

Fast Watermarking of MPEG-1/2 Streams Using Compressed-Domain Perceptual Embedding and a Generalized Correlator Detector

Dimitrios Simitopoulos

Information Processing Laboratory, Electrical and Computer Engineering Department, Aristotle University of Thessaloniki, 54006 Thessaloniki, Greece

Informatics and Telematics Institute, Centre for Research and Technology Hellas, 1st Km Thermi-Panorama Road, 57 001 Thermi-Thessaloniki, Greece
Email: dsim@iti.gr

Sotirios A. Tsiftaris

Electrical and Computer Engineering Department, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA
Email: s-tsiftaris@northwestern.edu

Nikolaos V. Boulgouris

The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, ON, Canada M5S 3G4
Email: nikos@comm.toronto.edu

Alexia Briassouli

Beckman Institute, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
Email: briassou@ifp.uiuc.edu

Michael G. Strintzis

Information Processing Laboratory, Electrical and Computer Engineering Department, Aristotle University of Thessaloniki, 54006 Thessaloniki, Greece
Email: strintzi@eng.auth.gr

Informatics and Telematics Institute, Centre for Research and Technology Hellas, 1st Km Thermi-Panorama Road, 57 001 Thermi-Thessaloniki, Greece

Received 9 January 2003; Revised 18 September 2003; Recommended for Publication by Ioannis Pitas

A novel technique is proposed for watermarking of MPEG-1 and MPEG-2 compressed video streams. The proposed scheme is applied directly in the domain of MPEG-1 system streams and MPEG-2 program streams (multiplexed streams). Perceptual models are used during the embedding process in order to avoid degradation of the video quality. The watermark is detected without the use of the original video sequence. A modified correlation-based detector is introduced that applies nonlinear preprocessing before correlation. Experimental evaluation demonstrates that the proposed scheme is able to withstand several common attacks. The resulting watermarking system is very fast and therefore suitable for copyright protection of compressed video.

Keywords and phrases: MPEG video watermarking, blind watermarking, imperceptible embedding, generalized correlator detector.

1. INTRODUCTION

The compression capability of the MPEG-2 standard [1, 2] has established it as the preferred coding technique for audiovisual content. This development, coupled with the ad-

vent of the digital versatile disc (DVD), which provides enormous storage capacity, enabled the large-scale distribution and replication of compressed multimedia, but also rendered it largely uncontrollable. For this reason, digital watermarking techniques have been introduced [3] as a way to

protect the multimedia content from unauthorized trading. Watermarking techniques aim to embed copyright information in image [4, 5, 6, 7], audio [8], or video [9, 10, 11] signals so that the lawful owner of the content is able to prove ownership in case of unauthorized copying. A variety of image and video watermarking techniques have been proposed for watermark embedding and detection in either the spatial [12, 13], Fourier-Mellin transform [14], Fourier Transform [15], discrete cosine transform (DCT) [4, 16], or wavelet [17] domain. However, only a small portion of them deal with video watermarking in the compressed domain [9, 13, 18, 19].

In [13] a technique was proposed that partially decompresses the MPEG stream, watermarks the resulting DCT coefficients, and reencodes them into a new compressed bitstream. However the detection is performed in the spatial domain, requiring full decompression. Chung et al. [19] applied a DCT domain-embedding technique that also incorporates a block classification algorithm in order to select the coefficients to be watermarked. In [18], a faster approach was proposed, that embeds the watermark in the domain of quantized DCT coefficients but uses no perceptual models in order to ensure the imperceptibility of the watermark. This algorithm embeds the watermark by setting to zero some DCT coefficients of an 8×8 DCT block. The embedding strength is controlled using a parameter that defines the smallest index of the coefficient in an 8×8 DCT block which is allowed to be removed from the image data upon embedding the watermark. However, no method has been proposed for the automatic selection of the above parameter so as to ensure perceptual invisibility of the watermark. In addition, in [9, 18], this parameter has a constant value for all blocks of an image, that is, it is not adapted to the local image characteristics in any way.

The important practical problem of watermarking MPEG-1/2 multiplexed streams has not been properly addressed in the literature so far. Multiplexed streams contain at least two elementary streams, an audio and a video elementary stream. Thus, it is necessary to develop a watermarking scheme that operates with multiplexed streams as its input.

In this paper, a novel compressed domain watermarking scheme is presented, which is suitable for MPEG-1/2 multiplexed streams. Embedding and detection are performed without fully demultiplexing the video stream. During the embedding process, the data to be watermarked, are extracted from the stream, watermarked, and placed back into the stream. This leads to a fast implementation, which is necessary for real-time applications, such as video servers in video on demand (VoD) applications. Implementation speed is also important when a large number of video sequences have to be watermarked, as is the case in video libraries.

The watermark is embedded in the intraframes (I-frames) of the video sequence. In each I-frame, only the quantized AC coefficients of each DCT block of the luminance component are watermarked. This approach leads to very good resistance to transcoding. In order to reach a satisfactory tradeoff between robustness and imperceptibility of the embedded watermark, a novel combination of perceptual

analysis [20] and block classification techniques [21] is introduced for the selection of the coefficients to be watermarked and for the determination of the watermark strength. Specifically block classification leads to an initial selection of the coefficients of each block that may be watermarked. In each block, the final coefficients are selected and the watermark strength is calculated based on the perceptual analysis process. In this way, watermarks having the maximum imperceptible strength are embedded into the video frames. This leads to a maximization of the detector performance under the watermark invisibility constraint.

A new watermark detection strategy in the present paper operates in the DCT domain rather than the quantized domain. Two detection approaches are presented. The first uses a correlation-based detector, which is optimal when the watermarked data follow a Gaussian distribution. The other, which is optimal when the watermarked data follow a Laplacian distribution, uses a generalized correlator, where the data is preprocessed before correlation. The preprocessing is nonlinear and leads to a locally optimum (LO) detector [22, 23], which is often used in communications [24, 25, 26] to improve the detection of weak signals.

The resulting watermark detection scheme is shown to withstand transcoding (bitrate change and/or coding standard change), as well as cropping and filtering. It is also very fast and therefore suitable for applications where watermark detection modules are incorporated in real-time decoders/players, such as broadcast monitoring [27, 28].

The paper is organized as follows. In Section 2, the requirements of a video watermarking system are analyzed. Section 3 describes the processing in the compressed stream. The proposed watermark embedding scheme is presented in Section 4. In Section 5 the detection process is described, and in Section 6 two implementations of watermark detectors for video are presented. In Section 7 experimental results are discussed, and finally, conclusions are drawn in Section 8.

2. VIDEO WATERMARKING SYSTEM REQUIREMENTS

In all watermarking systems, the watermark is required to be imperceptible and robust against attacks such as compression, cropping, filtering [7, 10, 29], and geometric transformations [14, 30]. Apart from the above, compressed video watermarking systems have the following additional capability requirements.

- (i) *Fast embedding/detection.* A video watermarking system must be very fast due to the large volume of data that has to be processed. Watermark embedding and detection procedures should be efficiently designed in order to offer fast processing times using a software implementation.
- (ii) *Blind detection.* The system should not use the original video for the detection of the watermark. This is necessary not only because of the important concerns raised in [29] about using the original data in the detection process, but also because it is sometimes impractical to keep all original sequences in addition to the watermarked ones.

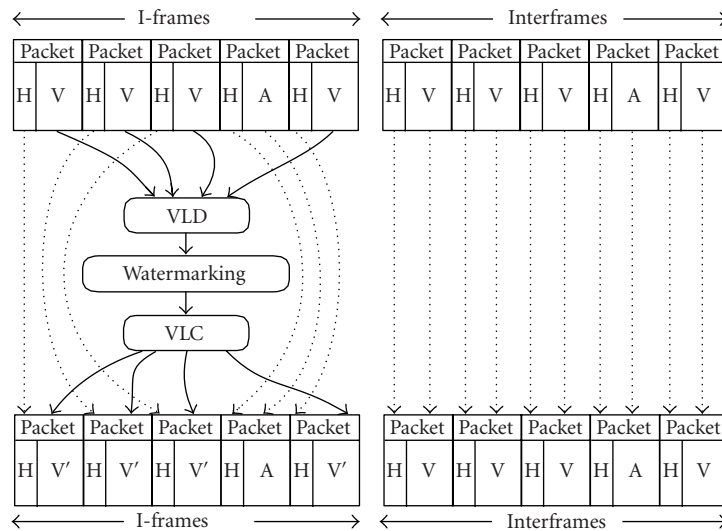


FIGURE 1: Operations performed on an MPEG multiplexed stream (V: encoded video data, A: encoded audio data, H: elementary stream packet header, Packet: elementary stream packet, V': watermarked encoded video data, VLC: variable length coding, VLD: variable length decoding).

- (iii) *Preserving file size.* The size of the MPEG file should not be altered significantly. The watermark embedding procedure should take into account that the total MPEG file size should not be significantly increased, because an MPEG file may have been created so as to conform to specific bandwidth or storage constraints. This may be accomplished by watermarking only those DCT coefficients whose corresponding variable length code (VLC) words after watermarking will have less than or equal length to the length of the original VLC words, as in [13, 18, 19, 31].
- (iv) *Avoiding/compensating drift error.* Due to the nature of the interframe coding applied by MPEG, alterations of the coded data in one frame may propagate in time and cause alterations to the subsequent decoded frames. Therefore, special care should be taken during the watermark embedding, to avoid visible degradation in subsequent frames. A drift error of this nature was encountered in [13], where the watermark was embedded in all video frames (intra- and interframes) in the compressed domain; the authors of [13] proposed the addition of a drift compensation signal to compensate for watermark signals from previous frames. Generally, either the watermarking method should be designed in a way such that drift error is imperceptible, or the drift error should be compensated, at the expense of additional computational complexity.

In the ensuing sections, an MPEG-1/2 watermarking system is described which meets the above requirements.

3. PREPROCESSING OF MPEG-1/2 MULTIPLEXED STREAMS

It is often preferable to watermark video in the compressed rather than the spatial domain. Due to high storage capac-

ity requirements, it is impractical or even infeasible to decompress and then recompress the entire video data. Decoding and reencoding an MPEG stream would also significantly increase the processing time, perhaps even to the point of rendering it prohibitive for use in real-time applications. For these reasons, in the present paper the video watermark embedding and detection methods are carried out entirely in the compressed domain.

MPEG-2 program streams and MPEG-1 system streams are multiplexed streams that contain at least two elementary streams, that is, an audio and a video elementary stream. A fast and efficient video watermarking system should be able to cope with multiplexed streams. An obvious approach to MPEG watermarking would be to use the following procedure. The original stream is demultiplexed to its comprising elementary video and audio streams. The video elementary stream is then processed to embed the watermark. Finally the resulting watermarked video elementary stream and the audio elementary stream are multiplexed again to produce the final MPEG stream. However, this process has a very high computational cost and a very slow implementation, which render it practically useless.

In order to keep complexity low, a technique was developed that does not fully demultiplex the stream before the watermark embedding, but instead deals with the multiplexed stream itself. The elementary video stream packets are first detected in the multiplexed stream. For those that contain I-frame data, the encoded (video) data are extracted and variable length decoding is performed to obtain the quantized DCT coefficients. The headers of these packets are left intact. This procedure is schematically described in Figure 1. The quantized DCT coefficients are first watermarked. Then the watermarked coefficients are variable length coded. The video encoded data are partitioned so that they fit into video packets that use their original headers.

