A Robust Visual-Audio Stream Watermarking Scheme for Video Protection

Jawhara Aljabri  
jaljabri0004@stu.kau.edu.sa

Anas Fattouh  
afattouh@kau.edu.sa

Abdullah Almalaise Alghamdi  
almalaise@kau.edu.sa

Department of Computer Science  
College of Computing and Information Technology  
King Abdul-Aziz University

Abstract—Watermarking techniques are modern and strong methods of copyright protection for authors and companies as well as for ensuring the authority and identification. They hide the company’s motto or policy of using for any digital content that exchanged via the internet. Various requirements are needed to embed the watermark in digital media like transparency, capacity, security and robustness against attacks. However, applying watermark on video, with visual and audio stream parts, is still a resource of concern for scientists and interested people to ensure the integrity of the whole video and guarantee that no manipulation was made on the video. In this research, a system is presented for video copyright protection taking into account its integrity and achieving all requirements in watermarking. The system will embed the watermark in both audio and visual streams allowing the transparency control in the watermarked video. Comparing to previous works, our experimental results showed better performance under the terms of robustness against attacks.

Index Terms—Watermarking, Discrete cosine transform, Discrete wavelet transform, Copyright protection, Robustness, JPEG compression, Law pass filter, Additive white Gaussian noise.

I. INTRODUCTION

Facing the ever-growing quantity of digital videos transmitted, shared and exchanged over the Internet, illegal copies and unreliable distribution of digital contents have become a serious, alarming problem in preserving the copyright of digital videos.

Digital watermarking is a powerful technique to solve this issue as well as the security of digital videos. Comparing to audio or image watermarking, video watermarking produces many requirements to be achieved such as transparency, which means how do we embed data such as that without perceptually degrade the underlying content. The digital watermark hide within the video data must be invisible by the human eyes, the robustness that refers to how do we mask and extract data such as that would make it survive from malicious attempts to remove it. As many attacks can affect the watermark, it must be impossible to remove the watermark by manipulating it through manual or automatic operations on the video, at the same time, a capacity which reflexes the answer to the following questions; what is the optimum amount of data that can be embedded in the host? What is the typical way to hide it and then extract this information later? In-depth, a watermarking model, should allow enough amount of copyright data to be embedded into a video. The amount of information that could we tend to hide in a watermarked video is identified as data-load. The data-load in video watermarking refers to the number of bits encoded within the video. The data-load of the embedded watermark data must be enough to create the imagined application and security that explains how do we determine that the information embedded has not been tampered or forged. The watermark must be discovered by an authorized person only. Also, the confidentiality, reliability, and integrity must be achieved during the motion [1]. Furthermore, as video consists of visual and audio streams, there are difficulties in manipulating combined visual-audio watermarking as many attacks can affect the watermarked video, especially in the audio stream [2]. This is because the human audio system (HAS) is more sensitive than human visual system (HVS).

Motivation. By combing visual and audio stream (the whole video), watermarking gives more protection in preventing unauthorized usage of copyrighted videos, ownership proof, authentication, and tampering detection. This may be utilized in video conferencing, distance learning, videophone, and many other applications. Moreover, few studies about the watermarking of the whole video. The underlying reasons behind this are summarized in the following points, i) The manipulation complexity of a video comparing to audio or image watermarking; ii) Video manipulation needs resources for high performance. This will be clearer when we want to watermark a video, even with a short period, on a computer that has excellent features, where watermarking a video with only two minutes may need several hours. Moreover, the required time to apply attacks and extracting the values of used metrics, for performance evaluation, will be very long; iii) The internet was suffering from the low speed of transferring and availability to send and exchanging the videos through it. This means that the Internet was a suitable environment for applying audio and images watermarking, which encouraged the researchers to increase the researches that related to audio and image watermarking on the account of video watermarking; v) Few researchers are proposed to improve and control the requirements of the whole video against attacks.

For integrity, a recognized motivation could be illustrated depending on the truth that says the typical multimedia is
consist of visual and audio streams. The integration between these two parts will form the typical video that we deal with in our real life. Integration could be identified as hiding the same watermark in the whole video with its two parts (visual and audio streams) to guarantee the delivering of the watermarked video to the receiver side, completely protected, so that the watermarked video will not be altered by saving the watermarked visual stream and replacing the original audio of the video by another audio. In other words, preventing the attack that depends on video theft relying on the audio stream only.

