Multimedia Tools and Applications (2019) 78:16159–16175
https://doi.org/10.1007/s11042-018-6892-6

A fast video watermarking algorithm using dual tree complex wavelet transform

Reza Esfahani[1] · Mohammad Ali Akhaee[2] · Zynolabedin Norouzi[1]

Received: 14 April 2018 / Revised: 24 October 2018 / Accepted: 13 November 2018 /
Published online: 14 December 2018

Abstract

Illegal camcording as the most common source of video piracy, is the main concern of the film production companies. With the increased speed and convenience of access to the Internet and video sharing services, the distribution of pirated video copies is more than easy. Digital video watermarking is one of the possible ways to prevent the illegal recording and distribution of videos. In this paper, a blind digital video watermarking method based on Dual-Tree Complex Wavelet Transform (DTCWT) is proposed. Here, the generated watermark is embedded in the low frequency coefficients of the chrominance channel of the video frames. The watermark is generated in a way that the detection system does not need to know the size of the original video. To reduce the computational complexity of the proposed method, the embedding process is equivalently done in the spatial domain. The proposed method is compared to the state-of-art methods against different attacks. Experimental results show that the proposed method is more robust against various attacks while maintaining the perceptual quality of the original video significantly better than the compared methods. Also the proposed method is evaluated in terms of computational complexity which reveals that the proposed method performs faster than the compared methods.

Keywords Video watermarking · Video piracy · Dual-tree complex wavelet transform · Robustness against comcording

✉ Reza Esfahani
resfahani@ihu.ac.ir

Mohammad Ali Akhaee
akhaee@ut.ac.ir

Zynolabedin Norouzi
zy_norozi@ihu.ac.ir

[1] Information Technology and Communication Faculty, Imam Hossein University, Tehran, Iran

[2] Schools of Electrical and Computer Engineering, Faculty of Engineering, University of Tehran, Tehran, Iran

1 Introduction

Nowadays, online video piracy is one of the major concerns and challenges of the film production companies. Illegal camcording at the cinema theaters is the most common source of the video piracy, in which video captured from the screen by a camcorder is widely distributed through Internet without observing the copyright laws. Today, watching the pirate videos online is much more facilitated, taking into account the ease of access to the Internet and the number of streaming servers. Although strict laws have been passed in various countries to stop the online piracy, they have been ineffective in practice. For example, the Dark Knight movie was illegally downloaded 7-million times within only six months of its original distribution at 2008; despite the fact that the Warner Brothers Company had prepared itself to combat this challenge from several months ago [30]. One of the most important techniques capable of tackling such disputes is the video watermarking.

In video watermarking, the copyright information is embedded into the video signal. A watermarking system is normally composed of two steps: embedding and detection. At the embedding stage, a watermark is generated and embedded into the original signal to produce the watermarked signal. In the detection phase, the input signal is examined to detect whether it contains the watermark or not [21]. In the most cases that are known as the blind watermarking, the original signal is not available at the detection phase and detection must be carried out without the reference signal. One of the main issues of the blind watermarking is the synchronization; which can be defined as the process of extracting the relation between the spatial and temporal coordinates; or in other words, locating the watermark. Therefore, any problem in the synchronization will trouble the watermark detection [21].

Since the watermark is actually an extra information added to the host signal, it can degrade the visual quality of the video. Therefore, maintaining the video quality at an acceptable level after being watermarked – known as the imperceptibility – is a major characteristic of a watermarking system. A watermarked video may suffer from various kinds of distortion. For instance, consider the video camcording at the cinema theaters. In case that this video is captured from a large screen via a high quality video camera capable of HD-quality recording, the video resolution might be preserved. However, if this video is supposed to be played on the portable devices such as mobile phones, its format or resolution might be changed. On the other hand, due to the improper camera position, distortions like upscaling, rotation and cropping might happen which results in a trouble for the synchronization. Moreover, the video might lose some information due to transmission on the Internet. Frame rate conversion is some other attack often applied to break the synchronization. Altogether, watermark must be detectable whether it has suffered from maliciously or unintentionally applied distortion. Therefore, robustness is another major characteristic of a watermarking system [21].