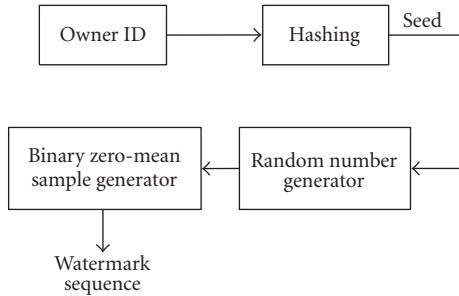


FIGURE 2: Watermark generation.

Audio packets and packets containing interframe data are not altered. The stream structure remains unaffected and only the video packets that contain coded I-frame data are altered. Note that the above process produces only minor variations in the bitrate of the original compressed video and does not impose any significant memory requirements to the standard MPEG coding/decoding process.

4. IMPERCEPTIBLE WATERMARKING IN THE COMPRESSED DOMAIN

4.1. Generation of the embedding watermark

We will use the following procedure for the generation of the embedding watermark. The values of the watermark sequence $\{W\}$ are either -1 or 1 . This sequence is produced from an integer random number generator by setting the watermark coefficient to 1 when the generator outputs a positive number and to -1 when the generator output is negative. The result is a zero-mean, unit variance process. The random number generator is seeded with the result of a hash function. The MD5 algorithm [32] is used in order to produce a 128 bit integer seed from a meaningful message (owner ID). The watermark generation procedure is depicted in Figure 2. As explained in [29], the watermark is generated so that even if an attacker finds a watermark sequence that leads to a high correlator output, he or she still cannot find a meaningful owner ID that would produce the watermark sequence through this procedure and therefore cannot claim to be the owner of the image. This is ensured by the use of the hashing function included in the watermark generation.

4.2. Imperceptible watermark embedding in the quantized DCT domain

The proposed watermark embedding scheme (Figure 3) modifies only the quantized AC coefficients $X_Q(m, n)$ of a luminance block (where m, n are indices indicating the position of the current coefficient in an 8×8 DCT block) and leaves chrominance information unaffected. In order to make the watermark imperceptible, a novel method is employed, combining perceptual analysis [10, 20] and block classification techniques [19, 21]. These are applied in the DCT domain in order to adaptively select which coefficients are best for watermarking. The product of the em-

bedding watermark coefficient $W(m, n)$, that is, the value of the pseudorandom sequence for the position (m, n) , with the corresponding values of the quantized embedding strength $S_Q(m, n)$ and the embedding mask $M(m, n)$ (which result from the perceptual analysis and the block classification process, respectively), is added to each selected quantized coefficient. The resulting watermarked quantized coefficient is given by $X'_Q(m, n)$:

$$X'_Q(m, n) = X_Q(m, n) + M(m, n)S_Q(m, n)W(m, n). \quad (1)$$

In order to select the embedding mask M , each DCT luminance block is initially 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 calculation of energy distribution and the subsequent block classification are performed as in [19], returning the class of the block examined. For each block class, the binary embedding mask M determines which coefficients are the best candidates for watermarking. Thus

$$M(m, n) = \begin{cases} 0, & \text{the } (m, n) \text{ coefficient} \\ & \text{will not be watermarked,} \\ 1, & \text{the } (m, n) \text{ coefficient} \\ & \text{can be watermarked (if } S_Q(m, n) \neq 0), \end{cases} \quad (2)$$

where $m, n \in [0, 7]$. The perceptual analysis that follows the block classification process leads to the final choice of the coefficients that will be watermarked and defines the embedding strength.

Figure 4 depicts the mask M for all the block classes. As can be seen, the embedding mask for all classes contains “zeros” for all high frequency AC coefficients. These coefficients are not watermarked because the embedded signal is likely to be eliminated by lowpass filtering or transcoding to lower bitrates. The rest of the zero $M(m, n)$ values in each embedding mask (apart from the low activity block mask) correspond to large DCT coefficients, which are left unwatermarked, since their use in the detection process may reduce the detector performance [19].

The perceptual model that is used is a new adaptation of the perceptual model proposed by Watson [20]. A measure $T''(m, n)$ is introduced to determine the maximum just noticeable difference (JND) for each DCT coefficient of a block. This model is then adapted for *quantized* DCT coefficients.

For a visual angle of $1/16$ pixels/degree and a 48.7 cm viewing distance, the *luminance masking* and the *contrast masking* properties of the human visual system (HVS) for each coefficient of a DCT block are estimated as in [20]. Specifically, two matrices, T' (*luminance masking*) and T'' (*contrast masking*) are calculated. Each value $T''(m, n)$ is compared with the magnitude of the corresponding DCT coefficient $|X(m, n)|$ and is used as a threshold to determine whether the coefficient will be watermarked or not. The values $T''(m, n)$ determine the embedding strength of

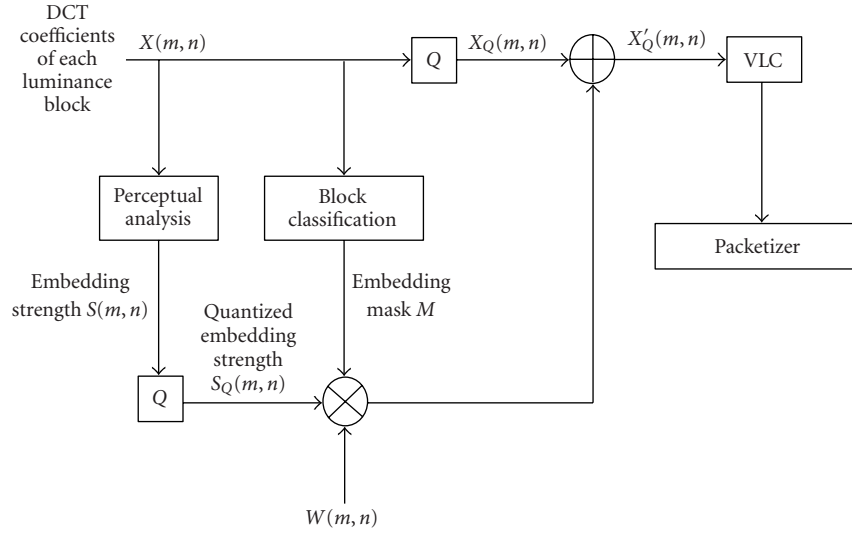


FIGURE 3: Watermark embedding scheme.

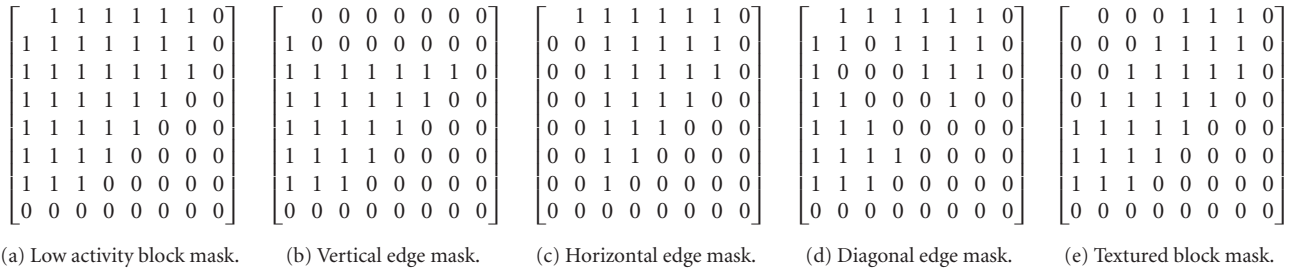


FIGURE 4: The embedding masks that correspond to each one of the five block classes.

the watermark $S(m, n)$ when $|X(m, n)| > T'(m, n)$:

$$S(m, n) = \begin{cases} T''(m, n), & \text{if } |X(m, n)| > T'(m, n), \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

Another approach would be to embed the watermark in the DCT coefficients $X(m, n)$, before quantization is applied; then the watermark embedding equation would be

$$X'(m, n) = X(m, n) + M(m, n)S(m, n)W(m, n). \quad (4)$$

However, as our experiments have shown, the embedded watermark, that is, the last term in the right-hand side of (4), is sometimes entirely eliminated by the quantization process. If this happens to a large number of coefficients, the damage to the watermark may be severe, and the watermark detection process may become unreliable. This is why the watermark is embedded directly in the quantized DCT coefficients. Since the MPEG coding algorithm performs no other lossy operation after quantization (see Figure 5), any information embedded as in Figure 5 does not run the risk of being eliminated by the subsequent processing. Thus, the watermark

remains intact in the quantized coefficients during the detection process when the quantized DCT coefficients $X_Q(m, n)$ are watermarked in the following way (see Figure 3):

$$X'_Q(m, n) = X_Q(m, n) + M(m, n)S_Q(m, n)W(m, n), \quad (5)$$

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

$$S_Q(m, n) = \begin{cases} \text{quant}[S(m, n)], & \text{if } \text{quant}[S(m, n)] > 1, \\ 1, & \text{if } \text{quant}[S(m, n)] \leq 1 \text{ and } \\ & S(m, n) \neq 0, \\ 0, & \text{if } S(m, n) = 0, \end{cases} \quad (6)$$

where $\text{quant}[\cdot]$ denotes the quantization function used by the MPEG video coding algorithm.

Figure 6 depicts a frame from the video sequence table tennis, the corresponding watermarked frame, and the difference between the two frames, amplified and contrast-enhanced in order to make the modification produced by the watermark embedding more visible.

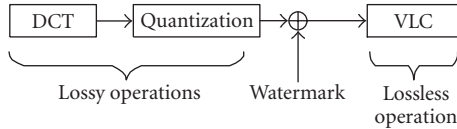
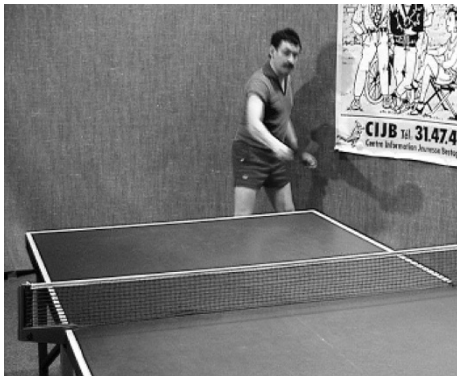


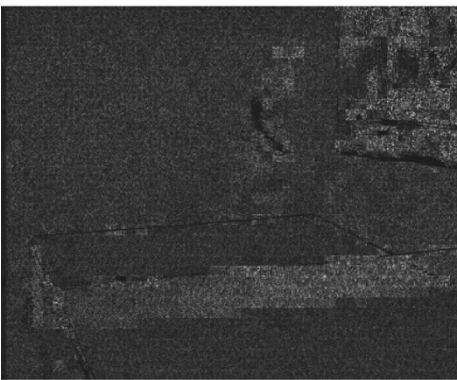
FIGURE 5: MPEG encoding operations.



(a)



(b)



(c)

FIGURE 6: (a) Original frame from the video sequence table tennis, (b) watermarked frame, (c) amplified difference between the original and the watermarked frame.