In this work, we mainly utilized wavelet and DCT transform to hide the watermark within the video where we exploited the power of those transformed to add auxiliary information (KSA logo, as a watermark) to ensure the copyright protection. We recognized our work through adding a new feature that could be used to control the transparency of the watermark over the watermarked video during running. Also, we are focused on the robustness of the embedded watermark against attacks where the total requirements of the video watermarking will be insured.

In general, the contribution of this work could be listed as follows:

- Presenting a sophisticated and robust watermarking technique for a video that could be directly applied to its two parts (visual and audio stream) to guarantee not to manipulate the video and detect any forgery during transmitting via the internet for copyright protection.
- Allowing the control of possibility of viewing or hiding the watermark under various levels of HVS (seen or non-seen levels).
- Evaluating the proposed technique by practical experiments and comparing it with the previous ones.

The remainder of the paper is organized as: Section II contains a literature review about the modern proposed works in this field. The proposed solution is described in chapter three where it illustrates, in depth, the various steps that ensure the achievement of our goals. The results of this paper are presented and discussed in section four and five respectively. Finally, section six concluded the paper and contained the future work.

II. RELATED WORK

Recently, several works are proposed to deal with video watermarking. The previously proposed works could be classified into three major categories.

The first category, dealing with a visual stream of video (i.e., manipulating only the frames of video to embed the watermark depending on frequency domain). A method for the video watermarking was introduced by Dug-Ryung Kim and Sung-Han Park [3]. They used DCT in the processing of hiding where the watermark can be presented as a sequence of zeros and ones (watermark pixels pattern) and embedded in all frames of video. Each frame is divided into blocks equal to the number of watermark bits to be embedded. The core of embedding process based on a search table of the unique relationship between the pixel patterns and their sign sequences of eight low DCT coefficients. The extraction stage depends on IDCT transform. The results show that this method is better than the previous one [3] in term of label bit error rate. Ahmed A. Baba’a Al-Deen and others proposed a scheme to enhance invisibility using blind watermarking video [4]. They applied YUV converting on the selected frames, where these frames were selected randomly depending on a particular key, DWT transform then is employed, and the embedding process is done in the luminance layer to ensure the transparency. To make the process blind, the watermarked video itself was manipulated without the need to the original video. They used a valid key to select the frames, which represent the place to hide within, and IDWT transform is applied to extract the watermark. PSNR showed enhancement in invisibility that it equal to 41.59 dB when compared with direct embedding method that has 38.48 dB.

V. K. Agrawal proposed a model in his thesis based on static 3D-DCT to hide a watermark in video depending on the motion where he chooses all kinds of frames in hiding [5]. This model gives good results for low motion activity videos otherwise there is a noticeable distortion. Similar to the previous work, V. Agrawal and S. Gupta developed the previous model so that they used dynamic 3-D DCT to get the benefit of the frequency of the video sequences that in turn gives more robustness against attacks [6].

Another method, for blind watermarking in the video, is presented by Ce Wang, Chao Zhang, and Pengwei Hao [7]. They found that the parameters in the high-frequency bandwidth of temporal wavelet transform (TWT) were orthogonally very close to the normally distributed watermark. Therefore, they do not use the low-frequency band in the DWT procedure in the embedding process. The extraction method is depending on calculating the similarity between TWT high or mid frequency frame after 3D-DWT and the watermark.

Majid Masoumi and Shervin Amiri proposed a Scheme for Copyright Protection of Video [8]. They used DWT transform in the hiding process where it was implemented on both watermark and the I-frames that represented the location of hiding information. To achieve the aim imperceptible, RGB video stream was converted into YCbCr color space, and the luminance layers were chosen to embed the watermark. After decomposing the watermarked video, IDWT transform is employed to extract the watermark. Test experiments, using various kinds of attacks showed strong resistance compared with a method that embedding in all frames as well as high transparency with satisfied values of PSNR. The same authors developed the previous study focusing on the capacity of hiding and more security [9]. They also used DWT transform in the process of hiding where the watermark was embedded in the key-frames that were included in each spot. To ensure the security, the watermark was encrypted before hiding.

