The ISC Int'l Journal of Information Security

ISeCure

January 2013, Volume 5, Number 1 (pp. 83–95)

http://www.isecure-journal.org

Robust Multiplicative Video Watermarking Using Statistical Modeling

Abolfazl Diyanat [1,*], Mohammad Ali Akhaee [1], and Shahrokh Ghaemmaghami [2,3]

[1] *Department of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran*

[2] *Department of Electrical Engineering, Sharif University of Technology Tehran, Iran*

[3] *Electronics Research Institute, Sharif University of Technology Tehran, Iran*

ARTICLE INFO.

Article history:

Received: 5 January 2013

Revised: 26 March 2013

Accepted: 15 April 2013

Published Online: 4 August 2013

Keywords:

Multiplicative Video Watermarking, Maximum Likelihood Decoding, 3D Wavelet Transform.

ABSTRACT

The present paper is intended to present a robust multiplicative video watermarking scheme. In this regard, the video signal is segmented into 3-D blocks like cubes, and then, the 3-D wavelet transform is applied to each block. The low frequency components of the wavelet coefficients are then used for data embedding to make the process robust against both malicious and unintentional attacks. The hidden message is inserted through multiplying/dividing these coefficients by a constant parameter which controls the power of the watermark. The watermark extraction relies on a maximum likelihood-based procedure, observing the distribution of the watermarked coefficients. The performance of the proposed scheme has been verified via simulations and found to be superior to some of the well-known existing video watermarking methods.

1 Introduction

Watermarking has been proposed as an elegant solution for the purpose of copyright protection, where it has also been found to be an efficient solution to several other problems in copy control, broadcast monitoring, fingerprinting, signal and data authentication, etc. [1]. Among the media types, image signals have been of special concern for copyright protection and authentication through watermarking. Nevertheless, by development of new handsets and their ability in transmitting and capturing video signals over the webs, the task of video watermarking is getting more demanding.

As a video is known as a moving picture signal in nature, therefore, the methods used for watermarking of still image may actually be extended to video watermarking as well. However, this extension is technically rejected for certain reasons, as: i) a video signal generally contains sequences of highly correlated frames, ii) there exist some video-based attacks such as MPEG compression, spatial desynchronization, frame collision, etc., and iii) video signals are often used in real-time applications and hence require real-time watermarking methods in most cases [2]

In this regard, there are two basic approaches to video watermarking, namely, watermarking in the compressed domain [3–8] or in the uncompressed domain [9–11]. The compressed domain watermarking applies to the embedding procedures in compressed domain, without any decompression/recompression . For example, Belhaj in [3] has embedded a message in the MPEG-4 using QIM (quantization index modulation) scheme based on the perceptual masking. Similarly, Langelaar et al. in [6] proposed two methods for embedding the message bits directly into an MPEG compressed video bitstream. The first method water-

* Corresponding author.

Email addresses: a.diyanat@ut.ac.ir (A. Diyanat), akhaee@ut.ac.ir (M.A Akhaee), ghaemmagh@sharif.edu (S. Ghaemmaghami).

ISSN: 2008-2045

marks the signal by changing the variable length codes in the bitstream, while the second discards some of the high frequency DCT (discrete cosine transform) coefficients of the bitstream for data hiding. Video watermarking using motion vectors (MV) has been discussed in [12, 13]. These watermarking algorithms are suitable for real-time video applications, though they are mostly restricted to specific video compression standards. Biswas et al. have embedded several binary images decomposed from a single watermarked image into different scenes in a video sequence [8]. Barakli has also proposed a new reversible watermaking algorithm based on motion compensated interpolation error [14].

Despite these kinds of watermaking, embedding the watermark in the uncompressed domain enjoys the advantage that the watermarked video can usually undergo standard compression processes, within a reasonable range of different data rates, without losing the mark. However, the embedded watermark must be resistant to compression attacks. The spatial domain [9–15] or the transformed domains [16], such as DWT (discrete wavelet transform) [10], [17–20] DFT (discrete Fourier transform)[11], and DCT [21, 22], can be used to watermark data in uncompressed domain. The spatial domain schemes are the simplest watermarking methods yet the watermark may be easily erased by lossy video compression. Conversely, in the transform domain watermarking, embedding the watermark into the transform coefficients can yield higher robustness against watermarking attacks [23]. The method proposed here falls into this category.

Given the present study, the 3D-DWT coefficients were used to embed the watermarks. Among image and video watermarking methods, several schemes take advantage of this transform. Chan and Lyu, for instance, have embedded different parts of a single watermark into different scenes of a video in the wavelet domain. The watermark is embedded into video frames by changing the positions of some DWT coefficients according to specific rules [10]. Guo-juan in [24] proposes a blind video watermarking based on a combination of Zernike moments and singular value decomposition (SVD). In this method, the SVD is applied to the low DWT coefficients and the message is embedded by modifying the maximum singular value in each frame. Elsewhere, Wang et al. apply the DWT to each frame of a video signal, and then use a QIM algorithm to embed the message into some of the high frequency DWT coefficients [25]. Kothari in [16], extracts the frames from the video and then uses the frequency domain characteristics of the frames for watermarking.

Watermarking systems can be categorized into additive and non-additive methods based on the embedding rule. In the additive case, the watermark is added to a set of image features, such as gray level values of pixels or frequency coefficients [5, 26, 27]. In the non-additive watermarking though, depending on the host characteristics, the embedding process is performed which results in better robustness and better use of the human visual system characteristics [28]. These approaches often make use of the video data in a transform domain [29]. Multiplicative watermarking methods are well-known examples of non-additive data hiding methods.

A correlation detector is used for multiplicative watermarking in [29]; nonetheless, this type of detection is not suitable for the transform domain watermarking. Hence, several alternative decoders have been proposed [30–33] to overcome this constraint. In [30], for example, in order to improve the performance of the correlation-based watermark recovery in the DFT domain, a new watermark Neyman-Pearson criterion. In a similar vein, Wang in [31] employed the DWT coefficients detection algorithm is proposed that is optimal under modeled with the Generalized Gaussian Distribution (GGD). He proposed a locally optimum detector for the $Barni\hat{a}\in^{TM}s$ multiplicative watermarking. In most cases, the transform domain coefficients are assumed to be i.i.d., while it is not usually true.

The present paper aims to introduce a video watermarking technique which is highly robust to video-based attacks in the uncompressed domain. To insert the watermark, the scaling based rule proposed in [28, 34, 35] has been used for low frequency components of the 3-D DWT video blocks. Considering the distribution of the watermarked approximation coefficients, the maximum likelihood (ML) decoder has been applied for data extraction. To this end, a number of assumptions have also been made on both the embedding parameter and the channel noise to simplify the detection process, leading to a real-time decoding scheme that is highly demanding. It is noteworthy that embedding in the low frequency components of video signals, as well as optimal detection make this algorithm favorably robust against typical attacks.

Thus the rest of the paper is organized as follows. In Section 2, we introduce our watermarking scheme. In fact, we attempt to describe how embedding and detection procedures are conducted in our proposed system. Performance analysis of the proposed method is presented in Section 3. In Section 4 the simulation results are reported the robustness of the proposed approach against common attacks are being discussed in detail. Finally, Section 5 concludes the paper.