Various video sequences were watermarked and viewed in order to evaluate the imperceptibility of the watermark embedding method. The viewers were unable to locate any degradation in the quality of the watermarked videos. Table 1 presents the mean of the PSNR values of all the frames of some commonly used video sequences. In addition, Table 1 shows the mean of the PSNR values of the I-frames (watermarked frames) of each video sequence. Additionally, the good visual quality of the various watermarked video sequences that were viewed showed that the proposed I-frame embedding method does not cause any significant drift error. The effect of the watermark propagation was also measured, in terms of PSNR values, for the table tennis video sequence. Figure 7 presents the PSNR values of all frames of a typical group of pictures (GOP) of the video sequence. As can be seen, the PSNR values for all P- and B-frames of the GOP are higher than the PSNR value of the I-frame. Generally, due to the motion compensation process, the watermark embedded in the macroblocks of an I-frame is transferred to the macroblocks of the P- and B-frames, except for the cases where the macroblocks of the P- and B-frames are intra-coded. Therefore, the quality degradation in the inter-frames should not exceed the quality degradation of the I-frame of the same GOP or the next GOP.¹

4.3. The effect of watermark embedding on the video file size

The absolute value of $X'_Q(m, n)$ in (5) may increase, decrease or may remain unchanged in relation to $|X_Q(m, n)|$, depending on the sign of the watermark coefficient $W(m, n)$ and the values of the embedding mask and the embedding strength. 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 $X_Q(m, n)$; the inverse is true when $|X'_Q(m, n)| < |X_Q(m, n)|$. Since the watermark sequence has zero mean, the number of cases where $|X'_Q(m, n)| > |X_Q(m, n)|$ is expected to be roughly equal to the number of 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 videos resulted in bitstreams whose size differed slightly (up to 2%) compared to the original. Table 2 presents the effect of watermark embedding in the file size for some commonly used video sequences.

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. However, this reduces 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.

¹This case may hold for the last B-frame(s) in a GOP, which are decoded using information from the next I-frame. These frames may have a lower PSNR value than the PSNR value of the I-frame of the same GOP but their PSNR is higher than the PSNR of the next I-frame.

TABLE 1: Mean PSNR values for the frames of 4 watermarked video sequences (MPEG-2, 6 Mbits/s, PAL).

Video sequence	Mean PSNR for all video frames	Mean PSNR for I-frames only
Flowers	38.6 dB	36.5 dB
Mobile and calendar	33.1 dB	30 dB
Susie	45.6 dB	40.4 dB
Table tennis	35.6 dB	33.2 dB

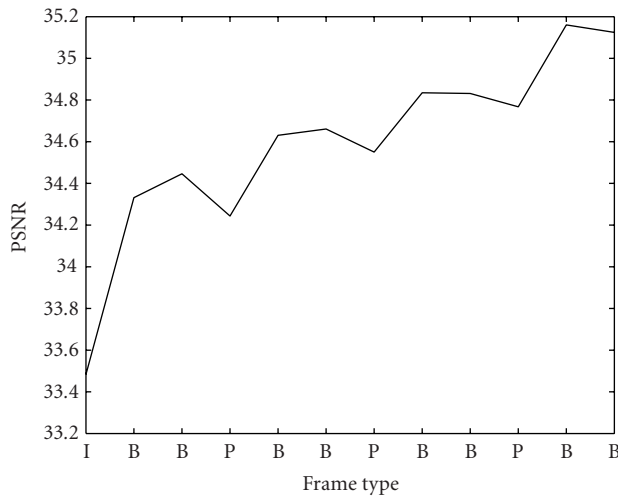


FIGURE 7: The PSNR values of all frames of a typical GOP of the video sequence table tennis (GOP size = 12 frames).

TABLE 2: The file size difference between the original and the watermarked video file as a percentage of the original file size.

Video sequence	Percentage (%)
Flowers (MPEG-2, 6 Mbits/s, PAL)	0.4
Mobile and calendar (MPEG-2, 6 Mbits/s, PAL)	1
Susie (MPEG-2, 6 Mbits/s, PAL)	1.1
Table tennis (MPEG-2, 6 Mbits/s, PAL)	1.4

5. WATERMARK DETECTION

The detection of the watermark is performed *without* the use of the original data. The original meaningful message that produces the watermark sequence W is needed in order to check if the specified watermark sequence exists in a copy of the watermarked video. Then, a correlation-based detection approach is taken similar to that analyzed in [29].

In Section 5.1, the correlation metric calculation is formulated. Section 5.2 presents the method used for calculating the threshold to which the detector output is compared, in order to decide whether a video frame is watermarked or not. In addition, the probability of detection is defined as a measure for the evaluation of the detection performance. Finally, in Section 5.3 a novel method is presented, for improving the performance of the watermark detection procedure

by preprocessing the watermarked data before calculating the correlation.

5.1. Correlation-based detection

The detection can be formulated as the following hypothesis test:

- (H_0) the video frame is not watermarked,
- (H_1) the video frame is watermarked with watermark W .

Another realistic scenario in watermarking would be the presence of a watermark different from W . In that case, the two hypotheses become

- (H'_0) the video frame is watermarked with watermark W'
- (H_1) the video frame is watermarked with watermark W .

Actually, this setup is not essentially different from the previous one: in fact, in (H_0) and (H_1) the data may be considered to be watermarked with $W' = 0$ under (H_0) , while in (H'_0) and (H_1) , under (H'_0) we may have $W' \neq 0$.

In order to determine which of the above hypotheses is true, for either (H_0) and (H_1) , or (H'_0) and (H_1) , a correlation-based detection scheme is applied. Variable length decoding is first performed to obtain the quantized DCT coefficients. The DCT coefficients for each block, which will be used in the detection procedure, are then obtained via inverse quantization. The block classification and perceptual analysis procedures are performed as described in Section 4 in order to define the set $\{X\}$ of the N DCT coefficients that are expected to be watermarked with the sequence $\{W\}$. Only these coefficients will be used in the correlation test (since the rest are probably not watermarked) leading to a more efficient detection scheme.

Each coefficient in the set $\{X\}$ is multiplied by the corresponding watermark coefficient of the correlating watermark sequence $\{W\}$, producing the data set $\{X_W\}$. The correlation metric c for each frame is calculated as

$$c = \frac{\text{mean} \cdot \sqrt{N}}{\sqrt{\text{variance}}}, \quad (7)$$

where

$$\text{mean} = \frac{1}{N} \sum_{l=0}^{N-1} X_W(l) = \frac{1}{N} \sum_{l=0}^{N-1} X(l)W(l) \quad (8)$$

is the sample mean of $\{X_W\}$, and

$$\begin{aligned} \text{variance} &= \frac{1}{N} \sum_{l=0}^{N-1} (X_W(l) - \text{mean})^2 \\ &= \frac{1}{N} \sum_{l=0}^{N-1} (X(l)W(l) - \text{mean})^2 \end{aligned} \quad (9)$$

is the sample variance of $\{X_W\}$.

The correlation metric c is compared to the threshold T : if it exceeds this threshold, the examined frame is considered watermarked. The calculation of the threshold is discussed in the following subsection.

5.2. Threshold calculation and probability of detection for DCT domain detection

After the correlation metric c is calculated, it is compared to the threshold T . However, in order to define the optimal threshold in either the Neyman-Pearson or Bayesian sense, a statistical analysis of the correlation metric c is required.

The correlation metric c of (7) is a sum of a large number of independent random variables. The terms of the sum are products of (watermarked or not) DCT coefficients with the corresponding values of the watermark. The DCT coefficients are independent random variables due to the decorrelating properties of the DCT. The watermark values are also independent by their construction, since we are examining spread-spectrum watermarking. The corresponding products can then be easily shown to be independent random variables as well. Then, for large N , and by the central limit theorem (CLT) [33], the distribution of the correlation metric c can be approximated by the normal distribution $N(m_0, \sigma_0)$ under (H_0) and $N(m_1, \sigma_1)$ under (H_1) . Also, under (H'_0) it can easily be shown that the correlation metric still follows the same distribution $N(m_0, \sigma_0)$ as under (H_0) . Based on [29], the means and standard deviations of these distributions are given by

$$m_0 = m'_0 = 0, \quad (10)$$

$$\sigma_0 = \sigma'_0 = 1, \quad (11)$$

$$m_1 = \frac{E[\text{quant}^{-1}[S_Q(l)]]\sqrt{N}}{\sqrt{\text{variance}}} \quad (12)$$

$$\begin{aligned} &\approx \frac{\sum_{l=0}^{N-1} \text{quant}^{-1}[S_Q(l)]}{\sqrt{\text{variance} \cdot N}}, \\ &\sigma_1 = 1, \end{aligned} \quad (13)$$

where $E[\cdot]$ denotes the expectation operator, $\text{quant}^{-1}[\cdot]$ denotes the function that MPEG uses for mapping quantized coefficients to DCT values, and $S_Q(l)$ is the quantized embedding strength that was used for embedding the watermark in the l th of the N DCT coefficients of the set $\{X\}$.

The error probability P_e for equal priors ($P_{(H_0)} = P_{(H_1)} = 1/2$) is given by $P_e = (1/2)(P_{FP} + P_{FN})$, where P_{FP} is the false positive probability (detection of the watermark under (H_0)) and P_{FN} is the false negative probability (failure to detect the watermark under (H_1)). The analytical expressions of P_{FP} and P_{FN} are then given by

$$P_{FP} = Q\left(\frac{T - m_0}{\sigma_0}\right) = Q(T), \quad (14)$$

$$P_{FN} = 1 - Q\left(\frac{T - m_1}{\sigma_1}\right) = 1 - Q(T - m_1), \quad (15)$$

where T is the threshold against which the correlation metric is compared and $Q(x)$ is defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-t^2/2} dt. \quad (16)$$

Since $\sigma_0 = \sigma_1$, it can easily be proven that the threshold selection T_{MAP} which minimizes the detection error probability P_e (maximum a posteriori criterion) is given by

$$T_{MAP} = \frac{m_0 + m_1}{2} = \frac{m_1}{2}. \quad (17)$$

In practice, this is not a reliable threshold, mainly because in case of attacks the mean value m_1 is not accurately estimated using (12). In fact, experimental results have shown that in case of attacks the experimental mean of the correlation value under (H_1) is smaller than the theoretical mean m_1 calculated using (12). The Neyman-Pearson threshold T_{NP} is preferred, as it leads to the smallest possible probability P_{FN} of false negative errors while keeping false positive errors at an acceptable predetermined rate. By solving (14) for T we obtain

$$T_{NP} = Q^{-1}(P_{FP}). \quad (18)$$

Equation (18) will be used for the calculation of the threshold for a fixed P_{FP} since the mean and the variance of the correlation metric under (H_0) have constant values. Furthermore, to evaluate the actual detection performance, the probability of detection P_D as a function of the threshold T_{NP} is calculated using the following expression:

$$P_D = Q\left(\frac{T_{NP} - m_1}{\sigma_1}\right). \quad (19)$$

5.3. Nonlinear preprocessing of the watermarked data before correlation

The correlation-based detection presented in this section would be optimal if the DCT coefficients followed a normal distribution. However, as described in [34, 35], the distribution of image DCT coefficients is more accurately modeled by a heavy-tailed distribution such as the Laplace, Cauchy, generalized Gaussian, or symmetric alpha stable (SaS) [36] with the maximum likelihood detector derived as shown in [16, 37] for the Laplacian distribution and in [38] for the Cauchy distribution. This detector outperforms the correlator in terms of detection performance, but may not be as simple and fast as the correlation-based detector. Also, modeling of the DCT data to acquire the parameters that characterize each distribution is required, thus increasing the detection time. This is why, in many practical applications, the suboptimal but simpler correlation detector is used.