Anita Jadhav and Megha Kolhekar [10] showed that most of the video watermarking schemes that used 3D-DCT work well in correlated videos but ignore the scene change in the video. Therefore, they presented a dynamic and static size 3D-DCT technique depending on scene change detection to select the
frames for embedding the watermark. Embedding steps include converting all original frames from RGB into YCbCr, choosing Y channel, as it has the least effect, dynamic and static 3D-DCT was applied on these frames, and scene change algorithm finds the frames which the embedding process was done within. 3D-IDCT transform is wanted for extraction. Results showed that there is a noticeable change in visual quality.

**The second category, dealing with the audio stream of video** (i.e., for ignoring the visual stream and picking up the audio stream and then apply the embedding process). Rade Petrovic and Dai Tracy Yang presented a technique for watermarking depending on audio stream based-compressed domain [11]. Their study addressed what modifications, in audio watermarking technology are required to suit it to a particular perceptual audio technique in the compression domain, such as AAC. Two main steps are used preprocessing and marking. In the preprocessing stage, a host signal is signaled by one or more hiders. Each hider embeds a series of identical letters. In the second step of the process, two or more separate productions of host signal are taken from the storage into a multiplexer (MUX) simultaneously when a creation of marked production is requested. This proposed technique has little resistance against attacks.

Shervin Shokri, Mahamod Ismail and Nasharuddin Zainal proposed an approach for speech watermarking using spread spectrum, depending on the reason that because of small capacity transmission limitation, the speech signal is seldom used [12]. To evaluate the system they used the quality of voice (speech) by the Mean Opinion Score (MOS) from 20 respondents. The results showed that the average MOS is 2.75 out of 5. Therefore, weakness is noticeable.

To enhance the capacity of embedding, a modified scheme using the audio of video watermarking is suggested by Maha Charfeddine, Eya Mezghani, and Chokri Ben Amar [13], as they noticed that most of the schemes do not deal with the audio stream. They used the classic method of embedding within an audio stream. Nevertheless, depending on silence deletion, they made the method robust against MPEG compression through deleting the silence before hiding and add them back to it. Therefore, the watermark is embedding in the section of audio that does not affect by compression. In their research, Rohit Tanwar and Monika Bisla described various methods employed to hide information within the audio [14]. These methods are LSB, echo hiding and spread spectrum. They explained the advantages and disadvantages of each method. For LSB method, the main advantage is the capacity for embedding data is very suitable, and disadvantage is, in addition to low robustness against attacks, it is also not impervious to manipulation. The message can be easily extracted. For echo hiding, the advantage is that HAS is not easily able to detect the presence of additional data, and disadvantage is that embedding capacity is less, and this method is less secure too. Therefore, this method is not used in recent researchers. Finally, for spread spectrum, the advantage is allowing a wide range of distortion tolerance. Resistant to compression and disadvantage is not immune to low-pass filtering and low capacity.

The third category combined visual and audio streams (i.e., dealing with both visual and audio streams in the video). Jana Dittmann and Martin Steinebach suggested a method that uses audio-visual data of video to hide the watermark [2]. Their aim was to present a solution to robust and fragile aspects to ensure both authorization and integrity based on watermark embedding with audio and visual data. After getting the audio and video streams, features are extracted where the watermark will be hidden. They used direct embedding method for video/visual stream and seed-based method for the audio stream. Then, merge stage is done. However, they do not mention anything about watermark extraction. Qiang Cheng, Thomas S. Huang, and Hao Pan [16] introduced another new method. Their method depending on generating the watermark from the features extracted from the audio stream from the same video using the mel-frequency cepstra. After generating this watermark, it is embedded into the visual stream of the same video, using DCT transform, in all frames. This ensures the integrity of the whole video. The concluded results say that the introduced method has a little loss of perceptual quality and has no resistance to video compressions attack.