2 Proposed Scheme

Consider a binary message (M) to be embedded into a host video signal of the uncompressed AVI format. At the encoder side, using the multiplicative rule, we embed the message bits in the host video [28, 29], as detailed in Section 2.1. The host signal is then sent through a communication channel and is assumed to be corrupted by additive white Gaussian noise (AWGN). At the decoder side, an ML decoder is designed to achieve the optimal detection in the presence of AWGN (See Section 2.2). For convenience, the notations are listed in Table 1.

Table 1. Notation.

Notation	Description
$B \times B \times T$	Non-ovelappint 3D block Dimensions
t	Total video frames
$W \times H$	Video frame dimensions
N	Number of wavelet approximation coefcients
M	Message
L	Message length
U_i	LLL DWT coefficient before embedding
$U_i^{'}$	LLL DWT coefficient after embedding
α	Strength factor
σ^2	Variance of 3D block before embedding
μ	Average of 3D block before embedding
W_i	LLL WDT coefficient In the receiver side
σ_n^2	Guassian white noise variance in the channel
$\sigma^2_{W\|x}$	Variance of W_i with condition x
$\mu^2_{W\|x}$	Average of W_i with condition x

2.1 Watermark Embedding

2.1.1 Video Segmentation

To embed the watermark, first the video signal was segmented into non-overlapping 3D blocks. Each 3D block has the size of $B \times B \times T$, where $B \times B$ denotes the non-overlapping pixel size in the spatial domain within each frame and T is the number of consecutive frames over the temporal domain. Accordingly, for a signal of the total length of t with a frame size of $H \times W$, the capacity of our watermarking system is achieved as:

$$Capacity = \lfloor \frac{H}{B} \rfloor \times \lfloor \frac{W}{B} \rfloor \times \lfloor \frac{t}{T} \rfloor \qquad (1)$$

2.1.2 3D Block Selection

In the next step, high entropy blocks of the video signal, which are more suitable for data hiding, are identified. Insertion of a random message into the host signal may be modeled by adding noise to the signal, according to the model introduced in [36]. Watermark embedding in high entropy blocks can reduce both visual and statistical footprint effects. The other reason for choosing such blocks is to take the advantage of the entropy masking of the Watson's visual model that accounts for lower sensitivity of the human visual system to the crowded regions [37]. However, as will be discussed later, we prefer to embed the message bits in all blocks with different powers that are adjusted based on an optimality criterion. This guarantees highest robustness against the desynchronization attacks, while maintaining the visual imperceptibility of the watermark.

2.1.3 Data Embedding

As mentioned earlier, to achieve higher robustness, the low frequency components are used to embed the watermark, as they suffer the least from the changes made by the compression or filtering attacks. However, special care is to be taken to keep the watermark invisible, due to the high sensitivity of the visual system to modification of these components. As for the proposed scheme, taking advantage of the wavelet transform, we have decomposed each 3-D block into different subbands and the approximation coefficients are used for data embedding. The 3D wavelet coefficients are computed by applying the 1D wavelet transform to the wavelet coefficients of consecutive frames at the same scale/position. From now on,the low frequency subband of the 3D wavelet transform will be referred to as LLL. Denoting these LLL approximation coefficients by u_i, the embedding process is performed using the following scaling based rule [28]

$$\begin{aligned} W_i &= U_i.\alpha, \qquad M = 0 \\ W_i &= U_i.\frac{1}{\alpha}, \qquad M = 1 \end{aligned} \qquad (2)$$

In the above equation, the parameter α is called the *strength factor* which controls the power of the watermark. The index i denotes the i'th coefficient of a 3D block. Should this parameter be adjusted appropriately, the blocking effect does not occur and consequently the watermarking would be kept transparent [34]. In fact, α makes a trade-off between the quality of the watermarked signal and its robustness against attacks. Figure 1 demonstrates the block diagram of the proposed data embedding process.

2.2 Watermark Extraction

2.2.1 Coefficients Distribution

A similar procedure is to be followed for data extraction. First, the received video signal is segmented into the 3D blocks. Then, applying the 3D wavelet transform to the blocks, the watermarked approximation coefficients are attained. To detect the watermark bits, the ML detector is used based on the distribution of the watermarked approximation coefficients of the 3D video blocks. In [38] these coefficients are modeled with the GGD. Mihcak, et al. show that the wavelet coefficients of an image can be modeled as a Gaussian process [39]. Besides, using Kolmogrov-Smirnove test, Akhaee et al. have illustrated that the approximation coefficients of image signals can be well-modeled with an i.i.d Gaussian distribution [28]. The same model can be assumed for the present work as well, as we use the uncompressed AVI signals that contain consecutive images. In the same vein, Petrosian and Meyer have shown in [40], that this assumption can be regarded as quite accurate for this work, as can also be viewed in Figure 2

The decoding procedure is represented in Figure 3. Let's assume u_i be the approximation coefficients of the video blocks with the mean μ and variance σ^2. Embedding the message bit in each block, these parameters are multiplied or divided by σ and σ^2 respectively. After transmitting the signal over the channel, the watermarked signal might be contaminated with AWGN. Since the noise is Gaussian and uncorrelated to the watermarked video signal, the received signal remains Gaussian with the following distribution:

$$W_i|0 = \alpha.U_i + n_i \Rightarrow W_i|0 \backsim \mathcal{N}(\mu_{|0}, \sigma^2_{W|0}) \quad (3)$$

$$W_i|1 = \alpha^{-1}.U_i + n_i \Rightarrow W_i|1 \backsim \mathcal{N}(\mu_{|1}, \sigma^2_{W|1}) \quad (4)$$

Where $\sigma^2_{W|0} = \alpha^2\sigma^2 + \sigma^2_n$, $\sigma^2_{W|1} = \alpha^{-2}\sigma^2 + \sigma^2_n$, $\mu_{|1} = \alpha^{-1}\mu$, and $\mu_{|0} = \alpha\mu$. These parameters must be known at the decoder side, and some of these parameters should be sent as side information, along with the watermarked signal. The other parameters, like the noise variance, can be estimated at the decoder.

2.2.2 Noise Estimation

There are several techniques to estimate the noise variance through the channel [42, 43]. However, due to the wavelet transform used in this work, we choose the technique proposed by *Donoho* in [44] employing detail coefficients. To this end, the variance of the noise is estimated, applying the median filter to detail coefficients (HH) of a given image as:

$$\hat{\sigma} = \frac{Median(|W_i|)}{0.6745}, \qquad W_i \in subbandHH \quad (5)$$

The noise variance is estimated using (5) for all wavelet subbands, where just the part of the noise added to LL coefficients is of concern to us. In other words, we have to estimate the noise variance in the subband to which the decoder is applied. Consequently, the standard deviation of the noise for the LL coefficients is computed by multiplication of the norm of the LL filter impulse response as:

$$\sigma_{\mathrm{n}} = \| \mathrm{LL} \| \hat{\sigma} \qquad \| \mathrm{LL} \| = \sqrt{\Sigma_{\mathrm{l}}\Sigma_{\mathrm{k}}\mathrm{LL}^2(\mathrm{l},\mathrm{k})} \quad (6)$$

It should be noted that in our case, the average of the noise variance in each frame is used to estimate the effective noise variance for each video segment.

