COMPRESSED-DOMAIN VIDEO WATERMARKING OF MPEG STREAMS

D. Simitopoulos[1], *S. A. Tsaftaris*[2], *N. V. Boulgouris*[1], *and M. G. Strintzis*[1]

[1]Informatics and Telematics Institute
1st Km Thermi-Panorama Road
Thermi-Thessaloniki, Greece

[2]Electrical and Computer Engineering Dept.
Northwestern University
Evanston, IL, 60208

ABSTRACT

A new technique for watermarking of MPEG compressed video streams is proposed. The watermarking scheme operates directly in the domain of MPEG program streams. Perceptual models are used during the embedding process in order to preserve the video quality. The watermark is embedded in the compressed domain and is detected without the use of the original video sequence. Experimental evaluation demonstrates that the proposed scheme is able to withstand a variety of attacks. The resulting watermarking system is fast and reliable, and is suitable for copyright protection and real-time content authentication applications.

1. INTRODUCTION

In parallel with the development and the introduction of Digital Versatile Disc (DVD) as the ultimate medium for the digital storage and distribution of audiovisual content, the MPEG-2 standard was established as the coding scheme for such content. These developments made the large-scale distribution and replication of multimedia very easy but at the same time also to a large extent uncontrollable. In order to protect multimedia content from unauthorized trading, many digital watermarking techniques have been introduced. However, few of them deal with the very important issue of compressed domain watermarking for video [1, 2].

In most watermarking systems the watermark is required to be imperceptible and robust against attacks such as compression, cropping, filtering, etc [3]. Apart from the above, video watermarking systems have additional requirements, such as fast embedding and detection, blind detection and file size preservation after the watermark is embedded.

In this paper, a novel compressed domain watermarking scheme is presented which is suitable for MPEG multiplexed streams (video and audio). Embedding and detection are performed without fully de-multiplexing the video stream. During the embedding process, the data that are going to be watermarked are extracted from the stream. After perceptual analysis and block classification is performed, the data are watermarked and placed back into the stream. This approach leads to a fast implementation which is necessary for real-time applications and also when a large number of video-sequences have to be watermarked, as is the case in video libraries. The detection is so fast that it can be incorporated to real-time content authentication systems. The resulting watermarked video sequences are shown to withstand transcoding, as well as cropping and filtering.

2. PERCEPTUAL WATERMARKING IN THE COMPRESSED DOMAIN

2.1. Perceptual embedding

The proposed watermark embedding scheme (see Fig. 1) alters only the quantized AC coefficients $X_Q(m,n)$ of luminance blocks (where m, n are indices indicating the position of the current coefficient in an 8×8 DCT block) that belong in intra frames. In order to make the watermark as imperceptible as possible, a novel method is employed, combining perceptual analysis [3, 4] and block classification techniques [5, 6]. These are applied in the DCT domain in order to adaptively select which coefficients are best for watermarking. The watermark coefficients $W(m,n)$ are the values of a pseudo-random sequence of ± 1 that is created as in [7]. The product of $W(m,n)$ with the corresponding parameters of the quantized embedding mask $M_Q(m,n)$ and the classification mask $C(m,n)$ (which result from the perceptual analysis and the block classification process respectively), is added to each selected quantized coefficient. The resulting watermarked quantized coefficient $X'_Q(m,n)$ is given by:

$$X'_Q(m,n) = X_Q(m,n) + C(m,n)M_Q(m,n)W(m,n) \quad (1)$$

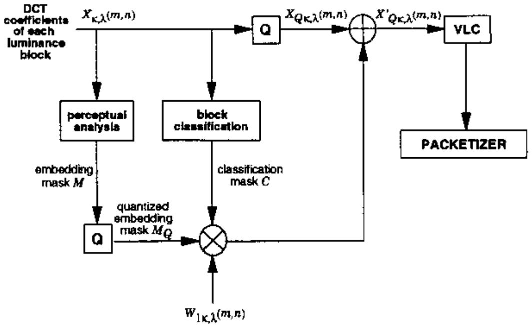


Fig. 1. Watermark embedding scheme.

Initially, each discrete cosine transformed (DCT) luminance block is classified with respect to its energy distribution to one of five possible classes: *low activity, diagonal edge, horizontal edge, vertical edge* and *textured block*. The calculations of the energy distribution and the block classification are performed as in [6].