Another approach used in signal detection to improve the correlation detector's performance is the use of LO detectors [22, 23], which achieves asymptotically optimum performance for low signal levels. In the watermarking problem, the strength of the embedded signal is small, so an LO test is appropriate for it. These detectors originate from the log-likelihood ratio, which can be written as

$$l(X) = \sum_{l=0}^{N-1} \ln \left(\frac{f_X(X(l) - W(l))}{f_X(X(l))} \right), \quad (20)$$

where $f_X(X)$ is the pdf of the video or image data. The watermark strength is small, so we have the following Taylor series approximation:

$$\begin{aligned} l(X(l)) \big|_{W(l)} &= l(X(l)) \big|_{W(l)=0} + \frac{\partial l(X(l))}{\partial X(l)} \bigg|_{W(l)=0} \cdot W(l) \\ &\quad + o(|W(l)|) \\ &\simeq -\frac{f'_X(X(l))}{f_X(X(l))} \cdot W(l) + o(|W(l)|) \\ &\simeq g_{LO}(X(l)) \cdot W(l), \end{aligned} \quad (21)$$

where we neglect the higher-order terms $o(|W(l)|)$ as they will go to zero. In this equation, $g_{LO}(X)$ is the "LO nonlinearity" [22, 23], defined by

$$g_{LO}(X) = -\frac{f'_X(X)}{f_X(X)}. \quad (22)$$

Thus, the resulting detection scheme basically consists of the nonlinear preprocessor $g_{LO}(X)$ followed by the linear correlator, which is why such systems are also known as generalized correlator detectors [22]. Such nonlinearities are often encountered in communication systems that operate in the presence of non-Gaussian noise, as they suppress the observations with high magnitude that cause the correlator's performance to deteriorate.

In an LO detection scheme (i.e., correlation with preprocessing), the data set $\{X_W\}$ used in (8) and (9) for the calculation of the correlation metric of (7) is replaced by the values calculated by multiplying the elements $g_{LO}(X(l))$ of the preprocessed data (note that $X(l)$ is an element of the data $\{X\}$) with the corresponding watermark coefficient $W(l)$ of the correlating watermark sequence.

It is obvious from (22) that an appropriate nonlinear preprocessor can be chosen based on the distribution of the frame data (i.e., the host) and the signal to be detected (the watermark). The DCT coefficients used here can be quite accurately modeled by the Cauchy or the Laplacian distributions. Table 3 depicts the expressions for the density functions of these distributions and the corresponding nonlinear preprocessors.

Experiments were carried out to evaluate the effect of these nonlinearities on the detection performance. It was shown that the use of either nonlinearity significantly improved the performance of the detector, on both nonattacked and attacked videos.

TABLE 3

pdf of frame DCT data	Nonlinearity used for preprocessing
$f_X(x) = \frac{b}{2} \exp(-b x - \mu)$	$g_{LO}(x) = b \cdot \text{sgn}(x - \mu)$
$f_X(x) = \frac{1}{\pi} \frac{\gamma}{(x - \delta)^2 + \gamma^2}$	$g_{LO}(x) = \frac{2(x - \delta)}{(x - \delta)^2 + \gamma^2}$

In the case of Cauchy distributed data, the corresponding nonlinearity requires the modeling of the DCT data in order to obtain the parameters γ and δ . For the Laplacian nonlinearity, it may initially appear that the parameters b and μ of this distribution need to be estimated. However, after careful examination of the Laplacian preprocessor, it is seen that this is not really required. As we verified experimentally, we may assume that the mean value μ of the watermarked DCT coefficients is zero, so there is no need to calculate this parameter. Furthermore, after a little algebra, it is also seen that the Laplacian parameter b does not appear in the final expression for this nonlinearity. Specifically, if in (7), (8), and (9), we replace the watermarked data with the preprocessed watermarked data, we easily observe that b is no longer present in the final expression for c :

$$\text{mean} = \frac{1}{N} \sum_{l=0}^{N-1} g_{LO}(X(l)) \cdot W(l) \quad (23)$$

$$= \frac{1}{N} \sum_{l=0}^{N-1} b \cdot \text{sgn}(X(l)) W(l),$$

$$\begin{aligned} \text{variance} &= \frac{1}{N} \sum_{t=0}^{N-1} (g_{LO}(X(t)) \cdot W(t) - \text{mean})^2 \\ &= \frac{1}{N} \sum_{t=0}^{N-1} \left(b \cdot \text{sgn}(X(t)) W(t) - \frac{1}{N} \sum_{l=0}^{N-1} b \cdot \text{sgn}(X(l)) W(l) \right)^2, \end{aligned} \quad (24)$$

$$c = \frac{\frac{1}{N} \sum_{l=0}^{N-1} \text{sgn}(X(l)) W(l) \cdot \sqrt{N}}{\sqrt{\frac{1}{N} \sum_{t=0}^{N-1} \left(\text{sgn}(X(t)) W(t) - \frac{1}{N} \sum_{l=0}^{N-1} \text{sgn}(X(l)) W(l) \right)^2}}, \quad (25)$$

where $X(l)$ are the N DCT coefficients of the data set $\{X\}$ that are used in the detection process and $W(l)$ are the corresponding correlating watermark coefficients. Thus, we finally choose to use a generalized correlator detector corresponding to Laplacian distributed data because this detector does not actually add any computational complexity (by the estimation of b and μ) to the existing implementation.

In order to define the threshold in the case of the proposed generalized correlator detector, the statistics of the correlation metric c given by (25) need to be estimated again. Under either hypothesis (H_0) or (H_1), the assumptions made for estimating the statistics of c in Section 5.2 are still valid. Specifically, the correlation metric c is still a sum of independent random variables, regardless of whether or not

preprocessing has been used. Thus, by the CLT, and for a sufficiently large data set (a condition that is very easily satisfied in our application, since there are many DCT coefficients available from the video frame—typically $N > 25000$ for PAL resolution video frames), the test statistic c will follow a normal distribution. Therefore, the distribution of c under (H_0) and (H'_0) can still be approximated by $N(0, 1)$ and the same threshold (equation (18)) as in the case of the correlation-based detector proposed in the previous section, can also be used for the proposed generalized correlator detector.

Under (H_1) it is not possible to find closed form expressions for the mean m_1 and variance σ_1^2 of the correlation statistic c , due to the nonlinear nature of the preprocessing. Nevertheless, c still follows a normal distribution $N(m_1, \sigma_1)$. The mean and variance of c under (H_1) can be found experimentally by performing many Monte Carlo runs with a large number of randomly generated watermark sequences. Then, the probability of detection can be calculated using (19). Such experiments are described in Section 7, where the superior performance of the proposed generalized correlator detector can be observed.

6. VIDEO WATERMARK DETECTOR IMPLEMENTATION

The proposed correlation-based detection (with or without preprocessing) described in Section 5 can be implemented using two types of detectors.

The first detector (*detector-A*) detects the watermark only in I-frames during their decoding by applying the procedure described in Section 5.1. Detector-A can be used when the video sequence under examination is the original watermarked sequence. It can also be used in cases where the examined video sequence has undergone some processing but maintains the same GOP structure as the original watermarked sequence. For example, this may happen when the video sequence is encoded at a different bit-rate using one of the techniques proposed in [39, 40]. This detector is very fast since it introduces negligible additional computational load to the decoding operation.

The second detector (*detector-B*) assumes that the GOP structure may have changed due to transcoding and frames that were previously coded as I-frames may now be coded as B- or P-frames. This detector decodes and applies DCT to each frame in order to detect the watermark using the procedure described in Section 5.1. The decoding operation performed by this detector may also consist of the decoding of non-MPEG compressed or uncompressed video streams, in case transcoding of the watermarked sequence to another coding format has occurred.

In cases where transcoding and I-frame skipping are performed on an MPEG video sequence, then detector-B will try to detect the watermark in previous B- and P-frames. If object motion in the scene is slow, or slow camera zoom or pan occurs, then the watermark will be detected in B- and P-frames as will be shown in the correlation metric plots for all frames of the test video sequence described in Section 7. Of course, the watermark may not be detected in any of the video frames. When this occurs, the transcoded video qual-

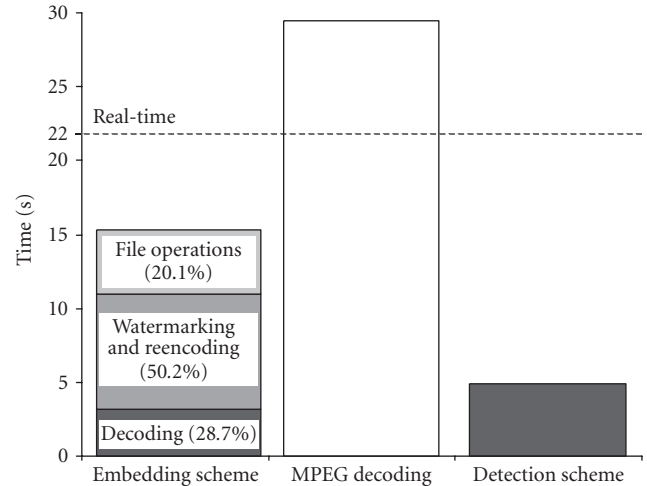


FIGURE 8: Speed performance of the embedding and detection schemes.

ity is severely degraded due to frame skipping (jerkiness will be introduced or visible motion blur will appear even if interpolation is used). Thus, it is very unlikely that an attacker will benefit from such an attack.

7. EXPERIMENTAL EVALUATION

The evaluation of the proposed watermarking scheme was based on experiments testing its speed and others testing the detection performance under various conditions. In addition, experiments were carried out to verify the validity of the analysis concerning the distributions of the correlation metric performed in Sections 5.2 and 5.3.

7.1. Speed performance of the watermarking scheme

The video sequence used for the first type of experiments was the MPEG-2 video *spokesman* which is part of a TV broadcast. This is an MPEG-2 program stream, that is, a multiplexed stream containing video and audio. It was produced using a hardware MPEG-1/2 encoder from a PAL VHS source. The reason for using such a test video sequence instead of more commonly used sequences like table tennis or *foreman* is that the latter are short video-only sequences that are not multiplexed with audio streams, as is the case in practice. Of course, the system also supports such video-only MPEG-1/2 streams. In general, the embedding and detection schemes support constant and variable bitrate main profile MPEG-2 program streams and MPEG-1 system streams, as well as video-only MPEG-1/2 streams (only progressive sequences in all cases).