2.2.3 ML Decoder

The distribution of N approximation coefficients of the video blocks after embedding the bit 0 or 1 can be calculated as:

$$\mathrm{P}(\mathrm{W}_1, \ldots, \mathrm{W}_\mathrm{N}|0) = \prod_{\mathrm{i}=1}^{\mathrm{N}} \frac{1}{\sqrt{2\pi\sigma^2_{\mathrm{W}|0}}} . \mathrm{e}^{-(\mathrm{W_i} - \alpha\mu_{|0})^2/2.\sigma^2_{\mathrm{W}|0}} \quad (7)$$

$$\mathrm{P}(\mathrm{W}_1, \ldots, \mathrm{W}_\mathrm{N}|1) = \prod_{\mathrm{i}=1}^{\mathrm{N}} \frac{1}{\sqrt{2\pi\sigma^2_{\mathrm{W}|1}}} . \mathrm{e}^{-(\mathrm{W_i} - \alpha^{-1}\mu_{|1})^2/2.\sigma^2_{\mathrm{W}|1}} \quad (8)$$

To extract the data, the ML rule can be used as:

$$P(W_1, W_2, \ldots, W_N|0) \gtrless_1^0 P(W_1, W_2, \ldots, W_N|1) \quad (9)$$

Substituting (7) and (8) in (9), we have:

$$\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2_{W|0}}} . e^{-(W_i - \alpha^2\mu)^2/2.\sigma^2_{W|0}} \gtrless_1^0 \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2_{W|1}}} . e^{-(W_i - \alpha^{-2}\mu)^2/2.\sigma^2_{W|1}} \quad (10)$$

Taking logarithmic function from both sides and doing some simplifications, (10) can be rewritten as:

$$\left(\frac{1}{\sigma^2_{W|1}} - \frac{1}{\sigma^2_{W|0}}\right) \sum_{i=1}^{N} W_i^2 - 2\mu \left(\frac{\alpha^{-2}}{\sigma^2_{W|1}} - \frac{\alpha^2}{\sigma^2_{W|0}}\right) \sum_{i=1}^{N} W_i \gtrless_1^0 2N \ln\left(\frac{\sigma_{W|0}}{\sigma_{W|1}}\right) - N\mu^2\left(\frac{\alpha^{-4}}{\sigma^2_{W|1}} - \frac{\alpha^4}{\sigma^2_{W|0}}\right) \quad (11)$$

For the case that the variance of the noise is far smaller than the variance of the approximation coeffi-

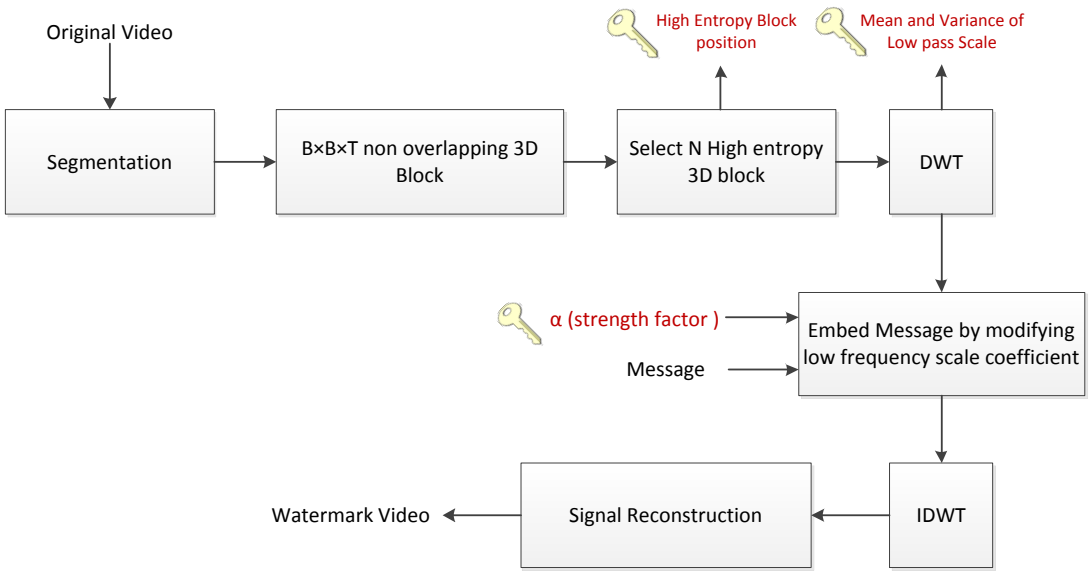


Figure 1. Block diagram of the proposed watermarking scheme.

cients of the video block, which usually happens, (11) can be simplified as:

$$(\alpha^4 - 1)(\Sigma_{i=1}^{N} W_i^2 - N\mu^2) \gtrless_1^0 4N\alpha^2\sigma^2 \ln(\alpha) \quad (12)$$

Besides, the transparency of the watermark limits the value of α, to be close to one. Thus, considering $\alpha = 1 + \epsilon$, where ϵ has a small value, equation (12) can be further simplified as:

$$\Sigma_{i=1}^{N} W_i^2 \gtrless_1^0 N(\sigma^2 + \mu^2) = N \times E\{U^2\} \quad (13)$$

According to (13), the proposed detector is independent of the strength factor in low noisy environments. Additionally, there is no need to send both mean and variance separately. In fact, the value of $\sigma^2 + \mu^2$, the second moment of the approximation coefficients, is enough to be sent through the secure channel.

2.3 Key Length

As mentioned earlier, in the detection process, some side information is required for data extraction. This side information, transmitted along with the watermarked video, includes the strength factor and the second moment of the approximation coefficients of the 3D video blocks. The size of this information can be reduced using a constant strength factor, at the expense of higher error rate in the watermark extraction and/or violating the watermark imperceptibility. Therefore, we prefer to choose the value of α among pre-set values which are available at the decoder. The total size of this information after compression and scrambling is 0.01% of the original video size on average. For instance, for a video size of 10 M-byte, just 1 K-byte scrambled data will transmit all the side information required for decoding the watermark.

3 Performance Analysis

Assume the bit 0 is sent with the probability of p_0 and 1 with that of p_1. We introduce a new random variable to calculate error probability, as:

$$\begin{aligned} & P\left(W_1, W_2, \ldots, W_N|1\right) \\ & \gtrless \_0\hat{}1\ P(W\_1, W\_2, \ldots, W\_N\ |0) \\ & \xi = \frac{P(W_i|1)}{P(W_i|0)} \gtrless_0^1 1 \end{aligned} \quad (14)$$

Using the above equation, the error probability can be computed as follows:

$$P_e = P_0 \int_1^{+\infty} f_\xi\left(\xi|0\right) \mathrm{d}\xi + P_1 \int_{-\infty}^{1} f_\xi(\xi|1)\mathrm{d}\xi \quad (15)$$

Since there is no closed-form solution for (15), the problem has to be investigated in the low noise condition where a closed-form relation can be found. The ML decoder in low noise environment can be obtained from (12). Following some simplification, we will arrive at:

$$\sum_{i=1}^{N} W_i^2 \gtrless_1^0 \left(\frac{4N\alpha^2\sigma^2 ln\ (\alpha)}{\alpha^4 - 1} + N\mu^2\right) \quad (16)$$