Within recent years, various blind watermarking algorithms are proposed to withstand different types of attacks. Using Harris detection in [12], watermark is embedded into the discrete cosine transform (DCT) domain of Y component of the cover video. Similar techniques are presented in [32] and [19] where the watermark is embedded into low frequency DCT coefficients; due to the effect of attacks such as downscaling on the high frequency coefficients. Downscaling a frame in the spatial domain is almost equivalent to removing the high frequency bands in the DCT domain. However, the human visual system (HVS) is more sensitive to the changes in the low frequency coefficients. Since the coefficients around DC normally are of high magnitudes, subtle modification of such coefficients results in the perceivable degradation in the video quality. Therefore, the watermark power cannot be increased in such methods due to the imperceptibility considerations. Although the

methods proposed in [12] and [32] are robust against the scaling and rotation, they face limitations in case of cropping attack. The method proposed in [19] is based on quantization index modulation (QIM) embedding and is robust against the downscaling, frame rate conversion and format conversion. Meanwhile, it is incapable to withstand geometric attacks such as rotation and cropping.

Rasti et al. [28] proposed to embed the watermark into low entropy parts of all three RGB color channels of each frame by combining QR decomposition, SVD, Chirp Z-transform and DWT. This method shows good performance against different attacks, however it uses a non-blind detection algorithm. In order to provide more security and to mislead the intruder, a frame selection algorithm is proposed in [5]. In this method a few frames are selected for embedding based on mathematical relationship between number of frames and capacity. However, this may make the system vulnerable to temporal attacks which is not considered in the proposed algorithm. In [29] a bi-orthogonal wavelet transform (BWT) based video watermarking method is proposed. Also, an improved artificial bee colony algorithm is employed to generate random frame for the embedding process. Experiments show a good performance against different attacks while the combined attacks are not considered at all.

It should be mentioned that DWT and DCT transforms have a variaty of usage in image processing area [20]. A combined DWT-DCT transform is utilized in [13]. The Arnold transform is also employed to enhance the security and robustness. To improve the robustness of the watermarking algorithm, key-frames could be selected as carries in the frame sequence [10]. They proposed a method based on boundary luminosity analysis to extract the key-frames. Then the watermark is embedded in the low frequency DCT coefficients of the key-frames. In contrast, the proposed method in [1], utilizes the P-frames of video for embedding purposes. The motion information is analyzed to find appropriate selections and then, nonzero quantized residuals are utilized for watermark embedding. One of the popular solutions for controlling the watermark strength and generate robust and imperceptible watermarked carriers is to consider Just Noticeable Distortion (JND). A saliency-modulated JND profile is proposed in [6] to improve video watermarking scheme. Their method tries to achieve the most robust possible scheme with an imperceptible watermark.

In [11] and [7] watermarking algorithms are proposed based on discrete wavelet transform (DWT). The watermark is embedded into the DWT coefficients of a sub-image in [11]. To detect the watermark, the sub-image is extracted from the entire image at first, and then the watermark detection algorithm is applied to it. This method is robust against cropping, but faces some problems in case of the rotation. The method in [7] also suffers from lack of robustness against geometric attacks. These weaknesses arise from two issues in the DWT domain. The first problem is the lack of shift invariance, that is, a subtle shift of the input signal causes significant alterations in the coefficients' distribution at various scales. The second problem is the poor directional sensitivity. Dual-tree complex wavelet transform (DTCWT) with intrinsic features like approximate shift invariance, good directional sensitivity and perfect reconstruction and efficient computation has been introduced to address these issues [14–16]. The magnitude of the low frequency coefficients remaining almost the same after rotation and scaling results in the robustness of watermarking techniques based on DTCWT against geometric attacks [9].