This procedure returns the class of the examined block. The binary mask values $C(m,n)$ corresponding to each class indicate the best coefficients to be altered (with an additive watermark

---

This work was largely completed while the second author was with Informatics and Telematics Institute.

0-7803-7304-9/02/$17.00 C2002 IEEE

whose strength is estimated by the perceptual analysis described in the sequel) without reducing the visual quality

$$C(m,n) = \begin{cases} 0 & \text{no alteration} \\ 1 & \text{alteration} \end{cases}$$

where $m, n \in [0, 7]$. For all block classes apart from the *low activity* class, the binary mask C is one of those depicted in Fig. 2. In the case of *low activity* blocks the binary mask contains "ones" for all AC coefficients.

$$\begin{bmatrix} 0&0&0&0&0&0&0&0 \\ 1&0&0&0&0&0&0&0 \\ 1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1 \end{bmatrix} \qquad \begin{bmatrix} 0&1&1&1&1&1&1&1 \\ 0&0&1&1&1&1&1&1 \\ 0&0&1&1&1&1&1&1 \\ 0&0&1&1&1&1&1&1 \\ 0&0&1&1&1&1&1&1 \\ 0&0&1&1&1&1&1&1 \\ 0&0&1&1&1&1&1&1 \\ 0&0&1&1&1&1&1&1 \end{bmatrix}$$

(a) Vertical edge mask (b) Horizontal edge mask

$$\begin{bmatrix} 0&1&1&1&1&1&1&1 \\ 1&1&0&1&1&1&1&1 \\ 1&0&0&0&1&1&1&1 \\ 1&1&0&0&0&1&1&1 \\ 1&1&1&0&0&0&1&1 \\ 1&1&1&1&0&1&1&1 \\ 1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1 \end{bmatrix} \qquad \begin{bmatrix} 0&0&0&0&1&1&1&1 \\ 0&0&0&1&1&1&1&1 \\ 0&0&1&1&1&1&1&1 \\ 0&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1 \end{bmatrix}$$

(c) Diagonal edge mask (d) Textured block mask

Fig. 2. Block classification masks

The perceptual model that will be used is a novel adaptation of the perceptual model proposed by Watson [4]. Specifically, a measure $T''(m,n)$ is introduced which determines the maximum Just Noticeable Difference (JND) for each DCT coefficient of a block and then this model is adapted in order to be applicable to the domain of *quantized* DCT coefficients.

For 1/16 *pixels/degree* of visual angle and 48.7 *cm* viewing distance, the *luminance masking* and the *contrast masking* properties of the Human Visual System (HVS) for each coefficient of a discrete cosine transformed block are estimated as in [4]. Specifically, two matrices, T' (*luminance masking*) and T'' (*contrast masking*) are calculated. Each one of the values $T'(m,n)$ is compared with the absolute value of each DCT coefficient $|X(m,n)|$. In this manner, they are used as thresholds in order to decide which coefficients to watermark. The values $T''(m,n)$ determine the embedding strength of the watermark. The embedding mask M contains the values $T''(m,n)$ for the coefficients that exceed the $T'(m,n)$ thresholds and zeroes for the remaining coefficients

$$M(m,n) = \begin{cases} T''(m,n), & \text{if } |X(m,n)| > T'(m,n) \\ 0, & \text{otherwise} \end{cases}$$

In order to achieve efficient perceptual embedding in the quantized domain, the *quantized* values $M_Q(m,n)$ of the perceptual mask values $M(m,n)$ are used as the embedding strength of the watermark, as explained in the ensuing Section 2.2.

2.2. Quantized domain embedding

Following the perceptual analysis described above, two options exist for the watermark embedding. The watermark will be added either to the DCT coefficients, or to the quantized DCT coefficients. We shall first examine in detail the case where the watermark coefficient $W(m,n)$ is embedded in the DCT coefficients $X(m,n)$, before quantization is applied; then the watermark embedding equation is given by:

$$X'(m,n) = X(m,n) + C(m,n)M(m,n)W(m,n) \quad (2)$$

Then, the MPEG coding algorithm quantizes the watermarked DCT coefficients $X'(m,n)$ using the *quant* function

$$quant\,[x(m,n)] = round\left(\frac{8x(m,n)}{q_s Q\,(m,n)}\right) \quad (3)$$

where $Q\,(m,n)$ denotes the (m,n) element of the quantization matrix used by MPEG [8] and q_s is the quantizer scale parameter (ranging from 1 to 31) that is selected by the MPEG encoding algorithm during the rate-control process in order to achieve a specific target bitrate for the entire video sequence. The *round* function performs rounding to the closest integer. The quantizer scale q_s has the same value for all DCT coefficients of a 8×8 block.