Based on (14), the distribution of two conditional random variables can be defined as:

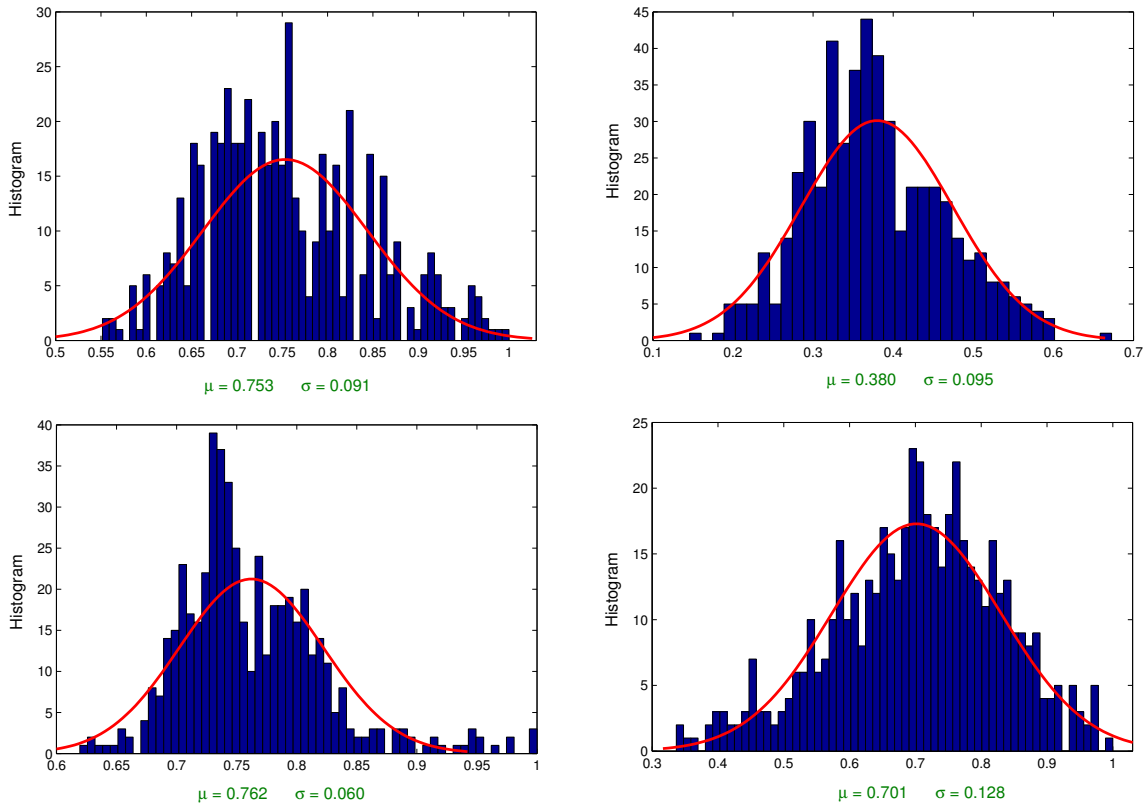


Figure 2. Histogram of approximation wavelet coefficients for 3D block of 4 video in *Hollywood2* database [41].

$$\begin{aligned}\xi|0 &= \sum_{i=1}^{N}(\alpha U_i + n_i)^2 \\ &= \alpha^2\sum_{i=1}^{N}U_i^2 + \sum_{i=1}^{N}n_i^2 + 2\alpha\sum_{i=1}^{N}U_i n_i \\ \xi|1 &= \sum_{i=1}^{N}(\alpha^{-1}U_i + n_i)^2 \\ &= \alpha^{-2}\sum_{i=1}^{N}U_i^2 + \sum_{i=1}^{N}n_i^2 + 2\alpha^{-1}\sum_{i=1}^{N}U_i n_i \end{aligned} \quad (17)$$

As mentioned earlier, W_i is of the Gaussian distribution. The distribution of $\xi|0$ or $\xi|0$ may seem to be a chi-square distribution with N degrees of freedom ($\chi^2(N)$). However, this is not true, since the term $\sum_{i=1}^{N} W_i^2$ is a known parameter at the receiver.

Here, we just derive the distribution of $\xi|0$, where the same procedure can be used to compute $\xi|1$. The first term in equation 12, in the case of '0' embedding, is equal to $N\alpha^2(\mu^2+\sigma^2)$. In the next step, the distribution of the second and third terms is found. According to the Central Limit Theorem (CLT) and since the number of samples in each block (N) is large enough, the second and third terms can be modeled by a Gaussian distribution, regardless of the type of the distribution of the host signal.

$\varphi = 2\sum_{i=1}^{N} U_i n_i$ with the Gaussian distribution comes with the following properties.

$$\begin{aligned} E(\varphi) &= \sum_{i=1}^{N}\mathcal{E}(U_i n_i) = 0 \\ Var(\varphi) &= 8N\alpha^2\sigma_n^2(\sigma^2+\mu^2) \end{aligned} \quad (18)$$

variable$= \sum_{i=1}^{N} n_i^2$, we have:

$$\begin{aligned} E(\phi) &= \sum_{i=1}^{N}\mathcal{E}(n_i^2) = N\sigma_n^2 \\ Var(\phi) &= 2N\sigma_n^4 \end{aligned} \quad (19)$$

where [45]:

$$\mathrm{E}\left[(X-\mu)^p\right] = \begin{cases} 0 & \text{if } p \text{ is odd,} \\ \sigma^p\,(p-1)!! & \text{if } p \text{ is even.} \end{cases}$$

Here $n!!$ denotes the double factorial. Suppose f($\xi|0$) and f($\xi|1$) are probability density functions of $\xi|0$ and $\xi|1$, respectively. According to (18) and (19), we have:

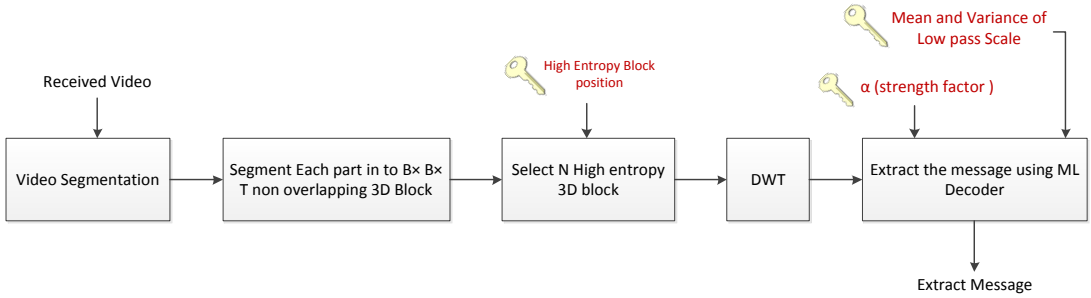


Figure 3. Block diagram of the proposed watermarking decoder.

Table 2. Effects of block size and message length (PSNR and BER (%) after some attack).