**III. PROPOSED SOLUTION**

In this work, we propose a robust combined visual-audio based video watermarking method to enhance the integrity of the video for copyright protection. Combined visual-audio means that both visual and audio streams, in a video, will be used and processed to embed the watermark. Robust means that the presented method will be resistant against various geometric and no geometric attacks. In addition, security will be enhanced by applying cryptography step to the watermark before embedding in the video. Figures 1 and 2 illustrate the embedding and extracting watermark processes respectively.

![Figure 1. Embedding process.](image)

As shown in Figure 1, the steps of the embedding process could be explained as follows, i) Getting the audio and visual...
streams from the splitter; ii) For the original audio stream, we deleted the silence spots to get the original cut audio stream. DCT transform is employed to embed the watermark. After embedding, we added the deleted silence. Thus, the watermarked audio is obtained. This will make the system robust against compression attack; iii) For the original visual stream, scene change detection algorithm is decided the location of embedding. Frames of scene change are converted from RGB to YUV. Watermark will be then embedded depending on DWT transform in the Y layer to ensure the transparency. Thus, watermarked visual is achieved; iv) The merger combined the two parts to get the watermarked video; v) To ensure more security, the watermark is encoded before embedding. Triple DES is used for encryption where generating the encryption key, encryption process, and decryption process will be the responsibility of the sender side.

![Watemarked video](image)

**Figure 2. Extraction process.**

The extraction process shown in Figure 2 includes the following steps, i) Splitting the watermarked video into audio and visual streams. ii) Applying IDCT transform on watermarked audio to get the encoded watermark. iii) Decode the watermark of audio stream. iv) Scene change detection algorithm will take the frames include the encoded watermark in the visual stream. v) IDWT transform is used to extract the encoded watermark. vi) Decode the watermark of the visual stream. vii) Original audio and visual streams are merged to form the original video.

### A. Audio Stream Hiding: DCT Transform Overview

In the following, we give a mathematical description of the DCT transform and its use in hiding watermarks. We based on two-dimensional discrete cosine transform (DCT) [17] to hide our watermark. The signal \( f(a, b) \), could be formed in two-dimensional discrete cosine transform formula as follows:

\[
F(r, p) = c(r)c(p) \left[ \frac{1}{2N} \sum_{a=0}^{N-1} \sum_{b=0}^{N-1} f(a, b) \cos \left( \frac{(2a + 1)\pi r}{2N} \right) \cos \left( \frac{(2b + 1)\pi p}{2N} \right) \right]
\]

(1)

Where,

\[
c(r) = \begin{cases} 
\frac{1}{\sqrt{2}} & \text{if } r = 0 \\
1 & \text{if } r > 0 
\end{cases}
\]

Inverse function will be as follows:

\[
f(a, b) = c(r)c(p) \left[ \frac{2N}{\pi} \sum_{r=0}^{N-1} \sum_{p=0}^{N-1} F(r, p) \cos \left( \frac{(2a + 1)\pi r}{2N} \right) \cos \left( \frac{(2b + 1)\pi p}{2N} \right) \right]
\]

(2)

Where,

\[
c(p) = \begin{cases} 
\frac{1}{\sqrt{2}} & \text{if } p = 0 \\
1 & \text{if } p > 0 
\end{cases}
\]

The key idea is to apply DCT and equalization on the audio signal. Then, the audio signal will be divided into blocks, and each block will be enforced to be adaptive with the DCT transform. The coefficients of the resultant blocks will be used to hide the watermark. In-depth, choosing two coefficients of the middle frequency and one watermark signal. According to this signal, updating the values of the two coefficients. A silence deletion will be the first step as it is shown in Figure 1. IDCT transform, then, will be used in extraction phase.

### B. Visual Stream Hiding: Wavelet Transform Overview

The power of the discrete wavelet transform over other transforms is that it can provide multi-resolution related to the human vision system (HVS). In general, we have two kinds of this transform. One-dimensional where it could be represented by the following formulas [8]:

\[
W(h, j) = \frac{1}{\sqrt{M}} \sum_{a} f(a) 2^j \psi(2^j a - h)
\]

(3)

\[
\psi = \begin{cases} 
1, & 0 \leq a \leq 0.5 \\
-1, & 0.5 \leq a \leq 1 \\
0, & \text{otherwise}
\end{cases}
\]

(4)

Where, \( W \) refers to the wavelet coefficients function, \( j \), \( k \) represents the dilation and translation parameters respectively, and \( M \) is the sequence length \( f \).