Similarly, the inverse mapping from quantized coefficients to DCT values is given by

$$quant^{-1}\,[x(m,n)] = \left\lfloor \frac{x(m,n)q_s Q\,(m,n)}{8} \right\rfloor \quad (4)$$

where the $\lfloor\cdot\rfloor$ operator denotes downward truncation.

The *quant* and $quant^{-1}$ functions in the above equations correspond to MPEG-1 quantization. MPEG-2 quantization is performed in the same way but the value 8 of the denominator is changed to 16 in both equations (3) and (4). Obviously, all analysis in the ensuing sections applies (with trivial and obvious alterations) to both MPEG-1 and MPEG-2 videos.

Using (2) and (3), the quantized watermarked coefficients are found by:

$$quant\left[X'(m,n)\right] = round\left(\frac{8X'(m,n)}{q_s Q\,(m,n)}\right) =$$

$$= round\left(\frac{8X(m,n)}{q_s Q\,(m,n)} + \frac{8C(m,n)M(m,n)W(m,n)}{q_s Q\,(m,n)}\right) \quad (5)$$

The decimal part of both fractions of equation (5) is uniformly distributed in [0,1). In the frequent case where

$$\left|\frac{8C(m,n)M(m,n)W(m,n)}{q_s Q\,(m,n)}\right| < 1 \quad (6)$$

as is easily seen, there is a 50% probability that the second term of (5), which contains the watermark, vanishes altogether and the right-hand part of (5) simply yields

$$round\left(\frac{8X(m,n)}{q_s Q\,(m,n)} + \frac{8C(m,n)M(m,n)W(m,n)}{q_s Q\,(m,n)}\right) =$$

$$= round\left(\frac{8X(m,n)}{q_s Q\,(m,n)}\right)$$

which is identical to the quantized value as if no watermark had been embedded. Therefore, it is clear that if (6) is valid the embedded watermark may be entirely eliminated by the quantization process. More generally, it is clear that the damage to the watermark may be very severe, and that potentially, the watermark detection process may become unreliable.

Thus, in order to avoid reduced detection performance due to MPEG quantization, the second option will henceforth be employed where the watermark is embedded in the quantized DCT coefficients. Since the MPEG coding algorithm performs no other lossy operation after quantization (see Fig. 3), any information embedded as in Fig. 3 does not run the risk of being eliminated by the subsequent processing. Thus, the watermark exists intact in the quantized coefficients when the detection process is carried out and the quantized DCT coefficients $X_Q(m,n)$ are watermarked in the following way (see Fig. 1)

$$X'_Q(m,n) = X_Q(m,n) + C(m,n)M_Q(m,n)W(m,n) \quad (7)$$

where $M_Q(m,n)$ is calculated by

$$M_Q(m,n) = quant\left[M(m,n)\right] = round\left(\frac{8M(m,n)}{q_s Q(m,n)}\right) \quad (8)$$

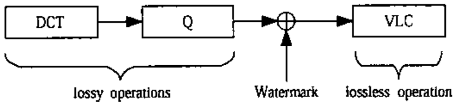
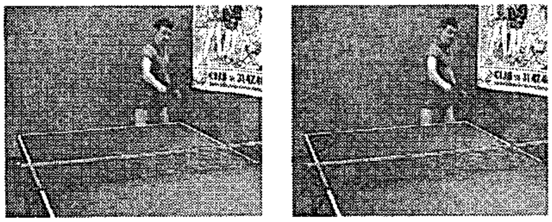


Fig. 3. MPEG encoding operations.

Whenever the value of $M(m,n)$ is non-zero but the value of $M_Q(m,n)$ becomes equal to zero due to the quantization, in order to increase the number of watermarked coefficients, we set $M_Q(m,n) = 1$ and the watermark coefficient $W(m,n)$ is embedded with its initial strength, which is equal to 1. In such a case, the corresponding watermark strength in the DCT domain ends up being higher than the strength allowed by the perceptual model. However, our experiments have shown that this modification is rare enough so as not to degrade the visual quality of the watermarked video frames. Fig. 4 depicts a frame from the video sequence *table tennis* and the corresponding watermarked frame.