	8 × 8 × 8				16 × 16 × 16				16 × 16 × 16			
$\frac{MessageLength}{Capacity}$	PSNR	Noise	Wiener	Median	PSNR	Noise	Wiener	Median	PSNR	Noise	Wiener	Median
0.2	54.08	23.07	18.34	14.05	53.82	5.40	5.12	7.56	54.16	0.51	0.80	6.03
0.3	52.77	24.52	19.76	13.13	52.54	7.28	6.03	7.16	52.74	0.48	1.64	6.33
0.4	51.88	25.55	19.70	12.21	51.68	7.90	7.11	6.58	51.84	1.00	1.83	5.96
0.5	51.23	26.93	20.82	11.15	51.06	9.01	7.60	6.15	51.20	1.49	2.38	5.95
0.6	50.70	27.24	20.58	10.63	50.60	10.17	8.12	5.73	50.73	1.46	3.14	6.60
0.7	50.30	28.08	20.46	9.88	50.22	11.13	8.77	5.54	50.34	2.22	3.09	5.89
0.8	49.97	28.82	20.10	9.25	49.90	12.09	8.77	5.26	50.02	2.51	3.50	5.51
0.9	49.69	29.24	19.24	8.68	49.63	13.00	8.77	4.99	49.75	3.77	4.01	5.87
1.0	49.47	29.91	18.80	8.52	49.39	14.03	8.58	4.83	49.48	4.81	4.34	5.42

$$f(\xi|0) = \mathcal{N}(N\sigma_n^2 + N\alpha^2(\mu^2+\sigma^2), \\ 8N\alpha^2\sigma_n^2(\sigma^2+\mu^2) + 2N\sigma_n^4) \\ f(\xi|1) = \mathcal{N}(N\sigma_n^2 + N\alpha^{-2}\left(\mu^2+\sigma^2\right), \\ 8N\alpha^{-2}\sigma_n^2\left(\sigma^2+\mu^2\right) + 2N\sigma_n^4) \quad (20)$$

Finally, the error probability can be calculated as:

$$P_e = P_0\frac{1}{\sqrt{2\pi\sigma_0^2}}\int_{-\infty}^{\theta}\exp(\frac{(\xi-\mu_0)^2}{2\sigma_0^2})\mathrm{d}\xi \quad (21)$$
$$+P_1\frac{1}{\sqrt{2\pi\sigma_1^2}}\int_{-\infty}^{\theta}\exp(\frac{(\xi-\mu_1)^2}{2\sigma_1^2})\mathrm{d}\xi$$

where:

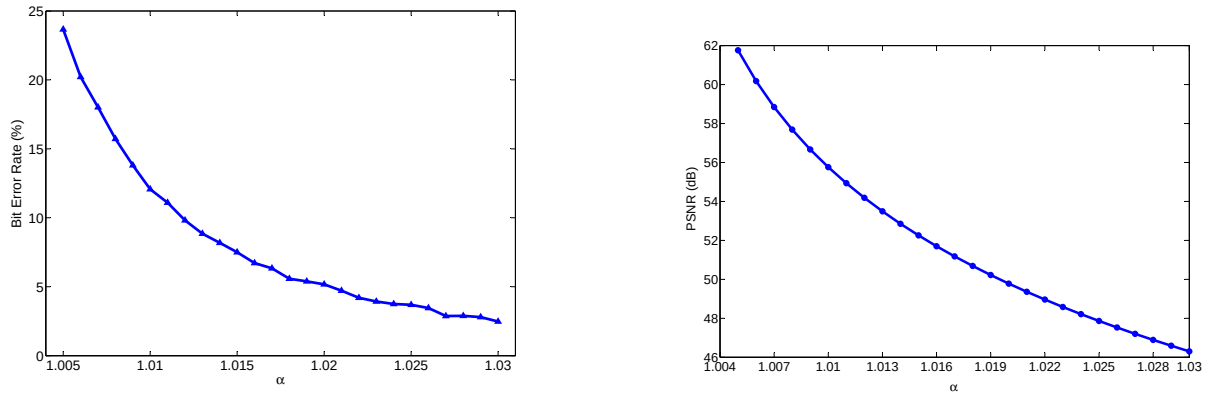
$$\mu_0 = N\sigma_n^2 + N\alpha^2(\mu^2+\sigma^2)$$
$$\sigma_0^2 = 8N\alpha^2\sigma_n^2(\sigma^2+\mu^2) + 2N\sigma_n^4$$
$$\mu_1 = N\sigma_n^2 + N\alpha^{-2}(\mu^2+\sigma^2)$$
$$\sigma_1^2 = 8N\alpha^{-2}\sigma_n^2(\sigma^2+\mu^2) + 2N\sigma_n^4$$
$$\theta = \frac{4N\alpha^2\sigma^2 ln\ (\alpha)}{\alpha^4-1} + N\mu^2$$

We can express the equation (21) with Q-function (the tailprobability of the standard normal distribution), as:

$$P_e = P_0(1-Q(\frac{\theta-\mu_0}{\sigma_0})) + P_1Q(\frac{\theta-\mu_1}{\sigma_1}) \quad (22)$$

4 Experiment Result

In this section, the performance of the proposed algorithm is being discussed using various video signals in the uncompressed AVI format. Hollywood2 has been made use of for the purpose of the experiments [41]. The 3D wavelet transform is computed through a 2D wavelet transform applied to each frame, followed by a 1D wavelet transform of the coefficients of similar positions along the time axis over a number of consecutive frames. The In this section, the performance of the proposed algorithm is being discussed using various video signals in the uncompressed AVI format. Hollywood2 has been made use of for the purpose of the experiments [41]. The 3D wavelet transform is computed through a 2D wavelet transform applied to each frame, followed by a 1D wavelet transform of the coefficients of similar positions along the time axis over a number of consecutive frames. The coefficients are then obtained from a three-level decomposition us-



(a) BER(%) (b) PSNR(db)

Figure 4. The effect of α on BER and PSNR after white noise with standard deviation of 30 attack.

ing a Daubechies filter of the length two. The message is then embedded into the approximation coefficients (LLL).

Letting the width and the height of a video signal be W and H, respectively, we segment the host signal into 3-D blocks of the size $B \times B \times T$. The capacity of our watermarking system can be obtained from (1). As stated earlier, t frames of the video signal are divided into several segments with T frames in each segment.

There exists a trade-off between the quality of the watermarked video and the robustness against attacks. Here, the quality is measured with the peak signal to noise ratio (PSNR) as:

$$PSNR = 10\log_{10}\left(\frac{255^2}{E(X-X')^2}\right) \qquad (23)$$

where X and X' represent the original and the watermarked signals, respectively. In addition, the error probability is calculated as the ratio of the number of error bits to the total number of the watermark bits are transmitted through the channel.

4.1 Variation of the Block Size and the Message Length

As the first experiment, the effect of message length is examined through changing the 3D block size. To measure the effect of attacks, three types of attack, namely, AWGN with the standard deviation of 30, median filtering (3×3 window), and Wiener filtering (5×5 window) are studied here. Table 2 represents the average bit error rate (BER) for 65 video signals. The strength factor α is set to 1.015. As displayed, the best performance (in terms of capacity, robustness, and transparency) is achieved when the block size is $16 \times 16 \times 16$. Therefore, this block size is used for the rest of our simulations.