From this one-dimensional, a two-dimensional could be obtained where \( \psi^H(a, b), \psi^P(a, b), \text{and } \psi^D(a, b) \) are required. Separable, “directionally sensitive” wavelets will be:

\[
\psi^H(a, b) = \psi(a) \phi(b)
\]

(5)

\[
\psi^P(a, b) = \phi(a) \psi(b)
\]

(6)

\[
\psi^D(a, b) = \psi(a) \phi(b)
\]

(7)

The expanded and translated basis functions are:

\[
\psi_{j, m, n}(a, b) = 2^j \psi \left( 2^j a - m, 2^j a - n \right)
\]

(8)

\[
\psi_{j, m, n}(a, b) = 2^j \psi \left( 2^j a - m, 2^j a - n \right)
\]

(9)

Where index \( j \) defines the directional wavelets in Eqs. (5) to (7). The discrete wavelet transform of function \( f(a, b) \) of size \( M \times N \) is given by:
Related to images, this transform emphasizes that the image preserves its main elements, and also shows that the major elements of the image have a strong ability to defense against inferences attacks. As a result, hiding a watermark in this place, we can get a better robustness against possible risks [18].

Using resultant layers of DWT we can divide a two-dimensional image into wavelet-like matrix C_{aij} and detail elements including C_{hi}, C_{vi}, C_{dj}, which are located in horizontal, vertical and diagonal routes respectively. Since we focused on the transparency control, we concern about those coefficients where our goal will be mathematically provided through the following modeling:

\[
\begin{align*}
C_a & = (1 + \alpha_1, l_{w}), l_{l_0} \\
C_n & = (1 + \alpha_2, h_{w}), h_{l_0} \\
C_r & = (1 + \alpha_3, h_{w}), h_{l_0} \\
C_d & = (1 + \alpha_4, h_{w}), h_{l_0}
\end{align*}
\]

(12)

\[i = \{H, D, V\}\]

D. Possible Attacks and Their Effects on Watermarked Video

Since we need our proposed approach to be robust against attacks, we should address the possible attacks that could be occurred by an attacker. In addition, we should define the measures that can be affected by those attacks.

In general, many intentional or unintentional attacks could be applied to the watermarked video [25]. However, all those attacks target the robustness requirement. In our work, we selected the most harmful attacks to be applied in our experiments. A description about those attacks could be summarized as follows:

**Compression attack.** This attack is considered as interference attack that contributes to increasing the noise that originally added to the video due to embedding the watermark. This attack directly impairs the similarity between the original frame and the watermarked one. As a result, it affects both the correlation to be decreased and the bit error rate to be increased.

**Low pass filter.** The victims of this attack are both the video flicker, which related to the quality of the video during the show, and the luminance of the frames [26]. This means that this attack has a bad impact on the bit error rate, correlation, and MSSIM metrics.

**Gaussian noise.** Since the embedded watermark itself is considered as a noise, this attack aims to increase the distortion caused by the watermark. In other words, increasing the noise to blur the watermarked video. Again, this attack will lead to poor values related to both bit error rate and correlation.

IV. RESULTS

In this section, experimental results will be provided to show the applicability of the proposed solution.

From one hand, because we need to split the videos into visual and audio streams, the following website is used [http://online-audio-converter.com/](http://online-audio-converter.com/). From another hand, since we deal with scene change as a place to hide the watermark in the visual stream, and this scene change differed from one video to another, we applied our proposed watermarking approach on various different color video sequences including cartoon video, animal video, elephant video, and rhinos video. These used videos are 30 fps and in a durations of them are 57, 126, 215, and 7 seconds respectively. Figures 4, 5 show the used four videos and the two binary images (butterfly logo and King Abdulaziz University logo) with the same size of the video frame as a watermark.

**For the visual stream,** we selected rhinos video to define the best value for alpha that is used to control the transparency to be \(\alpha = 0.001\) depending on the values of the following metrics: MSE, MSSIM [20], and PSNR [21] as shown in Table 1.

![Figure4. Frame samples of the used videos.](image-url)
Figure 5. The used watermarks.