(a) (b)

Fig. 4. (a) Original frame from the video sequence *table tennis*, (b) Watermarked frame.

Depending on the sign of the watermark coefficient $W(m,n)$ and on the values of the perceptual and block classification masks, the absolute value of $X'_Q(m,n)$ in equation (7) may be increased, decreased or may remain unchanged in relation to $|X_Q(m,n)|$. Due to the monotonicity of MPEG codebooks, when $|X'_Q(m,n)| > |X_Q(m,n)|$ the codeword used for $X'_Q(m,n)$ contains more bits than the corresponding codeword for $X_Q(m,n)$, and the opposite is true when $|X'_Q(m,n)| < |X_Q(m,n)|$. Since the watermark sequence has zero mean, the number of the cases where $|X'_Q(m,n)| > |X_Q(m,n)|$ is expected to roughly equal the number of the cases where the inverse inequality holds. Therefore, the MPEG bitstream length is not expected to be significantly altered. Experiments with watermarking of various MPEG-2 bitstreams resulted in bitstreams slightly larger (0-2%) than the original. The appearance of such slightly larger bitstreams is related to the *run-level* based variable length coding of MPEG [5] and the way it is affected by the watermarking process.

As mentioned above, when the *level* (i.e. the absolute value) of a quantized DCT coefficient is increased by the addition of the watermark, existing codewords of the bitstream are replaced by longer codewords. On the other hand, when the level of a quantized coefficient is decreased, there is a possibility that the coefficient may be set to zero. This results in the elimination of the codeword corresponding to the zeroed coefficient from the bitstream and in an increase of the *run* (the number of zero coefficients preceding a non-zero coefficient, in the scan order) for the next coded coefficient, which leads in a longer codeword. Therefore, it is reasonable to assume that decreases of the level that do not result in setting a coefficient to zero are generally compensated by increases of the level in other coefficients. For the cases where quantized DCT coefficients are set to zero, the effect on the bitstream length is undefined and this may lead to leaving some increases of the level uncompensated. However, the coefficients that are set to zero due to the watermark embedding are very few because block classification and perceptual analysis generally do not permit small coefficients to be watermarked or to be assigned a large watermark value. This explains the slight increase in the bitstream length that was observed.

In order to ensure that the length of the watermarked bitstream will remain smaller than or equal to the original bitstream, the coefficients that increase the bitstream length may be left unwatermarked. This will, however, reduce the robustness of the detection scheme because the watermark can be inserted and therefore detected in fewer coefficients. For this reason, such a modification was avoided in our embedding scheme.

3. DETECTION

The detection can be formulated as the following hypothesis test:

H_0: the video sequence is not watermarked

H_1: the video sequence is watermarked

In order to determine which of the above hypotheses is true, a correlation-based detection scheme as in [9] is applied. Variable length decoding is first performed to obtain the quantized DCT coefficients of intra frames. Then, inverse quantization provides the DCT coefficients for each block. The block classification and perceptual analysis procedures are performed as described in Section 2 in order to define the set $\{X\}$ of the N DCT coefficients that are expected to be watermarked with the sequence W.

Each coefficient in the set $\{X\}$ is multiplied by the corresponding watermark coefficient of the correlating watermark sequence $W(m,n)$ producing the data set $\{X_W\}$. The statistical characteristics (mean and variance) of the data set $\{X_W\}$ are calculated as follows

$$mean = E\{X_W\} = \frac{1}{N}\sum_{l=0}^{N-1} X_W(l) \quad (9)$$

$$variance = E\left\{(X_W - mean)^2\right\} = \frac{1}{N}\sum_{l=0}^{N-1}(X_W(l) - mean)^2 \qquad (10)$$

Finally, the statistical correlation metric c for each frame is calculated as

$$c = \frac{mean \cdot \sqrt{N}}{\sqrt{variance}} \qquad (11)$$

The correlation metric c is compared to the threshold T_c, which is an adaptive threshold calculated for each frame of the video sequence as in [9]. If the correlation metric c exceeds the threshold T_c, the examined frame is considered watermarked.

4. EXPERIMENTAL RESULTS

A software simulation of the proposed embedding algorithm was implemented and executed using a Pentium III (800 MHz) processor. All experiments were conducted using the standard video-only sequence *table tennis* (PAL resolution, 8 Mbit/sec, 15sec, 375 frames, 32 I-frames). The total execution time of the embedding and detection scheme is 72% and 23% of the real-time duration of the video sequence respectively, thus allowing the incorporation of the detector in real-time decoders/players.