4.2 The Effect of α

In order to investigate the effect of the watermark power on the robustness, we have changed the value of α from 1.005 to 1.035. The white noise with standard deviation of 30 is used as the attack to 16 frames of each video signal. The results, averaged over 100 video files, are reported in 4a and 4b. The total number of watermarked bits is $0.5 \times Capacity$ for each video signal. As expected, increasing the strength factor results in higher robustness against noise attack and less quality of the watermarked signal. Table 3 illustrates a comparison between several methods, based on the proposed data embedding scheme, using the Foreman video signal with $\alpha = 1.015$. [48] and [47] are two video watermarking which use DCT transfer to embed watermark into 4×4 video blocks. As can be observed, the presented method outperforms the aforementioned methods in terms of transparency.

Table 3. Comparison of the transparency.

method	Our method	method [46]	method [47]
PSNR (dB)	42.89	35.91	38.98

4.3 Noise Attack

In this experiment, white Gaussian noise with a mean of zero and different standard deviations (from 0 to 40) as well as uniform noise, with different standard deviations (from 30 to 80), are added to the watermarked signal. The average BER, over 77 video signals, is depicted in Figure 5 and Figure 6. As demonstrated, the proposed method is highly robust against such noise attacks. This is due to the use of optimal decoder as well as embedding the message bits in the low frequency components of the host signal. The value of α is set to 1.015, which guarantees the transparency of the watermark while keeping the performance at an

acceptable level (the tolerable BER in the multimedia applications). The length of the embedded message in this case is half of the total capacity of the video signal.

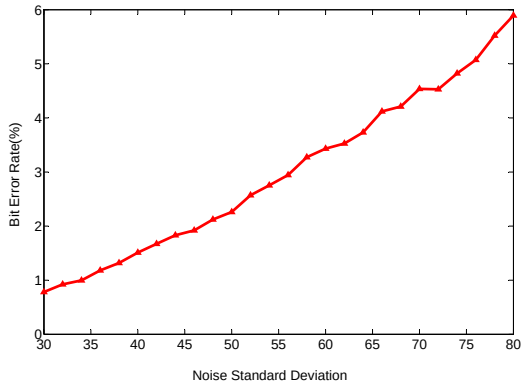


Figure 5. BER (%) after the Gaussian white noise attack.

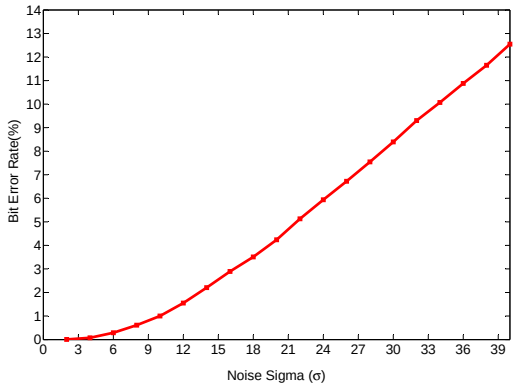


Figure 6. BER (%) after the uniform noise attack..

Table 4 allows a comparison between several methods of our scheme. The maximum capacity of Foreman video signal is used in each method which are then corrupted with the help of a Gaussian noise that leads to the PSNR of 30dB. [21], [49] and [50] employ wavelet transform, DCT and just noticeable difference (JND) respectively for embedding watermark in the video signal. Notice that all of the mentioned methods use an additive scheme for embedding. The superiority of our multiplicative video watermarking scheme over the abovementioned additive schemes canthus be easily observed.

Table 4. Results of Gaussian noise attack.

method	Our method	method [21]	method [49]	method [50]
BER (%)	0.8	6	4	1

4.4 Filtering Attacks

The resistance of the proposed method against several filtering attacks has also been investigated. In this experiment, the algorithm was tested over 80 video signals. The error rates are demonstrated in Table 5 after Gaussian filtering with different window sizes and sigma values was applied. Given the median, Wiener, and mean filtering, the results for several window sizes, i.e. $3 \times 3, 5 \times 5, 7 \times 7$ in are summarized in Table 6.

Table 5. BER (%) in case of Gaussian filter attack.

σ	3×3	5×5	7×7
0.1	0	0	0
0.2	0	0	0
0.3	0	0	0
0.4	0	0	0
0.5	0.115	0.118	0.075
0.6	0.377	0.353	0.419
0.7	0.895	1.16	1.28
0.8	1.57	2.32	2.46
0.9	1.91	3.40	3.86
1.0	2.17	4.88	5.39

Table 6. BER (%) after median and Wiener filtering attacks.

Filter Type	3×3	5×5	7×7
Median Filter	1.11	7.57	15.59
Average Filter	0.43	6.49	16.02
Wiener Filter	4.87	14.43	20.50

4.5 M-JPEG Compression Attack

Motion JPEG (M-JPEG) is a class of video formats where each video frame is separately compressed through JPEG compression. M-JPEG is used by many

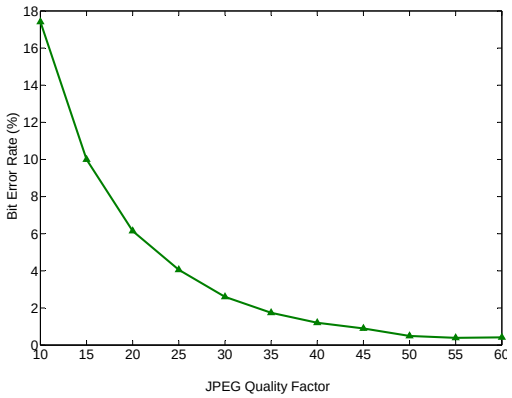


Figure 7. The BER (%) of the M-JPEG attack.

portable devices with video-capture capabilities, such as digital cameras. As for the purpose of the present study, we conducted a test over 75 video signals, in which the message bits were embedded and then the watermarked signals were compressed using M-JPEG, to evaluate the robustness of the proposed method against this compression attack. At the receiver, the message bits were extracted after decompressing the video signals. Figure 7 represents the bit error rate (BER) versus the JPEG quality factor between 10 and 60.

4.6 Runtime Analysis

To get an overall estimate of the latency of the proposed scheme and its major parts, we implemented the method using *MATLAB 2010a profiler* run on a 2.27 GHz, 64-bit quad-core Sony Vaio notebook, model VPC-CW2GGXB, with a 4-GB RAM anda 4-MB cache memory. Table 5 reports the time taken to run each of the stages dipicted in Figure 1 and Figure 3.

From Table 8, it is observed that the DWT and its inverse transform are the bottlenecks in our scheme, in terms of the processing time. However, there are real-time implementation algorithms found in the literature for these transforms including the ones introduced in [51, 52]. The computational complexity of the other parts of the proposed method is far from critical, making a real-time implementation of the system quite attainable.

Table 7. The BER (%) after MPEG compression attack.

α	1.015	1.016	1.017	1.019	1.020	1.021
BER (%)	11.13	7.22	3.90	0.59	0	0

5 Conclusion

The present paper attempted to present a new multiplicative watermarking scheme suitable for video signals in the AVI format. The data embedding is performed by slightly modifying the approximation coefficients of the 3D video blocks. Through modeling the modified noisy coefficients with Gaussian distribution, the optimal model was designed which implemented ML decoder for the watermark extraction. The performance of the proposed method is then analytically investigated. Experimental results over several video files revealed that the proposed method is highly robust against common attacks, including popular video compression methods such as MPEG and M-JPEG. As an extension of the current work, embedding the watermark directly in the compressed domain can be studied to reduce the computational cost while improving the robustness.