Recently, many watermarking algorithms have been proposed based on DTCWT [2–4, 18, 22, 23, 26, 27, 31]. In [9], all video frames are transformed at first, using a 4-level DTCWT. The watermark coefficients are derived by applying a 1-level DTCWT to the watermark and are added to the third and fourth level coefficients of frame with the proper weights. This method is robust against the upscaling, cropping, rotation and lossy com-

pression, but its performance is poor against downscaling in resolution. Moreover, in this method, the luminance channel (Y) of frames is applied for watermark embedding. Since the HVS is more sensitive to luminance channel than the chrominance one [17, 25], the imperceptibility is reduced by increasing the watermark power. A DTCWT and Singular Value Decomposition (SVD) based hybrid watermarking system is proposed in [4] that improves the imperceptibility through applying the U channel of frames. However, the performance of this method is not evaluated against frame rate conversion. SVD is also combined with secrete sharing to obtain robust video watermarking scheme [8]. Another DTCWT based algorithm is proposed in [3] where the watermark is embedded into the all six sub-bands of the third level of a 3-level DTCWT transform. This method embeds the watermark into the U channel too, and is robust against the frame rate conversion attack, in addition to the geometric attacks but the computational complexity of the proposed method is hight and it has poor performance against large combined geometric distortions.

In this paper, a video watermarking technique based on DTCWT is presented. For better detection, the watermark generation method somehow differs from common approaches; in the sense that every watermark row is generated with its own specific key which is different from those of others. The U channel is applied for the watermark embedding to increase the imperceptibility. Moreover, to increase the robustness against attacks, high frequencies of the third level of a 3-level DTCWT are used. Only the coefficients at the first and sixth subbands are applied to improve the imperceptibility. Moreover, to significantly decrease the embedding time, watermark embedding is not performed directly at the transform domain, but its equivalent mask is executed in the spatial domain. At the detection phase, the watermark is extracted from the received video at first. Then two-dimensional (2D) normalized cross correlation (NCC) between the original watermark (generated with the help of a given key) and the extracted watermark is done. By comparing the result to a preset threshold, the presence or absence of the watermark is detected. Many scenarios are considered to evaluate the performance of the proposed method and the comparison is performed with state of the art methods.

The rest of the paper is organized as follows. In Section 2 the details of different parts of the proposed method is described in details. Section 3 includes the evaluation of the method against different video attacks. Finally Section 4 concludes the paper.

2 Watermark algorithm

2.1 Watermark generation

Watermark is assumed to be a random 2D matrix of elements including -1 and +1. A pseudo-random number generator with the below key is applied to generate the watermark:

$$k = k_c + k_v \tag{1}$$

The overall key consists of two parts: k_c which is the fixed key and the variable key k_v which changes every v frames. k_v prevents a certain watermark from being repeatedly embedded into many frames. Since a large value of v means embedding a certain watermark into many consecutive frames and consequently the watermark being vulnerable against the estimation attacks, its value should not be too large. On the other hand, a small value for v helps the attacker to remove the watermark from the video by means of temporal frame averaging techniques [21]. Therefore, v must be properly set to avoid both the estimation and temporal frame averaging attacks.

At the detection phase, watermark will not be recoverable without knowing the size of the originally generated watermark at the embedding stage. We further discuss this issue through an example. Figure 1a shows the 2D NCC between two watermarks of sizes 60×120 and 55×115. Assume that the first watermark is embedded into the cover video, and the second one is produced using the key and the size of the input video. As can be observed, there exists no correlation between these two watermarks and the detector will be misled. Therefore, the detector fails without knowing the size of the original video since it must make watermark from the size of the manipulated video which is not correct. Following, we suggest a solution to this problem.

To solve this problem, the k_c key must change for every row (or column) while generating the watermark. With this approach, generating watermarks of different sizes does not trouble its synchronization, and always a watermark of smaller size is a sub-matrix of the larger one. Figure 1b shows 2D NCC between two watermarks of sizes 60×120 and 55×115 generated by the described approach. As can be observed, the detector simply detects the correlation between two watermarks; hence, the size of the original video is not required at the receiver side and the key for watermark generation suffices for the detection.

Due to the redundancy in the DTCWT, some information of the pseudo-random watermark lies in the null space and would be lost during inverse transform. To avoid this, the watermark is added in the transformed domain. The last issue is the watermark size. Since the watermark is a 2D array, its size is defined having the number of its rows and columns known. The horizontal and vertical dimensions of the watermark are chosen as one eighth of those of the U frame. The watermark embedding procedure is described with details at the following section.