Two correlation metric curves and the threshold curve for 32 I-frames of the MPEG-2 sequence *table tennis* are shown in Fig. 5. The upper correlation metric curve corresponds to the detector output when detection is performed using the copyright owner's watermark W, while the lower curve shows the correlation metric for the watermarked videosequence when a random watermark sequence W' is used for the detection. As seen, using the proposed watermarking system, the actual copyright owner can be clearly identified since watermarks provided by others that claim copyright ownership do not correlate with the content.

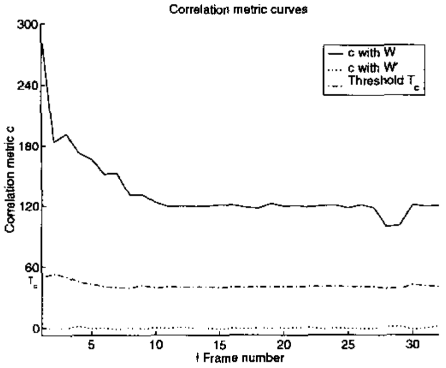


Fig. 5. Detector output for 32 I-frames of the *table tennis* MPEG-2 stream when the owner's watermark W and a random watermark sequence W' is used for the detection.

The robustness of the embedded watermark in the case of common video processing attacks was also tested. Table 1 shows the correlator output for the 15th I-frame of the *table tennis* video sequence when the owner's watermark W and a false watermark W' is used.

Attack	W	W'	Threshold
Original - No attack	121.1	-0.78	40.1
Blurring	67.7	-0.64	37.8
20% Clipping	71.2	-0.03	28.0
Transcoding to 4 Mbit/sec	75.3	-0.36	33.5

Table 1. Correlator output results for watermark detection on the 15th I-frame (frame 168) of the MPEG-2 *table tennis* sequence.

5. CONCLUSION

A novel and robust way for embedding watermarks in MPEG multiplexed streams was presented. The proposed scheme operates directly in the compressed domain and is able to embed copyright information without causing noticeable degradation to the quality of the video. Due to its speed, the resulting system is suitable for real-time content authentication applications. Experimental evaluation showed that the proposed watermarking scheme is able to withstand a variety of attacks.

6. REFERENCES

[1] G. C. Langelaar, R. L. Lagendijk, and J. Biemond, "Real-time labeling of MPEG-2 compressed video," *J. Visual Commun. Image Representation*, vol. 9, no. 4, pp. 256–270, Dec. 1998.

[2] F. Hartung and B. Girod, "Watermarking of Uncompressed and Compressed Video," *Signal Processing*, vol. 66, no. 3, pp. 283–301, May 1998.

[3] R. B. Wolfgang, C. I. Podilchuk, and E. J. Delp, "Perceptual watermarks for digital images and video," *Proc. IEEE*, vol. 87, no. 7, Jul. 1999.

[4] A. B. Watson, "DCT quantization matrices visually optimized for individual images," in *Proc. SPIE Conf. Human Vision, Visual Processing and Digital Display IV*, Feb. 1993, vol. 1913, pp. 202–216.

[5] K. R. Rao and J. J. Hwang, *Techniques and standards for Image, Video and Audio coding*, Prentice Hall PTR, 1996.

[6] T.-Y. Chung, M.-S. Hong, Y.-N. Oh, D.-H. Shin, and S.-H. Park, "Digital watermarking for copyright protection of MPEG-2 compressed video," *IEEE Trans. Consumer Electronics*, vol. 44, no. 3, Aug. 1998.

[7] D. Simitopoulos, S. A. Tsaftaris, N. V. Boulgouris, and M. G. Strintzis, "Digital watermarking of MPEG-1 and MPEG-2 multiplexed streams for copyright protection," in *Proc. Second USF International Workshop on Digital and Computational Video (DCV'01)*, Feb. 2001.

[8] ISO/IEC 13818-2, *Information technology - Generic coding of moving pictures and asscociated audio: Video.*

[9] W. Zeng and B. Liu, "A Statistical Watermark Detection Technique Without Using Original Images for Resolving Rightful Ownerships of Digital Images," *IEEE Trans. Image Processing*, vol. 8, no. 11, pp. 1534–1548, Nov. 1999.