Table 1. Values of the used metrics as a function to Alpha.

<table>
<thead>
<tr>
<th>Alpha</th>
<th>MSE</th>
<th>PSNR</th>
<th>MSSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>99.00</td>
<td>65.23</td>
<td>0.934</td>
</tr>
<tr>
<td>0.002</td>
<td>123.00</td>
<td>63.01</td>
<td>0.922</td>
</tr>
<tr>
<td>0.003</td>
<td>151.00</td>
<td>60.70</td>
<td>0.911</td>
</tr>
<tr>
<td>0.004</td>
<td>197.00</td>
<td>58.70</td>
<td>0.898</td>
</tr>
<tr>
<td>0.005</td>
<td>247.00</td>
<td>56.92</td>
<td>0.873</td>
</tr>
<tr>
<td>0.006</td>
<td>298.00</td>
<td>54.92</td>
<td>0.859</td>
</tr>
</tbody>
</table>

On one hand, since alpha is used to control the transparency of the embedded watermark, the distortion of the watermarked video will be increased due to increasing the noise. On another hand, high values of alpha lead to more visible watermark that in turn leads to more noise comparing to the original signal. In depth, as a more noticeable watermark in the watermarked video, as high ability of the attacker to target the watermark. Therefore, a low alpha value is desirable to make the watermark hard to be detected by the attacker.

Many attacks are taken into account to prove the resistance ability against it. Because the most attacks that can affect the watermarked video are Compression, Low Pass Filter, and Gaussian Noise, they are applied. The next curves represent the results using bit error rate and correlation metrics to evaluate the robustness of the watermarked videos.

Figure 6. BER and Correlation values under JPEG compression attack under various compression levels.

Figure 7. BER and Correlation values under LPF attack under various filter sizes.

From the previous curves, it was noticed that we have good results related to BER and correlation since the hidden watermark is considered to have the highest transparency according to the selected alpha value. The reason behind this could be justified as the embedded watermark will be strongly affected by the applied attacks if its transparency has poor value. To provide a satisfied explanation, we asked the same attacks on the same used videos under $\alpha = 0.9$. The following tables summarize these results.

Table 2. Correlation and BER values under $\alpha = 0.001$.

<table>
<thead>
<tr>
<th>Video name</th>
<th>Attack type</th>
<th>(quality factor=70)</th>
<th>Size [3 3] LPF</th>
<th>variance=0.01 AWGN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cartoon</td>
<td>0.15 0.45</td>
<td>0.15 0.43</td>
<td>0.17 0.42</td>
<td></td>
</tr>
<tr>
<td>Animal</td>
<td>0.14 0.50</td>
<td>0.17 0.41</td>
<td>0.17 0.42</td>
<td></td>
</tr>
<tr>
<td>Elephant</td>
<td>0.15 0.5</td>
<td>0.16 0.42</td>
<td>0.17 0.43</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Correlation and BER values under $\alpha = 0.9$.

<table>
<thead>
<tr>
<th>Video name</th>
<th>Attack type</th>
<th>(quality factor=70)</th>
<th>Size [3 3] LPF</th>
<th>variance=0.01 AWGN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cartoon</td>
<td>0.08 0.73</td>
<td>0.098 0.7</td>
<td>0.1475 0.52</td>
<td></td>
</tr>
<tr>
<td>Animal</td>
<td>0.08 0.73</td>
<td>0.098 0.7</td>
<td>0.1462 0.517</td>
<td></td>
</tr>
<tr>
<td>Elephant</td>
<td>0.08 0.73</td>
<td>0.098 0.7</td>
<td>0.147 0.518</td>
<td></td>
</tr>
</tbody>
</table>

Figure 9 shows the extracted watermark after applying the attacks under $\alpha = 0.001$, and various factors related to each attack.
Comparing with previous works [8, 22, 23], we can that our method gives better numerical results taking in consideration the used attack level (i.e., different JPEG quality factor, AWGN variance, and type of used filter).

For JPEG quality factor, here it is considered the widest used attack for evaluation. Our results show that correlation value (for the three videos, under factor = 70) is close to one (0.934) and achieved better hit obtained on their side (0.8938). This means that our proposed method did not affect by JPEG attack and more robust than the previous work.

