefficients and thus leads to watermark detection error, we choose one of the quantized AC coefficients $X_q(u, v)$ in high frequency along the diagonal positions (i.e., $u = v$) for embedding. Our experiments show that the coefficients in diagonal positions are stabler than the others. Thus, the $X_q(u, v)$ is replaced by the watermarked coefficient X_q^*,

$$X_q^* = \begin{cases} \max\{X_q(u, v), M_q(u, v)\} & \text{if } w_n = 1 \\ 0 & \text{if } w_n = 0 \end{cases} \quad (2)$$

where w_n is the bit to be embedded. We notice the AC coefficient $X_q(u, v)$ is cleared if '0' is embedded. It can be justified by the fact that the $X_q(u, v)$ is zero in most cases. It will not introduce significant artifacts.

After watermarking, the best mode for a watermarked macroblock $\boldsymbol{S}_k^*$ is selected by minimizing the modified Lagrange optimization function:

$$J_{MODE} = D_{REC}(\boldsymbol{S}_k^*, \boldsymbol{I}_k) + \lambda_{MODE} R_{REC}(\boldsymbol{S}_k^*, \boldsymbol{I}_k) \quad (3)$$

where D_{REC} and R_{REC} represent the distortion and the number of bits, respectively, encoded for modes $\boldsymbol{I}_k \in$ {INTRA- 4×4, INTRA-16×16}. λ_{MODE} is Lagrange parameter.

2.3. Watermarking motion vectors

Here we describe how MV watermarking is performed during the motion estimation (ME) process.

In H.264/AVC, since individual MVs can be transmitted for blocks as small as 4×4, up to 16 MVs may be transmitted for a single macroblock[9]. For each block, the motion search proceeds first over integer-pixel positions and then the half-pixel locations around the best integer-pixel position. If motion compensation with quarter-pixel accuracy is used, an additional refinement search is performed among the quarter-pixels $\{1 \sim 8\}$ around the previous determined best half-pixel position A, as illustrated in Fig. 2. Quarter-pixel motion accuracy is assumed in this paper.

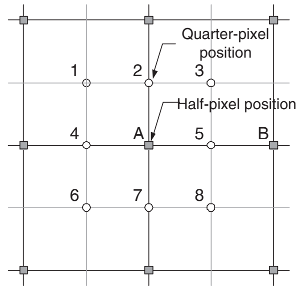


Figure 2. Fractional sample search positions.

During the above process, the criterion to find the optimum MV $\mathbf{m}_i$ is to minimize a Lagrangian cost function:

$$\mathbf{m}_i = \arg\min_{\mathbf{m}\in\mathcal{M}}\{D_{DFD}(\boldsymbol{S}_i, \mathbf{m}) + \lambda_{ME} R(\mathbf{m} - \mathbf{p})\} \quad (4)$$

where $\mathbf{m}$ denotes the current candidate MV, $\mathbf{p}$ denotes predicted MV estimated from neighboring blocks, λ_{ME} is the Lagrangian multiplier for motion estimation, D_{DFD} is used as a distortion measure, and the rate $R(\mathbf{m} - \mathbf{p})$ represents the number of bits required to encode the MV prediction error $\mathbf{d} = [d_x, d_y]^T$, which is the displacement between $\mathbf{m}$ and $\mathbf{p}$, i.e., $\mathbf{d} = \mathbf{m} - \mathbf{p}$.

Similar to [3], we insert the watermark into MVs by changing one of the components of a set of selected MVs. However, noting that it is the MV prediction error $\mathbf{d}$ which is to be encoded into video stream, in our method, we embed the watermark by changing the current MV $\mathbf{m}$ to make the MV prediction error $\mathbf{d}$ odd or even rather than modifying the MVs directly.