The proposed embedding algorithm was simulated using a Pentium 866 MHz processor. The total execution time of the embedding scheme for the 22-second MPEG-2 (5 Mbit/s, PAL resolution) video sequence *spokesman* is 72% of the real-time duration of the video sequence. Execution time is allocated to the three major operations performed for embedding: file operations (read and write headers, and packets), partial decoding, and partial encoding and watermarking as shown in Figure 8. In Figure 8 the embedding time is also

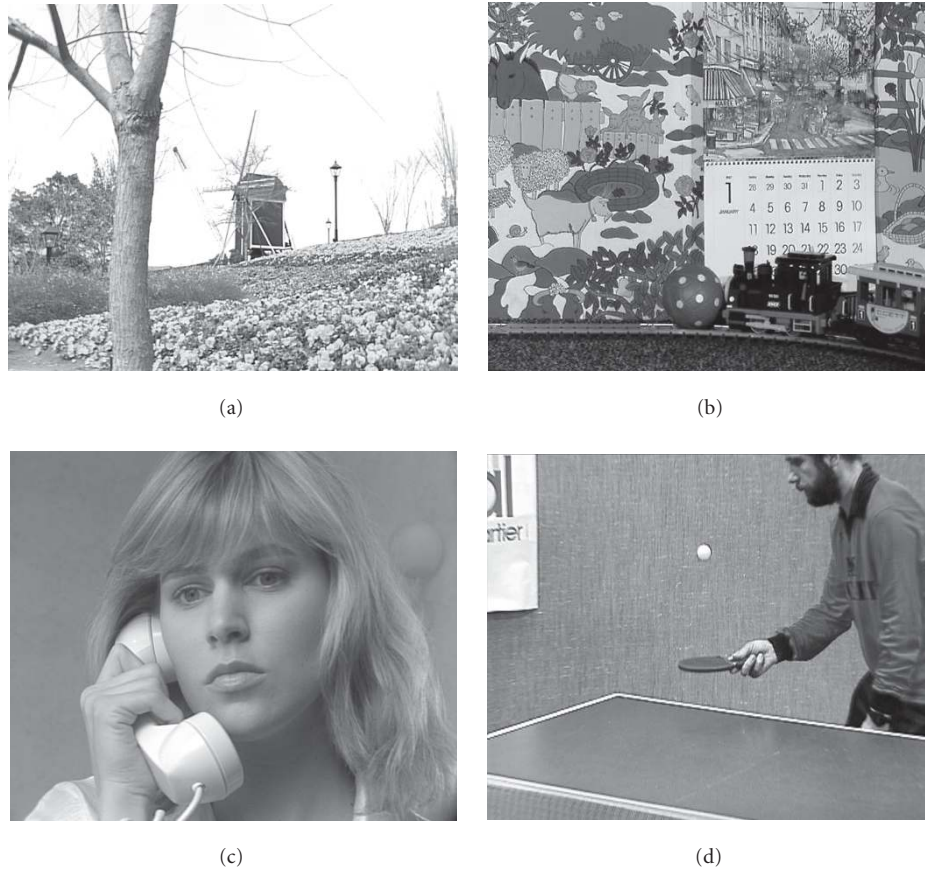


FIGURE 9: Selected frames for use in the experiments. The frames belong to the (a) flowers, (b) mobile and calendar, (c) Susie, and (d) table tennis video sequences (MPEG-2, 6 Mbits/s, PAL).

compared to the decoding time (without saving each decoded frame to a file) using the MSSG [41] decoder software. Clearly, the embedding time is significantly shorter than the decoding and reencoding time that would be needed if the watermark embedding were performed in the spatial domain. Figure 8 also presents the time required for detection using the detector-A described in Section 6. Detection time is only 23% of the real-time duration of the video sequence, thus enabling the detector to be incorporated in real-time decoders/players. Using detector-B, the detection process takes five times the real-time duration of the video. This makes detector-B more suitable for offline watermark detection.

7.2. Correlation metric distributions and probability of detection

For the rest of the experiments presented in this section, commonly used video-only sequences with various types of content were used. Specifically, the table tennis, flowers, mobile and calendar, and Susie (all PAL resolution, 15 seconds, 375 frames, 32 I-frames) were selected.

In the first experiment, a typical frame from each one of the above sequences was selected. Figure 9 presents the selected frames. Then, for each one of these unwatermarked

frames, the correlation metric c of (7) (DCT domain detection) and (25) (detection using preprocessing) was calculated for 5000 different correlating watermarks (H_0). Subsequently, the selected frames were watermarked with a specific watermark and using 5000 different watermarks, we conducted 5000 Monte Carlo runs to calculate the correlation metric for both types of detectors (H'_0). Using these results, the means and standard deviations of the correlation metric were calculated and are shown in Table 4. It is easily seen that the corresponding values under hypotheses (H_0) and (H'_0) are very similar. This was expected, since (H_0) and (H'_0) are equivalent, as we have already explained. In addition, the selected frames were watermarked with 5000 different watermarks and, using the same 5000 watermarks, the correlation metric was calculated (H_1). Its means and standard deviations for both types of detectors are shown in Table 5.

Figure 10 also presents the experimental pdfs (under (H_0) and (H_1)) of the correlation metric for the selected frame of the mobile and calendar video sequence, where the Gaussian nature of all pdfs can be observed. The Gaussian distribution of the correlation metric is indeed verified by the normal probability plots depicted in Figure 11. In all cases, the plots are almost linear, showing that c follows a normal

TABLE 4: Means and variances of the correlation metric under (H_0) and (H'_0) for DCT domain detection and detection using preprocessing.

Frame from video sequence	DCT domain detection				Detection using preprocessing			
	m_0	σ_0	m'_0	σ'_0	m_0	σ_0	m'_0	σ'_0
Flowers	0.01	0.99	0.00	0.99	-0.02	1.00	-0.02	1.00
Mobile and calendar	0.02	0.99	0.00	1.01	0.00	1.00	0.00	1.02
Susie	0.02	0.99	0.01	0.98	0.01	0.98	0.00	0.96
Table tennis	-0.01	1.00	-0.03	1.00	-0.01	0.99	-0.01	0.99

TABLE 5: Means and variances of the correlation metric under (H_1) for DCT domain detection and detection using preprocessing.

Frame from video sequence	DCT domain detection		Detection using preprocessing	
	m_1	σ_1	m_1	σ_1
Flowers	56.36	0.96	66.91	1.01
Mobile and calendar	127.56	1.1	148.21	1.38
Susie	208.17	1.14	289.91	2.14
Table tennis	203.69	1.11	275.40	2.11

distribution. In addition, as can be seen in Table 4, the correlation metric has $m_0 = m'_0 \approx 0$, $\sigma_0 = \sigma'_0 \approx 1$ under (H_0) and (H'_0) , for both types of detectors. Therefore, since under (H_0) and (H'_0) the experimental results for the distributions of the correlation metric match the theoretical, (18) can be used for the a priori determination of the threshold T_{NP} for both types of detectors. In order to achieve this, only the false positive probability P_{FP} that can be tolerated needs to be set for the Neyman-Pearson test. For all the experiments presented in the next subsection, the constant threshold $T_{NP} = 4.75$ was determined by using (18) and by selecting the false positive probability $P_{FP} = 10^{-6}$ since such a false positive probability is sufficient for copyright protection applications [42]. It should be noted that in the case of the experiment with the 5000 different correlating watermarks, under both (H_0) and (H'_0) , the detector output was always below the selected threshold. This shows that watermarks created by owner IDs different from the ID of the actual copyright owner cannot be used by others in order to claim copyright ownership of the video content.

We also directly compare the two detection methods by calculating the probability of detection P_D using (19). Since the threshold is determined in the same manner for both detection schemes, and since the distributions of the correlation metric may be assumed normal for both detectors, this comparison is meaningful. Figure 12a presents the resulting diagram for the selected frame of the mobile and calendar video sequence. In addition, Figure 12b presents the probability of detection diagram in the case where the MPEG video is transcoded from 6 Mbit/s to 3 Mbit/s. In both diagrams, it is easy to notice the superior performance of the detector that uses preprocessing.

7.3. Watermark detection under various attacks

In Section 5, DCT domain detection with and without preprocessing was presented. Experiments were conducted for both detection methods and their results were compared in

absolute terms and in cases of attacks. First, the 6 Mbit/s MPEG-2 video table tennis was watermarked and transcoded to 5, 4, and 3 Mbit/s video sequences. The original watermarked sequence and the transcoded sequences were correlated with the valid correlating watermark and a false watermark. The correlator output results for an I-frame (the 15th I-frame of the sequence and also the 168th frame of the same sequence) of this sequence are given in Table 6. It can be easily observed that both detection methods have a very good performance under transcoding. In addition, the correlator output in *case II* (detection using preprocessing) is much higher than in *case I* (DCT domain detection) when correlation is performed with a valid watermark. Finally, the correlator output is very low in both cases when correlation is performed with a false watermark.

In Figure 13a the correlation metric for both detection methods is evaluated for the 375 frames of the PAL resolution 8 Mbit/s MPEG-2 video sequence table tennis using detector-B. The constant threshold $T_{NP} = 4.75$, which is calculated as described in the previous subsection, is also plotted (in Figure 13a). The correlator output exceeds the threshold for all I-frames. The correlator output is also above the threshold for the P- and B-frames of scenes where slow motion occurs. For example, for the P- and B-frames between the 84th and the 312th frame, the correlator output is above the threshold. In the case where slow motion occurs, an attacker may remove the I-frames from the video sequence without causing severe degradation to its quality. In this case the watermark can still be detected in the rest of the frames of the video sequence for the range of frames where motion occurs, as depicted in Figure 13b.

In the rest of the experiments, various attacks were carried out on the watermarked MPEG-2 video table tennis to decrease the detectability of the watermark. The applied attacks were low-pass filtering (blurring), transcoding to lower bitrate MPEG streams, transcoding using the popular DivX (<http://www.divx.com/>) codec, and also transcoding using a

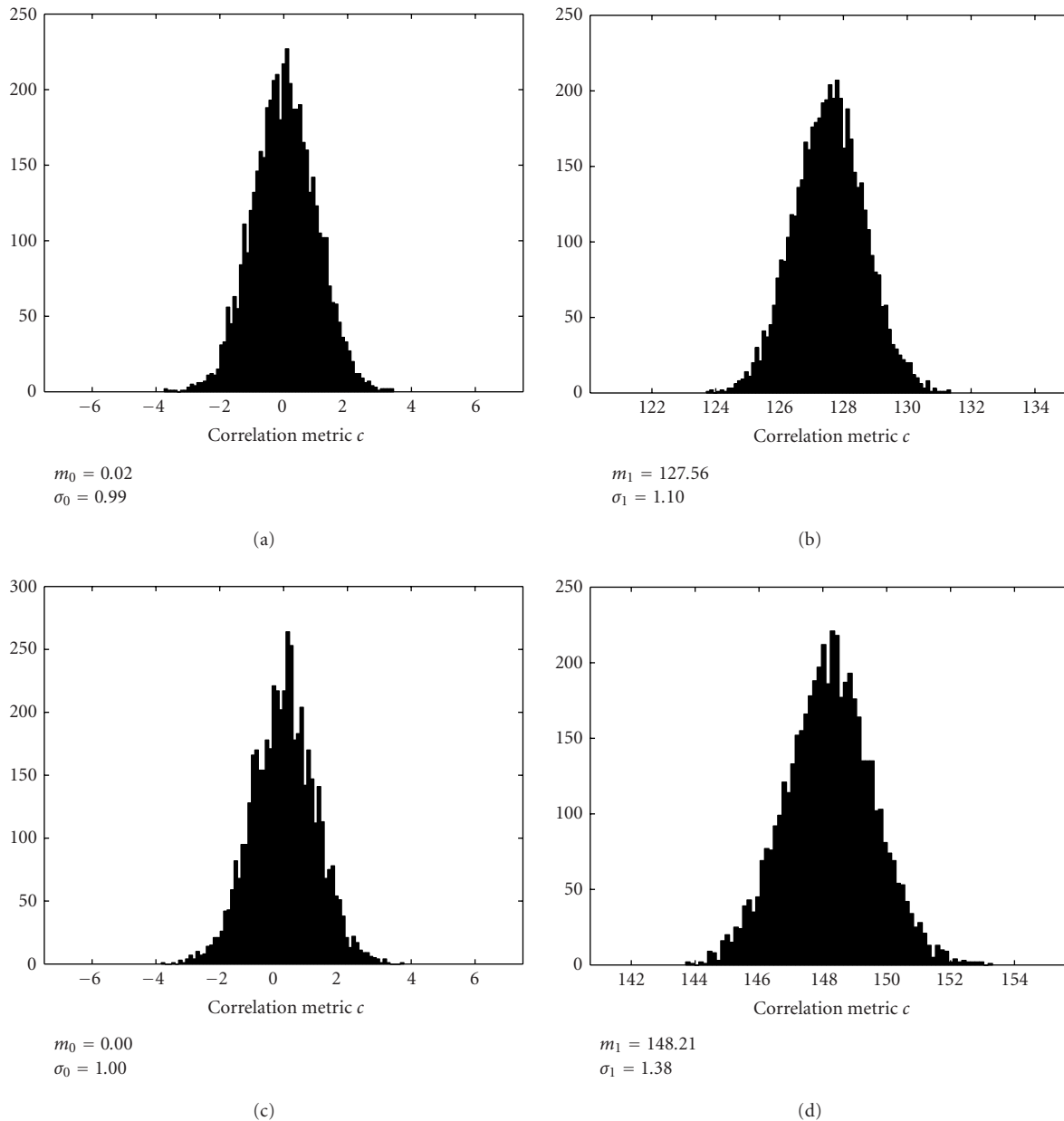


FIGURE 10: Experimentally evaluated pdfs of the correlation metric under (H_0) and (H_1) for the selected frame of the mobile and calendar video sequence: (a) DCT domain detection under (H_0) , (b) DCT domain detection under (H_1) , (c) detection using preprocessing under (H_0) , (d) detection using preprocessing under (H_1) .

wavelet-based coder (JPEG2000). In addition we tested the robustness of the proposed watermarking method to geometric attacks. Specifically, we first applied geometric attacks (cropping, scaling, and rotation) to the watermarked video. Then we reversed the geometric attacks, that is, we restored the attacked video to its original size, position, and orientation, and applied detection. These tests were performed in order to verify that techniques which can reverse the geometric transformation of the attack by using the nonattacked frame [43, 44] could be incorporated in the proposed scheme in order to offer resilience to geometric attacks.