2.2 Watermark embedding

Two important issues must be considered to design a proper watermark embedding algorithm. The embedded watermark must be robust against various attacks, and it should not cause such severe degradation in the video quality. The low frequency coefficients of a video frame are robust against the compression and geometric attacks, though the HVS is more sensitive to their changes [9]. Therefore, deciding where to add the watermark is a trade-off between the watermark imperceptibility and its robustness against attacks. In the proposed method, low frequency coefficients are applied to increase the watermark robustness, while

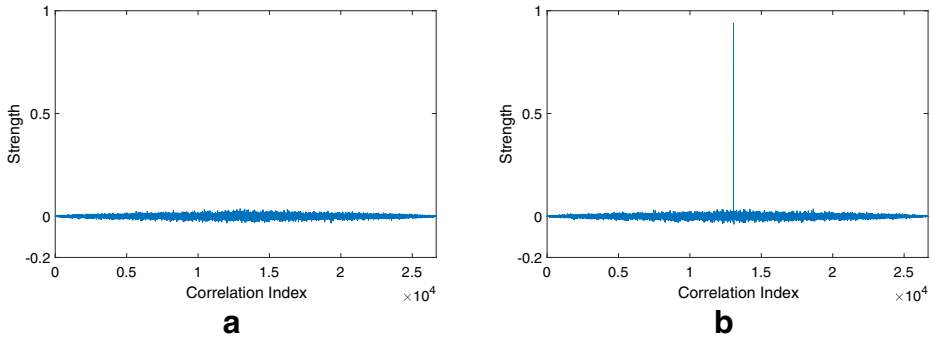

Fig. 1 The effect of the size of the generated watermark on the correlation strength. **a** The traditional way of watermark generation **b** The proposed method for watermark generation. The horizontal and vertical axes show the correlation index and 2D normalized correlation, respectively

the U frame of the YUV image presentation is chosen to improve the imperceptibility. The YUV model of a video frame is derived from its RGB one as below:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.2989 & 0.5866 & 0.1145 \\ -0.1688 & -0.3312 & 0.5 \\ 0.5 & -0.4184 & -0.0816 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2}$$

To increase the robustness of our method, the third level coefficients of a 3-level DTCWT (low frequency coefficients) are applied for the watermark embedding. Different regions of a frame have different frequency characteristics. Therefore, to improve the imperceptibility, a conceptual mask should be designed to adjust the watermark power at different regions. This mask must be designed such that it increases the watermark power at the regions with higher frequency components. Since this mask must be perfectly recoverable at the detection phase, it is extracted from the first level coefficients of the Y frame. The Y channel and first level coefficients are selected due to their negligible alteration at the detection stage [9], and the imperceptibility improvement, respectively. Before further discussion on this selection, we explain how the mask is made.

As noted before, every DTCWT level contains six high frequency sub-bands. Only the first and sixth sub-bands are applied in the proposed method for the sake of embedding. Two sub-bands are selected instead of all of them, to improve the watermark imperceptibility. The first and sixth level sub-band coefficients of the Y frame are used for the mask generation, and they are combined as below to produce the watermark:

$$\hat{M}_y = \left\lceil \frac{|Y^H_{1,1}| + |Y^H_{1,6}|}{2} \right\rceil \tag{3}$$

where $\hat{M}_y$ and $|Y^H_{1,i}|$ are the mask (dependent to the Y frame) and i'th sub-band of the first level of the transformed Y frame, respectively. To prevent the watermark power from being excessively increased, a constraint is considered for the generated mask as below:

$$M_y = \begin{cases} \hat{M}_y & \hat{M}_y \leq \beta \\ \beta & \hat{M}_y > \beta \end{cases} \tag{4}$$

In this way, the overgrow of the watermark power in some regions that causes the imperceptibility problems is prevented. The input video is assumed to be in YUV 4:2:0 format. In this format, the U frame horizontal and vertical dimensions are one half of those Y frame; hence, the generated mask is of the same size as the U frame. Now we discuss on the selection of the first level coefficients for the mask generation. As stated, for the sake of simplicity, we apply construct the mask from lower level coefficients. Thus, it is needed to upsample the mask to be the same size as U frame. However, with such expanding, the effect of every mask element is leaked to the neighboring elements. For example, the mask normally has higher values at the edges, resulting in the watermark power increases at those regions. Therefore, increasing the size of the mask causes its value that is proportional to the watermark power increases around edges, which yields leaving an undesirable artifact in those regions. This issue will be illustrated through an example in Section 4.