We use an example here for further clarification. Assume $w_n \in \{0, 1\}$ is the bit to be embedded into d_x, the $\mathbf{p}$ points at half-pixel location B, and the current best half-pixel MV is pointed at location A, shown in Fig. 2. Then, if $w_n = 1$, the current candidate MV $\mathbf{m}$ is selected from position {1, 3, 4, 5, 6, 8}; if $w_n = 0$, the $\mathbf{m}$ is selected from position {2, 7, A}. In other words, if '1' is to be embedded, the motion estimation process searches the positions where d_x gives odd value; the positions where d_x gives even value are searched if '0' is to be embedded.

The optimal MV after watermarking is selected in rate-distortion sense using equation (4). However, in order to take into account that the number of bits and the distortion introduced by the watermark, the equation (4) has to be rewritten as:

$$\mathbf{m}_i^* = \arg\min_{\mathbf{m}\in\mathcal{M}}\{D_{DFD}(\boldsymbol{S}_i, \mathbf{m}^*) + \lambda_{ME} R(\mathbf{m}^* - \mathbf{p})\} \quad (5)$$

where $\mathbf{m}^*$ denotes the watermarked MV. The above process can be seen as a motion estimation process constrained by the embedding rules. Finally, the residual block will be DCT transformed and quantized as usual.

Obviously, the above method simplifies the watermark extraction process and reduces memory consumption. The watermark bits embedded by our method can be read directly from the compressed video stream by analyzing the MV prediction errors. In contrast, in traditional MV watermark detection methods, MV prediction error must be firstly associated with predicted MVs in order to get the real MVs. Hence, all the decoded MVs of neighboring blocks need to be stored.

2.4. Detection Scheme

The watermark is extracted from the H.264/AVC compressed video stream by going through following steps:

- The bitstream is partially decoded to obtain the quantized DCT coefficients and the MV prediction errors.
- For each block containing MV prediction errors in the P-frames, the watermark bit $w_n = d_i \mod 2$;
- For each block containing DCT coefficients in the I-frames, the watermark bit $w_n = 1$ when the watermarked DCT coefficient $X_q(u, v) \geq 0$, otherwise $w_n = 0$.

3. Experiments and Results

We have integrated our proposed watermarking algorithm into H.264/AVC JM-7.3 reference software [11]. All experiments have been conducted on the video sequence *Foreman* (QCIF@15Hz) and *Container* (QCIF@10Hz). For both sequences, the I-frame refresh period is one I-frame per second. In each I-frame, one bit is inserted into a macroblock in order to spread the watermark bits over the whole frame. So totally 99 bits are embedded as copyright owner's signature in a frame of QCIF (176×144) resolution. In every P-frame, one bit is embedded to the horizontal component of a MV for MV watermarking.

Fig. 3 illustrates the coding efficiency loss results from watermarking. Both of the test sequences *Foreman* and *Container* are encoded with fixed quantization parameters $QP = [28, 32, 36, 40]$, corresponding to typical QPs for low bit-rate applications. While the watermarking causes the RD performance loss in both sequences, the loss in *Foreman* is more significant. It is because *Foreman* contains far more dynamic motions than *Container*. Since the number of MVs in a macroblock depends on the motion activities, more bits are embedded into the sequence *Foreman*, thus lead to bigger bit-rate.

The watermark embedded in MVs is fragile and easy to remove. This is demonstrated by a experiment which simply re-compresses the video with the same quantization parameters. On the other hand, the robustness of the watermark embedded in the DCT domain of the I-frames is evalu-

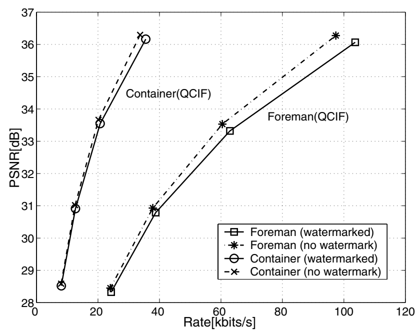
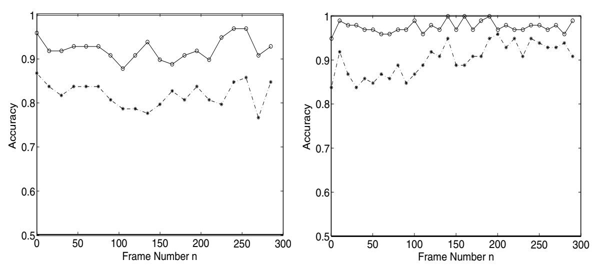


Figure 3. Rate-distortion curves for the H.264/AVC bit-streams with watermark and without watermark for the sequences *Foreman* and *Container* encoded with $QP = [28, 32, 36, 40]$.