The case in which the reversion of the geometric attack is not perfect was also investigated.

For all attacks except wavelet-based transcoding, *Adobe Premiere* (<http://www.adobe.com/products/premiere/main.html>) was used. Adobe Premiere is a video editing software widely used by video professionals. Built-in filters were used in order to perform the listed attacks on the watermarked sequence to simulate a possible scenario of video editing and processing.

DivX transcoding was performed using the DivX codec release 5.0.5. The coding parameters were chosen to simulate

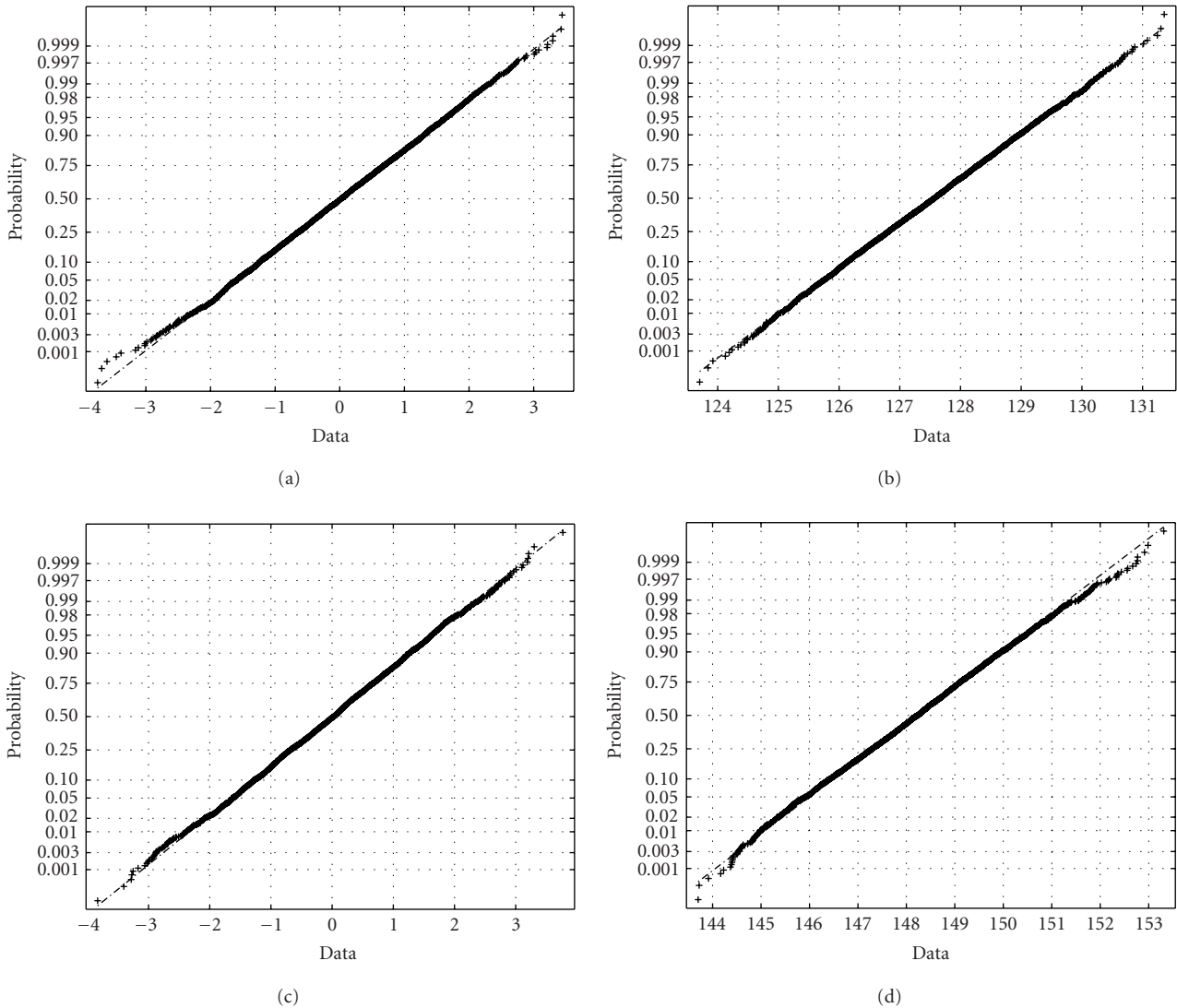


FIGURE 11: Normal probability plots for c using the *normplot* of Matlab: (a) DCT domain detection under (H_0), (b) DCT domain detection under (H_1), (c) detection using preprocessing under (H_0), (d) detection using preprocessing under (H_1).

a scenario where a movie of DVD quality is ripped and re-compressed in order to fit into one or two CDs (depending on the content and the duration of the movie). This attack was of particular interest since it is very easy to implement in practice.

Finally, in order to test the robustness of the watermarking method to non-DCT-based transcoding, we encoded each frame of the watermarked video using the wavelet-based JPEG2000 coder. We applied JPEG2000 compression at 0.4 bits/pixel (the original frame has 8 bits/pixel) and detected the watermark in each one of the decompressed frames after applying DCT and using the procedure described in Section 5.

The correlation metric plots for all frames of the video sequence for each one of the above attacks are given in Figures 14 and 15. After all the attacks presented in Figures 14a

to 14e, the watermark survived in all I-frames and was still detectable in interframes of scenes where slow motion occurred. The watermark also survived in nearly all I-frames after DivX transcoding (Figure 15a). Moreover, the watermark survived in most I-frames after a very severe attack combining DivX transcoding, cropping, and scaling (Figure 15b). In general, taking into account all conducted experiments, detection using preprocessing outperformed DCT domain detection in the majority of the detected frames.

8. CONCLUSIONS

A novel and robust way for embedding watermarks in MPEG-2 multiplexed streams was presented. The proposed scheme operates directly in the compressed domain and embeds copyright information without causing any visible

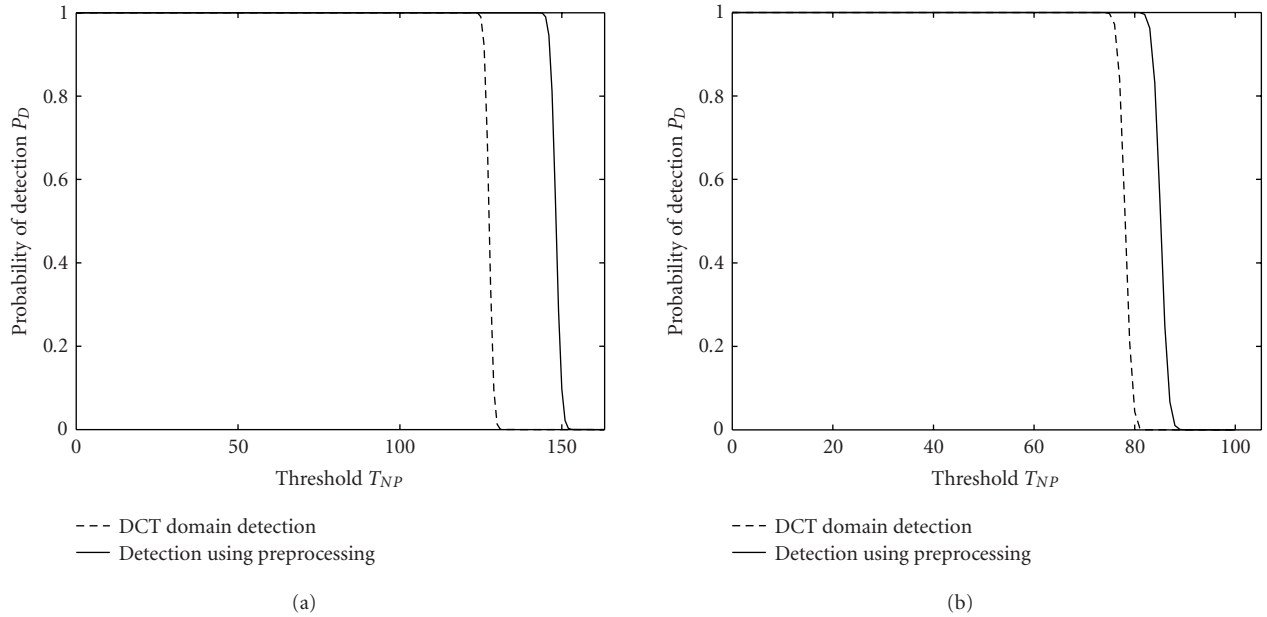


FIGURE 12: Probability of detection as a function of the threshold for both detection methods for: (a) the selected watermarked frame from mobile and calendar, (b) the transcoded watermarked frame from mobile and calendar.

TABLE 6: Correlator output results for watermark detection on the 15th I-frame (frame 168) of the MPEG-2 table tennis sequence using the correlation-based detection method and the detection using preprocessing.