Table 8. Runtime analysis.

Operation	Time (second)
Encoding	
Video segmentation	0.02
Calculate 3D block entropy	0.15
Wavelet transfer	1.49
embeding	0.02
Inverse wavelet transform	1.18
Combine 3D block	0.03
Decoding	
Recieve video segmentation	0.02
Choose 3D block	0.03
Wavelet transfer	1.42
Decode message	0.01

References

[1] G. Döerr, "A Guide Tour of Video Watermarking," Signal Processing:Image Communication, vol. 18, pp. 263-282, Apr. 2003.

[2] Y. Chen and H. Huang, "A New Shot-Based Video Watermarking," inComputer Communication Control and Automation (3CA), InternationalSymposium on, vol. 2, pp. 53-58, 2010.

[3] M. Belhaj, M. Mitrea, F. Preteux, and S. Duta, "MPEG-4 AVC robustvideo watermarking based on QIM and perceptual masking," in Communications (COMM), 8th International Conference on, pp. 477-480,2010.

[4] D. Xu, R. Wang, and J. Wang, "Low complexity video watermarking algorithm by exploiting CAVLC in H. 264/AVC," in WirelessCommunications, Networking and Information Security (WCNIS), IEEEInternational Conference on, pp. 411-415, 2010.

[5] L. Zhang, Y. Zhu, and L. L.-M. Po, "A novel watermarking scheme withcompensation in bitstream domain for H.264/AVC," IEEE InternationalConference on Acoustics, Speech and Signal Processing, pp. 1758-1761,Mar. 2010.

[6] G. Langelaar, R. Lagendijk, and J. Biemond, "Real-Time Labeling ofMPEG-2 Compressed Video," Journal of Visual Communication andImage Representation, vol. 9, no. 4, pp. 256-270, 1998.

[7] S. K. Bavipati and X. Su, "Secure Compressed Domain Watermarkingfor H.264 Video," Seventh International Conference on Information-Technology: New Generations, pp. 387-391, Apr. 2010.

[8] S. N. Biswas, S. Nahar, S. R. Das, E. M. Petriu,

M. H. Assaf, andV. Groza, "MPEG-2 digital video watermarking technique," in IEEEInternational Instrumentation and Measurement Technology ConferenceProceedings, pp. 225-229, IEEE, May 2012.

[9] R. Lancini, F. Mapelli, and S. Tubaro, "A Robust Video WatermarkingTechnique In the Spatial Domain," in Video/Image Processing andMultimedia Communications 4th EURASIP-IEEE Region 8 InternationalSymposium on VIPromCom, no. June, pp. 251-256, 2002

[10] P. Chan and M. Lyu, "A DWT-based digital video watermarkingscheme with error correcting code," in Proceedings of Fifth InternationalConference on Information and Communications Security, pp. 202-213,Springer, 2003.

[11] F. Deguillaume, G. Csurka, J. O'Ruanaidh, and T. Pun, "Robust 3D DFTVideo Watermarking," in Proceedings of IS & T/SPIE Electronic Imaging,vol. 3657, pp. 113-124, 1999.

[12] J. Zhang, J. Li, and L. Zhang, "Video watermark technique in motionvector," in Computer Graphics and Image Processing, Proceedings ofXIV Brazilian Symposium on, pp. 179-182, 2001.

[13] C. Kung, J. Jeng, Y. Lee, H. Hsiao, and W. Cheng, "Video watermarkingusing motion vector," in Proc. of 16th IPPR Conference on computervision, graphics and image processing, no. Cvgip, pp. 547-551, 2003.

[14] B. Barakli and C. Vural, "A new reversible video watermarking methodbased on motion compensated interpolation," in 20th Signal Processingand Communications Applications Conference (SIU), pp. 1-4, IEEE,Apr. 2012.

[15] B. Mobasseri, "Direct Sequence Watermarking of Digital Video UsingM-Frames," in Proceedings International Conference on Image Processing (ICIP-98), vol. 2, pp. 399-403, 1998.

[16] A. M. Kothari and V. V. Dwivedi, "Transform Domain Video Watermarking: Design, Implementation and Performance Analysis," inInternational Conference on Communication Systems and Network Technologies, pp. 133-137, IEEE, May 2012.

[17] S. a. M. Al-Taweel and P. Sumari, "Robust Video Watermarking BasedOn 3D-DWT Domain," in TENCON , IEEE Region 10 Conference,pp. 1-6, Nov. 2010.

[18] P. Campisi, "Video watermarking in the 3D-DWT domain using perceptual masking," in IEEE International Conference on Image Processing(ICIP), pp. 997-1000, 2005.

[19] R. Reyes, C. Cruz, M. Nakano-Miyatake, and H. Perez-Meana, "DigitalVideo Watermarking in DWT Domain Using Chaotic Mixtures," LatinAmerica Transactions, IEEE (Revista IEEE America Latina), vol. 8,no. 3, pp. 304-310, 2010.

[20] R. O. Preda, "Robust wavelet-based video watermarking scheme forcopyright protection using the human visual system," Journal of Electronic Imaging, vol. 20, p. 013022, Jan. 2011.

[21] M. Swanson and A. Tewfik, "Multiresolution Scene-Based Video Watermarking Using Perceptual Models," IEEE Journal on Selected Areasin Communications, vol. 16, pp. 540-550, May 2002.

[22] J. Sun, N. Yang, J. Liu, X. Yang, X. Li, and L. Zhang, "Videowatermarking scheme based on spatial relationship of DCT coefficients,"in Intelligent Control and Automation (WCICA), 8th World Congress on,pp. 56-59, 2010.

[23] E. E. Abdallah, A. Ben Hamza, and P. Bhattacharya, "Video watermarking using wavelet transform and tensor algebra," Signal, Image and VideoProcessing, vol. 4, pp. 233-245, Apr. 2009.

[24] X. Guo-juan and W. Rang-ding, "A Blind Video Watermarking Algorithm Resisting to Rotation Attack," International Conference onComputer and Communications Security, pp. 111-114, Dec. 2009.

[25] C.-X. Wang, X. Nie, X. Wan, W. B. Wan, and F. Chao, "A Blind VideoWatermarking Scheme Based on DWT," Fifth International Conferenceon Intelligent Information Hiding and Multimedia Signal Processing,vol. 1, pp. 434-437, Sept. 2009.

[26] R. C. Motwani, M. C. Motwani, B. D. Bryant, F. C. Harris Jr., and A. S.Agarwal, "Watermark Embedder Optimization for 3D Mesh ObjectsUsing Classification Based Approach," International Conference onSignal Acquisition and Processing, pp. 125-129, Feb. 2010.

[27] D. Pu, Y. Lu, and J. Dai, "Video watermarking approach based on temporal difference and discrete wavelet transform," in Computer Science andInformation Technology (ICCSIT), 3rd IEEE International Conferenceon, vol. 1, pp. 346-350, 2010.