When the mask is generated, the watermark must be added to the first and sixth sub-bands of the third level of the transformed U frame. In the proposed method this is equivalently performed in the spatial domain rather than directly in the transform domain, in order to speed up the watermarking process. More precisely, a 3-level DTCWT is required to insert

the watermark in the third level sub-bands of the U frame, from which we try to avoid since it needs more computation.

To prepare the watermark, a one-level transform is applied to the watermark W at first. The watermark W' is constructed such that the first and sixth sub-bands of its third level transform are equal to those of W; while the other sub-bands of its all levels are set to zero. For this sake, a 3-level DTCWT is first applied to a zero-valued 2D array of the same size as U frame. Then the first and sixth sub-bands of its third level are set to those of W, and the inverse transform is applied. Finally, the watermark is inserted into the U frame in the spatial domain based on the below formula:

$$\hat{U} = U + \alpha M.W' \tag{5}$$

where $M.W'$ represents the element by element matrix multiplication. α determines the watermark power. More precisely, α and β controls the trade-off between the robustness and imperceptibility of the watermark. As the consequence, their proper adjustment is critical for the system efficiency. The block diagram of the embedding method is shown in Fig. 2.

2.2.1 Parameter adjustment

Imperceptibility is one of the most important factors in a watermarking system design. An embedded watermark perceivable by the HVS is not suitable for the practical applications. The peak signal to noise ratio (PSNR) is a common metric for the visual evaluation of the videos. However, most of the time, it does not properly reflect the perceptual quality of the videos; due to the non-linear behavior of HVS. Therefore, universal standards such as JPEG and MPEG rely on the subjective tests to better evaluate the different algorithms.

Considering the above discussion, we have also adopted the subjective tests to adjust α and β parameters in our watermarking system. To this end, an experiment is conducted as described in the following. In each test, the watermarked and original videos of one-minute duration are simultaneously played for the subject at the rate of 30 frames per second (fps). The starting point of the video is chosen randomly. Then the subject votes for the perceived quality of the watched videos through a number ranging from one to five; where one and