Bitrate	Case I: DCT domain detection		Case II: detection using preprocessing	
	Valid watermark	False watermark	Valid watermark	False watermark
6 Mbit/s	115.6	0.8	212.1	0.3
5 Mbit/s	109.0	0.2	206.5	-0.4
4 Mbit/s	104.8	0.9	200.2	0.6
3 Mbit/s	100.2	1.5	176.2	1.1

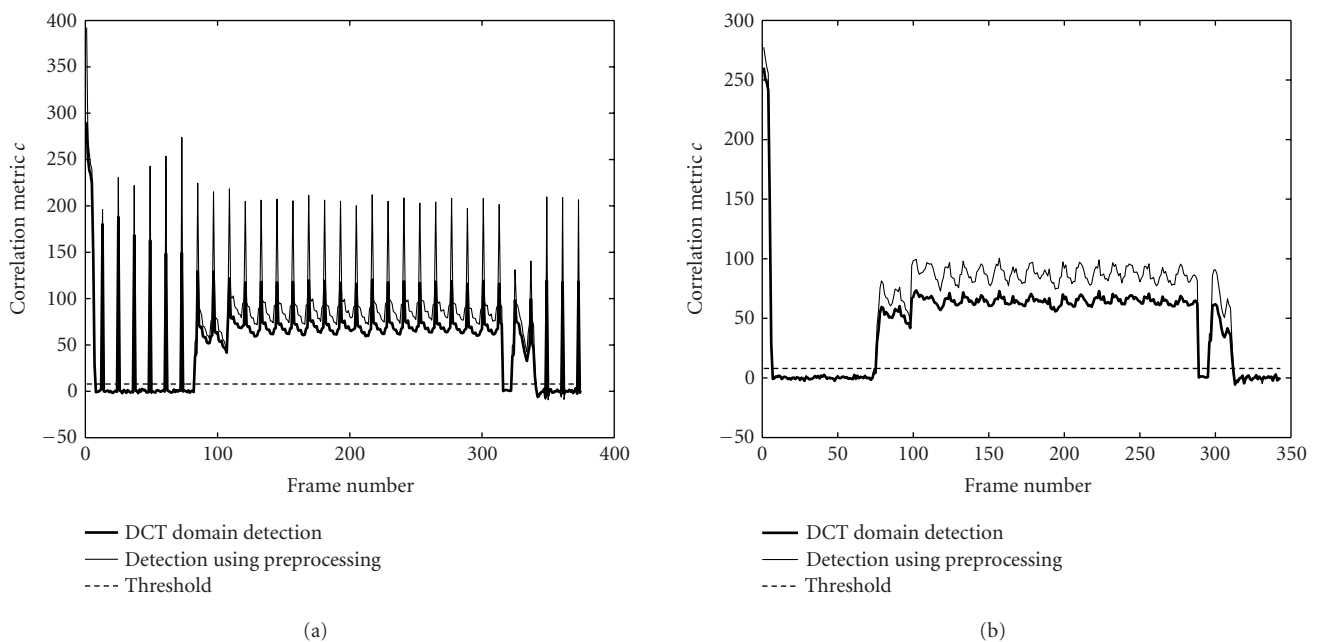


FIGURE 13: (a) Correlation metric plot for the 375 frames of the 8Mbit/s MPEG-2 video table tennis. (b) Correlation metric plot for the same video sequence with the I-frames skipped (this video sequence contains all but the frames that were encoded as I-frames before the skipping was performed).

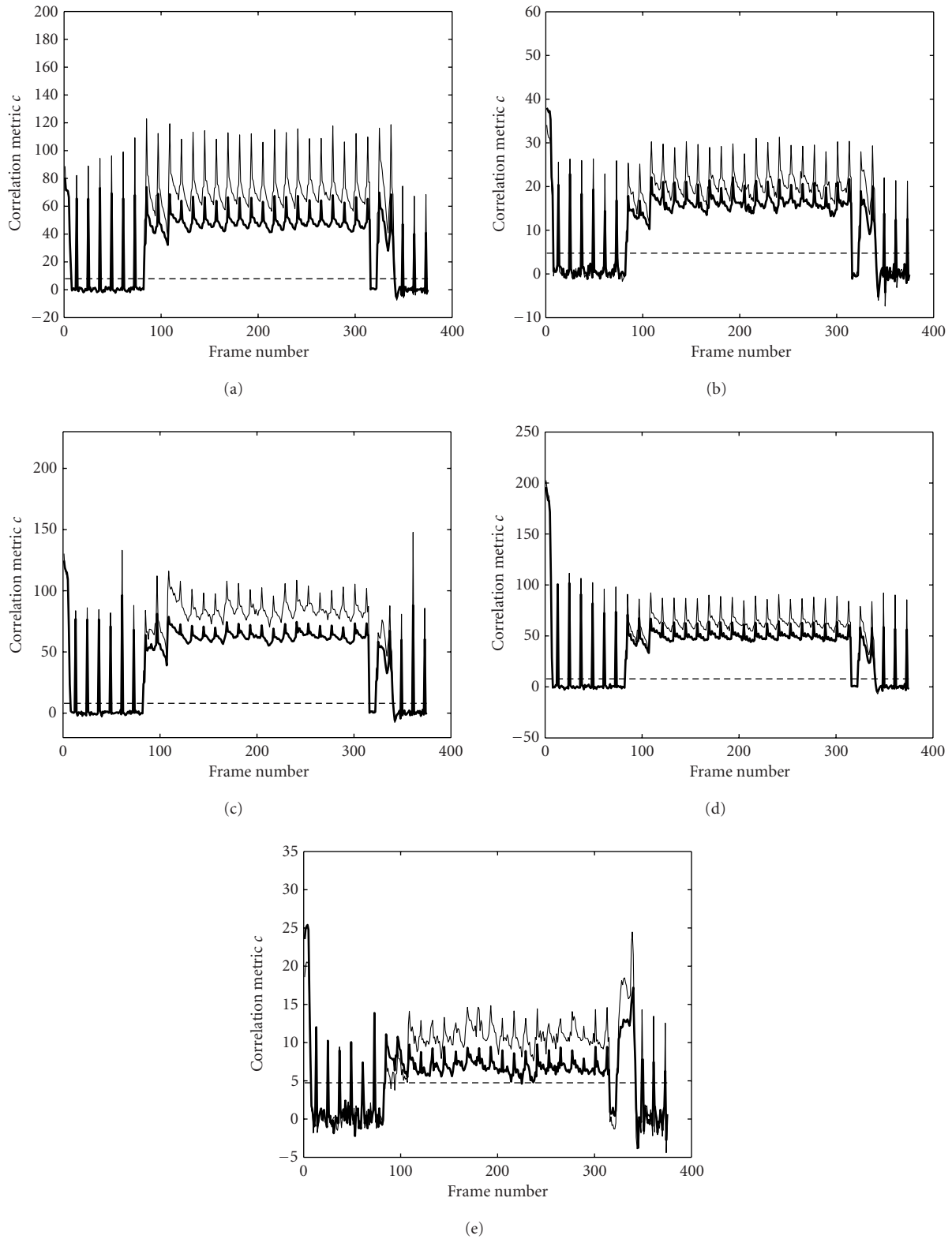


FIGURE 14: Plots of the correlation metric for all frames of the 8 Mbit/s MPEG-2 video table tennis after various attacks have been performed: (a) blurring, (b) cropping 40% of the frame area, rotation by 10° , and downscaling by 0.75, (c) MPEG transcoding to 4 Mbit/s, (d) transcoding with JPEG2000 to 0.4 bits/pixel, (e) downscaling by 0.75 (the reversion of the geometric attack was not perfect; the frames used for detection were larger by one row and one column than the original). The bold lines denote DCT domain detection, the solid ones denote detection using preprocessing, and the dashed, the threshold, as in Figure 13.

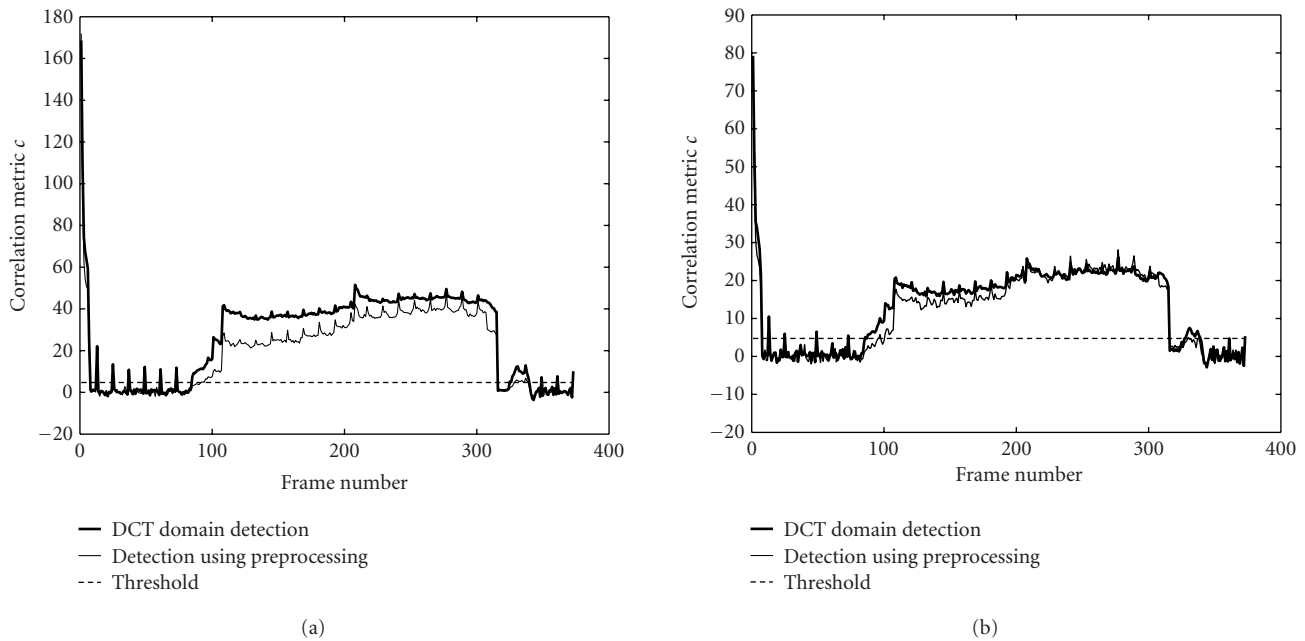


FIGURE 15: Plots of the correlation metric for all frames of the 8 Mbit/s MPEG-2 video table tennis after the following attacks have been performed: (a) DivX transcoding at 0.8 Mbit/s (two-pass run) and keyframes every 4 seconds, (b) DivX transcoding at 0.8 Mbit/s (two-pass run) and keyframes every 4 seconds, and cropping 40% of the frame area and downscaling by 0.75.

degradation to the quality of the video. The latter is achieved by combining perceptual and block classification techniques. Due to its speed, the resulting embedding scheme is suitable for real-time applications and also when a large number of video-sequences have to be watermarked, as is the case in video libraries.

An LO detector, the generalized correlator, was introduced and analyzed. This detector takes into account the Laplacian-like distribution of the DCT data by preprocessing the watermarked data before correlation. Experimental evaluation showed that this detector generally improves the detection results, leading to a watermarking scheme able to withstand attacks such as transcoding, and filtering, and even geometric attacks, if methods for reversing such attacks are incorporated.

Apart from being effective and reliable, the detection procedure used in the proposed scheme is fast, since it introduces negligible additional computational load to the decoding operation. This enables the proposed system to be useful not only for copyright protection but also as a component of real-time decoders/players that are used for applications such as broadcast monitoring.

ACKNOWLEDGMENT

This work was largely completed while all the authors were with Informatics and Telematics Institute, Thessaloniki, Greece.