(a) Foreman

(b) Container

Figure 4. The accuracy rate of watermark bits in DCT domain extracted after re-compression with quantization parameter $QP = 28$ ($\circ$) and $QP = 30$ ($*$), respectively. The original sequences are coded with $QP = 28$.

ated by transcoding with different quantization parameters. The results are shown in Fig. 4 in terms of bit accuracy rate, which is defined as the ratio of correct extracted bits among all embedded bits. In order to further improve the robustness, error correction code such as BCH code may be used on the top of information bits.

The robustness of watermarks depends on the DCT domain watermarking technique. While the smaller block-size of 4×4 integer DCT transform reduces the ringing artifacts, some researchers have noticed that the embedded watermark are more sensitive to various attacks [8]. We are working on this to further improve the robustness.

4. Conclusions

This paper presents a novel watermarking approach for H.264/AVC joint copyright protection and authentication. By combining the DCT domain watermarking technique and the MV watermarking technique, this approach achieves higher payload especially in low bit-rate video applications. It also consumes little computation and memory, which is necessary for real-time video applications. In addition, the optimal selection of prediction mode and motion vectors in H.264/AVC is supported by slightly modifying the Lagrangian optimization functions in order to curb the bit-rate increase. The experiment results show that our method is able to withstand transcoding and at the same time cause unnoticeable quality degradation.

References

[1] G. Doërr and J-L. Dugelay. A guide tour of video watermarking. *Signal Processing: Image Commun.*, 18:263-282, 2003.

[2] D. Simitopoulos, S.A. Tsaftaris, N.V. Boulgouris, and M.G. Strintzis. Compressed-domain video watermarking of MPEG streams. *Proc. IEEE Int. Conf. Multimedia and Expo*, 1:569-572, Aug. 2002.

[3] F. Jordan, M. Kutter, and T. Ebrahimi. Proposal of a Watermarking Technique for Hiding/Retrieving Datas in Compressed and Decompressed Video, ISO/IEC JTC1/SC29/WG11/MPEG97/M2281, 1997.

[4] J. Song and K.J.R. Liu. A data embedding scheme for H.263 compatible video coding. *Proc. IEEE Int. Symp. Circuits and Systems*, 4:390-393, June 1999.

[5] P. Yin and H.H. Yu. Semi-fragile watermarking system for MPEG video authentication. *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 4:3461-3464, 2002.

[6] C.-S. Lu, H.-Y.M. Liao, and C.-J. Sze. Combined watermarking for image authentication and protection. *Proc. IEEE Int. Conf. Multimedia and Expo*, 3:1415-1418, Aug. 2000.

[7] Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC). Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG JVT-G050, 2003.

[8] J. Zhang and A.T.S. Ho. An efficient digital image-in-image watermarking algorithm using the integer discrete cosine transform (IntDCT). *IEEE Joint Conf. of 4th Int. Conf. Info., Commun. and Signal Processing and 4th Pacific-Rim Conf. Multimedia*, Dec. 2003.

[9] T. Wiegang, G.J. Sullivan, G. Bjøntegaard, and A. Luthra. Overview of the H.264/AVC video coding standard. *IEEE Trans. Circuits Syst. Video Technol.*, 13(7):560-576, Jul. 2003.

[10] A.M. Alattar, E.T. Lin, and M.U. Celik. Digital watermarking of low bit-rate advanced simple profile MPEG-4 compressed video. *IEEE Trans. Circuits Syst. Video Technol.*, 13(8):787-800, Aug. 2003.

[11] H.264/AVC Joint Model 7.3 (JM-7.3). [On-line] Available: ftp://imtc.org/jvt-exports/reference_software/.