For AWGN variance, they obtained 0.6048 correlation value under variance = 0.001, where we achieved 0.6352 correlation value under variance = 0.001.

For low pass filter, the result obtained using their method gave 0.9112 correlation value using filter size = [3, 3] where our proposed method gave 0.934 correlation value using filter size = [3, 3]. The following table summarizes the comparison.

Table 4. Comparison with previous works.

<table>
<thead>
<tr>
<th>Attack type</th>
<th>Used level</th>
<th>Correlation value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Our Method</td>
<td>Method [8]</td>
</tr>
<tr>
<td>JPEG Factor =70</td>
<td>0.934</td>
<td>0.8938</td>
</tr>
<tr>
<td>LPF Size =3[3]</td>
<td>0.934</td>
<td>0.9112</td>
</tr>
<tr>
<td>AWGN Variance =0.001</td>
<td>0.6352</td>
<td>0.6048</td>
</tr>
<tr>
<td></td>
<td>Method [22]</td>
<td>Method [23]</td>
</tr>
<tr>
<td></td>
<td>0.7606</td>
<td>0.7503</td>
</tr>
<tr>
<td></td>
<td>0.9030</td>
<td>0.5512</td>
</tr>
<tr>
<td></td>
<td>0.5893</td>
<td>0.6029</td>
</tr>
</tbody>
</table>

For the audio stream, after applying our proposed method on the audios of the corresponding videos, we obtained the results shown in Table 5 below.

Table 5. Audio watermarking results.

<table>
<thead>
<tr>
<th>Video name</th>
<th>MSE value</th>
<th>PSNR value</th>
<th>Extracted watermark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carton</td>
<td>0.00021971</td>
<td>66.305</td>
<td></td>
</tr>
<tr>
<td>Animal</td>
<td>0.00015173</td>
<td>76.3785</td>
<td></td>
</tr>
<tr>
<td>Elephant</td>
<td>0</td>
<td>Infinity</td>
<td></td>
</tr>
</tbody>
</table>

The waveforms reflect similarity between the original and watermarked audios are shown in Figure 10.

We applied signal-processing operation (loosely compression) on watermarked audio digital signals. Although this action has no effect on the watermarked quality of the host signal, they deviate the watermark that was hidden within the signal. To evaluate the robustness of the proposed algorithm, we implemented a set of levels of lossy compression as it is shown in Table 6 below.

Table 6. Audio watermarking compression attack.

<table>
<thead>
<tr>
<th>Video name</th>
<th>Compression level</th>
<th>MSE value</th>
<th>PSNR value</th>
<th>Extracted watermark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carton</td>
<td>4</td>
<td>29.2x10^3</td>
<td>66.305</td>
<td></td>
</tr>
<tr>
<td>Animal</td>
<td>8</td>
<td>15.1x10^3</td>
<td>76.378</td>
<td></td>
</tr>
<tr>
<td>Elephant</td>
<td>16</td>
<td>0</td>
<td>Infinity</td>
<td></td>
</tr>
</tbody>
</table>

Comparing to previous works [24] we inferred that our proposed approach performs better under values of MSE and PSNR. This is because although [24] used several kinds of attacks, but we applied the most strong one where our obtained values are higher.

V. CONCLUSION

In this work, we presented a model for video copyright protection that was taking into consideration the two part of the video; visual and audio stream where the whole requirements; transparency, robustness, capacity, and security were achieved. In addition, we proposed watermark embedding and extraction for both visual and audio stream, where wavelet transform was used to hide the watermark in the scene change while DCT transform was used to cover up the watermark in the audio stream.

In our experiments, it was observed that the extracted watermark, in visual stream, was clearer than that extracted from the audio stream. Under various attacks (JPEG compression, Law pass filter, and AWGN Gaussian), and using several metrics (BER and Correlation), our model showed better performance and quality compared to previous works.

REFERENCES

[2] Jana Dittmann, Martin Steinebach, "JOIN HAND-REMOTE DATA HANDLING"., German National Research Center for Information Technology (FZI) – Research Institute for Interactive Media, Darmstadt, Germany, 0-7803-0252/01 $10.00 02001 IEEE.