REFERENCES

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

- [2] B. G. Haskell, A. Puri, and A. N. Netravali, *Digital Video: An Introduction to MPEG-2*, Kluwer Academic, Boston, Mass, USA, 1997.
- [3] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoan, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1673–1687, 1997.
- [4] M. Barni, F. Bartolini, V. Cappellini, and A. Piva, "A DCT-domain system for robust image watermarking," *Signal Processing*, vol. 66, no. 3, pp. 357–372, 1998.
- [5] D. Simitopoulos, D. Koutsonanos, and M. G. Strintzis, "Image watermarking resistant to geometric attacks using generalized Radon transformations," in *Proc. IEEE 14th International Conference on Digital Signal Processing*, vol. 1, pp. 85–88, Santorini, Greece, July 2002.
- [6] D. Tzovaras, N. Karagiannis, and M. G. Strintzis, "Robust image watermarking in the subband or discrete cosine transform domain," in *Proc. 9th European Signal Processing Conference*, pp. 2285–2288, Rhodes, Greece, September 1998.
- [7] D. Simitopoulos, N. V. Boulgouris, A. Leontaris, and M. G. Strintzis, "Scalable detection of perceptual watermarks in JPEG2000 images," in *Conference on Communications and Multimedia Security*, pp. 93–102, Darmstadt, Germany, May 2001.
- [8] M. D. Swanson, B. Zhu, A. H. Tewfik, and L. Boney, "Robust audio watermarking using perceptual masking," *Signal Processing*, vol. 66, no. 3, pp. 337–355, 1998.
- [9] G. C. Langelaar and R. L. Lagendijk, "Optimal differential energy watermarking of DCT encoded images and video," *IEEE Trans. Image Processing*, vol. 10, no. 1, pp. 148–158, 2001.
- [10] R. B. Wolfgang, C. I. Podilchuk, and E. J. Delp, "Perceptual watermarks for digital images and video," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1108–1126, 1999.
- [11] D. Simitopoulos, S. A. Tsiftaris, N. V. Boulgouris, and M. G. Strintzis, "Digital watermarking of MPEG-1 and MPEG-2 multiplexed streams for copyright protection," in *Proc. IEEE*

- International Workshop on Digital and Computational Video*, vol. 1, pp. 140–147, Tampa, Fla, USA, February 2001.
- [12] N. Nikolaidis and I. Pitas, “Robust image watermarking in the spatial domain,” *Signal Processing*, vol. 66, no. 3, pp. 385–403, 1998.
- [13] F. Hartung and B. Girod, “Watermarking of uncompressed and compressed video,” *Signal Processing*, vol. 66, no. 3, pp. 283–301, 1998.
- [14] J. O’Ruanaidh and T. Pun, “Rotation, scale and translation invariant spread spectrum digital image watermarking,” *Signal Processing*, vol. 66, no. 3, pp. 303–317, 1998.
- [15] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, “A new decoder for the optimum recovery of nonadditive watermarks,” *IEEE Trans. Image Processing*, vol. 10, no. 5, pp. 755–766, 2001.
- [16] J. R. Hernandez, M. Amado, and F. Perez-Gonzalez, “DCT-domain watermarking techniques for still images: detector performance analysis and a new structure,” *IEEE Trans. Image Processing*, vol. 9, no. 1, pp. 55–68, 2000.
- [17] H. Inoue, A. Miyazaki, and T. Katsura, “An image watermarking method based on the wavelet transform,” in *Proceedings of IEEE International Conference on Image Processing*, vol. 1, pp. 296–300, Kobe, Japan, October 1999.
- [18] G. C. Langelaar, R. L. Lagendijk, and J. Biemond, “Real-time labeling of MPEG-2 compressed video,” *Journal of Visual Communication and Image Representation*, vol. 9, no. 4, pp. 256–270, 1998.
- [19] T.-Y. Chung, M.-S. Hong, Y.-N. Oh, D.-H. Shin, and S.-H. Park, “Digital watermarking for copyright protection of MPEG2 compressed video,” *IEEE Transactions on Consumer Electronics*, vol. 44, no. 3, pp. 895–901, 1998.
- [20] A. B. Watson, “DCT quantization matrices visually optimized for individual images,” in *Proc. SPIE Conf. Human Vision, Visual Processing, and Digital Display IV*, vol. 1913 of *SPIE Proceedings*, pp. 202–216, San Jose, Calif, USA, February 1993.
- [21] K. R. Rao and J. J. Hwang, *Techniques and Standards for Image, Video and Audio Coding*, Prentice-Hall, Upper Saddle River, NJ, USA, 1996.
- [22] S. A. Kassam, *Signal Detection in Non-Gaussian Noise*, Springer-Verlag, New York, NY, USA, 1988.
- [23] V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York, NY, USA, 2nd edition, 1994.
- [24] K. R. Kolodziejski and J. W. Betz, “Detection of weak random signals in IID non-Gaussian noise,” *IEEE Trans. Communications*, vol. 48, no. 2, pp. 222–230, 2000.
- [25] S. Ambike, J. Ilow, and D. Hatzinakos, “Detection for binary transmission in a mixture of Gaussian noise and impulsive noise modeled as an alpha-stable process,” *IEEE Signal Processing Letters*, vol. 1, no. 3, pp. 55–57, 1994.
- [26] A. Swami and B. Sadler, “TDE, DOA and related parameter estimation problems in impulsive noise,” in *Proc. IEEE Signal Processing Workshop on Higher Order Statistics*, pp. 273–277, Banff, Canada, July 1997.
- [27] C. Busch, W. Funk, and S. Wolthusen, “Digital watermarking: from concepts to real-time video applications,” *IEEE Computer Graphics and Applications*, vol. 19, no. 1, pp. 25–35, 1999.
- [28] T. Kalker, G. Depovere, J. Haitsma, and M. Maes, “A video watermarking system for broadcast monitoring,” in *Proc. SPIE Electronic Imaging ’99, Security and Watermarking of Multimedia Contents*, vol. 3657 of *SPIE Proceedings*, pp. 103–112, San Jose, Calif, USA, January 1999.
- [29] 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, 1999.
- [30] 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 Trans. Image Processing*, vol. 10, no. 5, pp. 767–782, 2001.
- [31] K. Nahrstedt and L. Qiao, “Non-invertible watermarking methods for MPEG video and audio,” in *Proc. Security Workshop at ACM Multimedia*, pp. 93–98, Bristol, England, September 1998.
- [32] B. Schneier, *Applied Cryptography: Protocols, Algorithms, and Source Code in C*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1995.
- [33] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, New York, NY, USA, 3rd edition, 1991.
- [34] R. J. Clarke, *Transform Coding of Images*, Academic Press, New York, NY, USA, 1985.
- [35] E. Y. Lam and J. W. Goodman, “A mathematical analysis of the DCT coefficient distributions for images,” *IEEE Trans. Image Processing*, vol. 9, no. 10, pp. 1661–1666, 2000.
- [36] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes*, Chapman and Hall, New York, NY, USA, 1994.
- [37] Q. Cheng and T. S. Huang, “A DCT-domain blind watermarking system using optimum detection on Laplacian model,” in *Proc. IEEE International Conference on Image Processing*, vol. 1, pp. 454–457, Vancouver, BC, Canada, September 2000.
- [38] A. Briassouli, P. Tsakalides, and A. Stouraitis, “Hidden messages in heavy-tails: DCT-domain watermark detection using alpha-stable models,” to appear in *IEEE Transactions on Multimedia*.
- [39] A. Eleftheriadis and D. Anastassiou, “Constrained and general dynamic rate shaping of compressed digital video,” in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, pp. 396–399, Washington, DC, USA, October 1995.
- [40] R. J. Safranek, C. R. Kalmanek, and R. Garg, “Methods for matching compressed video to ATM networks,” in *Proceedings of IEEE International Conference on Image Processing*, vol. 1, pp. 13–16, Washington, DC, USA, October 1995.
- [41] “MPEG Software Simulation Group (MSSG),” <http://www.mpeg.org/MPEG/MSSG/>.
- [42] I. J. Cox, M. L. Miller, and J. A. Bloom, “Watermarking applications and their properties,” in *Proc. International Conference on Information Technology: Coding and Computing*, pp. 6–10, Las Vegas, Nev, USA, March 2000.
- [43] F. Davoine, “Triangular meshes: a solution to resist to geometric distortions based watermark-removal softwares,” in *Proc. European Signal Processing Conference*, vol. 4, Tampere, Finland, September 2000.
- [44] E. Izquierdo, “Using invariant image features for synchronization in spread spectrum image watermarking,” *EURASIP Journal on Applied Signal Processing*, vol. 2002, no. 4, pp. 410–417, 2002.

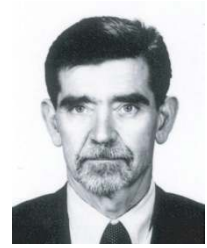
Dimitrios Simitopoulos was born in Greece in 1977. He received the Diploma and the Ph.D. degrees from the Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece, in 1999 and 2004, respectively. From 1999 to 2003, he held teaching and research assistantship positions in Aristotle University of Thessaloniki. Dr. Simitopoulos has participated in several research projects funded by the European Union (EU) and the General Secretariat for Research and Technology (GSRT). He is currently a researcher at the Informatics and Telematics Institute, Thessaloniki, Greece. His research interests include watermarking, multimedia security, and image indexing and retrieval. He is a Member of the Technical Chamber of Greece.



Sotirios A. Tsafaris was born in Thessaloniki, Greece in 1978. In June 2000, he received the Diploma of Electrical and Computer Engineering from the Department of Electrical and Computer Engineering of the Aristotle University of Thessaloniki. Since 2001, he has been with the Department of Electrical and Computer Engineering at Northwestern University in Evanston, Ill, working towards his Ph.D. degree. He received the M.S. degree in electrical and computer engineering from Northwestern University in June 2003. In the period 2000–2001, he held a research position in the Informatics and Telematics Institute working on copyright management of digital media. At Northwestern University, he is with the Image and Video Processing Laboratory working on biomolecular computing applications and bioinformatics. He held research and teaching assistantship positions and was the recipient of the Murphy Fellowship (2001–2002). His work is also funded by the Alexander S. Onassis Postgraduate Fellowship (2001–present) from the Alexander S. Onassis Public Benefit Foundation. His research interests include biomolecular computing and bioinformatics, microarray image compression, digital media rights, and image and video processing. He is a Member of the IEEE and the Technical Chamber of Greece.



Michael G. Strintzis received the Diploma degree in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1967, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, in 1969 and 1970, respectively. He then joined the Electrical Engineering Department at the University of Pittsburgh, Pittsburgh, Pa, where he served as Assistant Professor (1970–1976) and Associate Professor (1976–1980). Since 1980, he has been Professor of electrical and computer engineering at the University of Thessaloniki, Thessaloniki, Greece, and, since 1999, Director of the Informatics and Telematics Research Institute, Thessaloniki. His current research interests include 2D and 3D image coding, image processing, biomedical signal and image processing, and DVD and Internet data authentication and copy protection. Dr. Strintzis is currently serving as Associate Editor of the EURASIP Journal on Applied Signal Processing, the IEEE Transactions on Circuits and Systems for Video Technology, and the IEEE Transactions on Circuits and Systems I.



Nikolaos V. Boulgouris received the Diploma and the Ph.D. degrees from the Department of Electrical and Computer Engineering, University of Thessaloniki, Greece, in 1997 and 2002, respectively. Since September 2003, he has been a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, Canada. Formerly, he was a researcher in the Informatics and Telematics Institute, Thessaloniki, Greece. During his graduate studies, Dr. Boulgouris held several research and teaching assistantship positions. Since 1997, he has participated in research projects in the areas of image/video communication, pattern recognition, multimedia security, and content-based indexing and retrieval. Dr. Boulgouris is a Member of the IEEE and the Technical Chamber of Greece.



Alexia Briassouli obtained the Diploma degree in electrical engineering from the National Technical University of Athens (NTUA) in 1999 and the M.S. degree in image and signal processing from the University of Patras in 2000. From 2000 to 2001 she worked as a Research Assistant at the Informatics and Telematics Institute, Center of Research and Technology Hellas (CERTH), Thessaloniki, participating in a European funded research project. She is currently pursuing her Ph.D. degree in electrical engineering at the University of Illinois at Urbana-Champaign. Her research interests lie in the fields of statistical signal processing and image processing. She has worked on the design of optimal watermark embedding and detection systems for images and video that are robust to various attacks. Her current research interests lie in the areas of statistical image processing and computer vision, and include problems like motion estimation and segmentation for video.