[28] M. A. Akhaee, S. M. E. Sahraeian, B. Sankur, and F. Marvasti, "Robust Scaling-Based Image Watermarking Using Maximum-LikelihoodDecoder With Optimum Strength Factor," IEEE Transactions on Multimedia, vol. 11, pp. 822-833, Aug. 2009.

[29] I. I. Cox, J. Kilian, F. F. Leighton, and T. Shamoon, "Secure SpreadSpectrum Watermarking For Multimedia," Image Processing, IEEE-Transactions on, vol. 6, no. 12, pp. 1673-1687, 2002.

[30] M. Barni, F. Bartolini, A. D. Rosa, and A. Piva, "A new decoder forthe optimum recovery of non-

additive watermarks," IEEE transactionson image processing, vol. 10, pp. 755-66, Jan. 2001.

[31] J. Wang, G. Liu, Y. Dai, J. Sun, Z. Wang, and S. Lian, "Locally optimumdetection for Barni's multiplicative watermarking in DWT domain,"Signal Processing, vol. 88, pp. 117-130, Jan. 2008.

[32] S. Vassilios and P. Ioannis, "Optimal Detector for Multiplicative Watermarks Embedded in the DFT Domain of Non-White Signals," EURASIPJournal on Advances in Signal Processing, vol. 1900, no. 16, pp. 2522-2532, 2004.

[33] T. Ng and H. Garg, "Maximum Likelihood Detection in Image Watermarking Using Generalized Gamma Model," in Conference Record ofthe Thirty-Ninth Asilomar Conference onSignals, Systems and Computers, no. 2, pp. 1680-1684, 2006.

[34] M. Akhaee, N. KhademiKalantari, and F. Marvasti, "Robust Audio andSpeech Watermarking Using Gaussian and Laplacian Modeling," SignalProcessing, vol. 90, pp. 2487-2497, Aug. 2010.

[35] M. A. Akhaee, M. E. Sahraeian, and F. Marvasti, "Contourlet-basedimage watermarking using optimum detector in a noisy environment.,"IEEE transactions on image processing : (a publication of the IEEESignal Processing Society), vol. 19, pp. 967-80, Apr. 2010.

[36] R. Narayanan, S. Ponnappan, and SE, "Effects of uncorrelated andcorrelated noise on image information content," REMOTE SENSING,no. C, pp. 7031-7033, 2001.

[37] A. B. Watson, G. Y. Yang, J. a. Solomon, and J. Villasenor, "Visibilityof Wavelet Quantization Noise," Image Processing, IEEE Transactionson, vol. 6, pp. 1164-75, Aug. 2002.

[38] E. Lam, "Statistical modelling of the wavelet coefficients with differentbases and decomposition levels," in Vision, Image and Signal Processing,IEE Proceedings-, vol. 151, pp. 203-206, 2004.

[39] K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "LowComplexity Image Denoising Based On Statistical Modeling of WaveletCoefficients," Signal Processing Letters, IEEE, vol. 6, no. 12, pp. 300-303, 1999.

[40] A. Petrosian and F. Meyer, Wavelets in Signal and Image Analysis: FromTheory to Practice. Computational Imaging and Vision, Springer, 2001.

[41] [Available], "`http://www.irisa.fr/vista/actions/hollywood2.`"

[42] S. Tai and S. Yang, "A fast method for image noise estimation using laplacian operator and adaptive edge detection," in Communications,Control and Signal Processing, ISCCSP, 3rd International Symposiumon, no. March, pp. 1077-1081, Mar. 2008.

[43] C. Liu, W. Freeman, R. Szeliski, and S. Kang, "Noise estimation froma single image," in Computer Vision and Pattern Recognition, IEEEComputer Society Conference on, vol. 1, pp. 901-908, 2006.

[44] D. Donoho and J. Johnstone, "Ideal Spatial Adaptation By WaveletShrinkage," Biometrika, vol. 81, no. 3, p. 425, 1994.

[45] R. K. Pearson, Exploring data in engineering, the sciences, andmedicine. Oxford ; New York: Oxford University Press, 2011.

[46] X. Gong and H. Lu, "Towards fast and robust watermarking scheme forH. 264 video," in 10th IEEE International Symposium on Multimedia,pp. 649-653, Dec. 2008.

[47] L. Zhang, Y. Zhu, and L.-M. Po, "A novel watermarking scheme with compensation in bitstream domain for H.264/AVC," in IEEEInternational Conference on Acoustics, Speech and Signal Processing,pp. 1758-1761, IEEE, 2010.

[48] X. Gong and H.-M. Lu, "Towards Fast and Robust WatermarkingScheme for H.264 Video," in Tenth IEEE International Symposium onMultimedia, pp. 649-653, IEEE, Dec. 2008.

[49] F. Hartung and B. Girod, "Watermarking of uncompressed and compressed video," Signal Processing, vol. 66, pp. 283-301, May 1998.

[50] D. Xu, R. Wang, and J. Wang, "Video watermarking based on spatiotemporal JND profile," Digital Watermarking, vol. 5450, pp. 327-341,2009.

[51] M. Bahoura and H. Ezzaidi, "Real-time implementation of discretewavelet transform on FPGA," in IEEE 10th INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING PROCEEDINGS, pp. 191-194,IEEE, Oct. 2010.

[52] R. Jiang and D. Crookes, "FPGA implementation of 3D discrete wavelettransform for real-time medical imaging," in Circuit Theory and Design,ECCTD . 18th European Conference on, pp. 519-522, 2007.

Abolfazl Diyanat received his B.S. degree in Electrical and Computer Engineering from the University of Tehran, Iran, in 2009. He was an M.S. student at Sharif University of technology, Tehran, Iran. He is now a Ph.D. candidate in Electrical and Computer Engineering at University of Tehran. His main research interests include steganalysis, multimedia, telecommunication systems, distributed optimization, and analytical modeling of wireless ad hoc networks.

Mohammad Ali Akhaee received his B.S. degree in both Electronics and Communications Engineering from Amirkabir University of Technology, and the M.S. and Ph.D. degree in communication systems from Sharif University of Technology in 2005 and 2009, respectively. He has been awarded governmental Endeavour research fellowship from Australia in 2010. He is an author/coauthor of more than 35 papers and holds an Iranian patent. He served as the technical program chair of EUSIPCO'11-13. Dr. Akhaee serves as a faculty member at the College of Eng., University of Tehran, Tehran, Iran. His research interests include multimedia security, cryptography, network security, and statistical signal processing.

Shahrokh Ghaemmaghami (M'95) received the B.S. and M.S. degrees in Electrical Engineering from Sharif University of Technology, Tehran, Iran, and the Ph.D. degree from the Queensland University of Technology, Brisbane, Australia. He is an Associate Professor of Signal Processing, a Member of the Communications and System Engineering Group of the Electrical Engineering Department, and Director of the Electronics Research Institute, Sharif University of Technology, Tehran, Iran. He has been leading many large research projects and teaching courses in signal processing, including speech, audio, image, and video processing, and also in information hiding, e.g., watermarking, steganography and steganalysis. Dr. Ghaemmaghami is a member of the Association for Computing Machinery. He has served as a member of technical committees and advisory boards of several international conferences and journals.