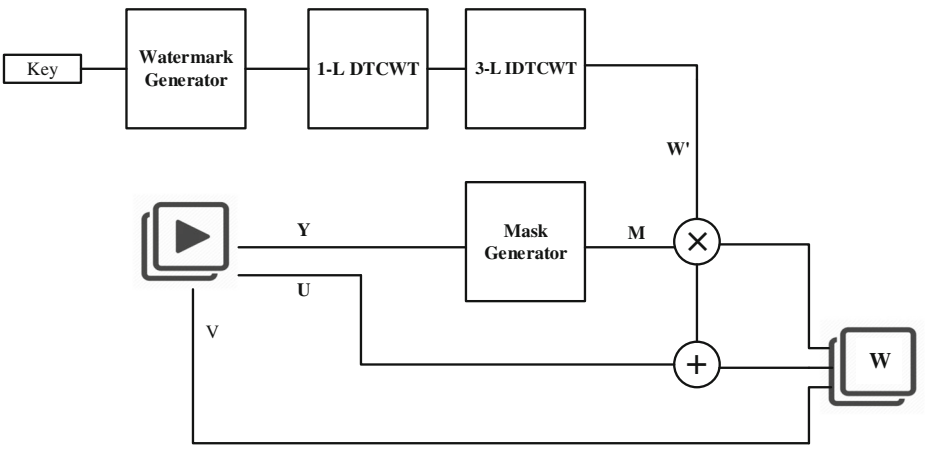

Fig. 2 The block diagram of the proposed watermark embedding algorithm

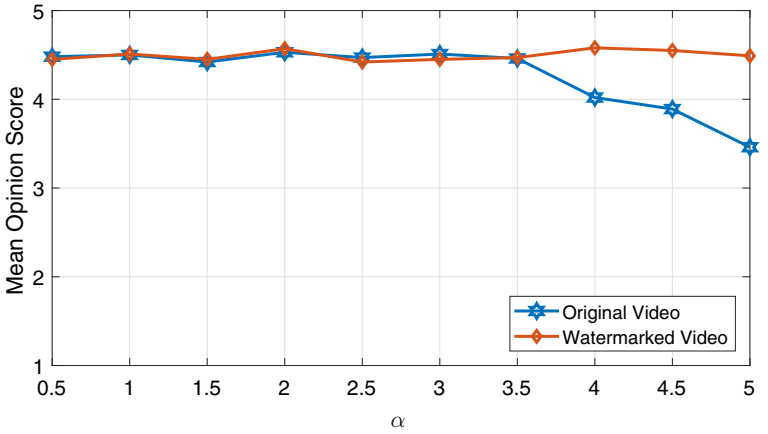

Fig. 3 Perceptual quality of the watermarked videos in terms of mean opinion score for different watermark strengths (α)

five represent the worst and best qualities, respectively. This experiment is repeated for 10 different videos.

Finding the best values for α and β requires extensive experiments. Considering the high correlation between these two parameters, β is fixed at 20 to avoid the large number of experiments. Therefore, only α is derived subjectively. For this sake, each video is watermarked by all α values ranging from 0.5 to 5 at step 0.5; thus every subject must make 100 different comparisons (10 videos each one watermarked at 10 different α values). In each experiment, the original video and the α value of the watermarked video are chosen randomly to fairly perform all 100 experiments. The mean opinion score (MOS) of this experiment is shown in Fig. 3.

As expected, subjects have been unable to distinguish between the original and watermarked videos for low α value; while they have correctly detected the quality loss of the watermarked video in some cases for α values exceeding 3.5. This is due to the fact that in the videos with more low frequency components, increasing α decreases the watermark imperceptibility rapidly. Therefore, we set $\alpha = 3.5$ and $\beta = 20$ hereafter.

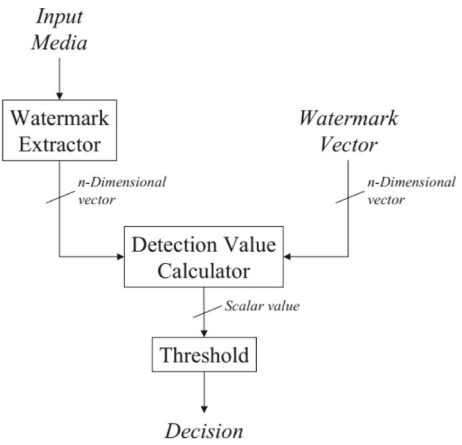

Fig. 4 The block diagram for the watermark detection system. The figure adapted from [24]

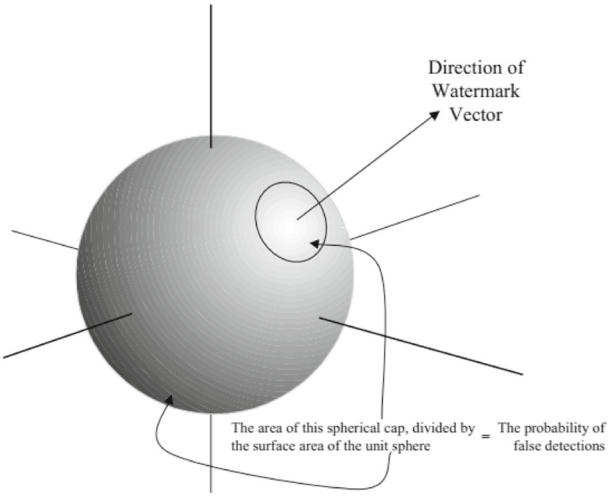

Fig. 5 Three dimensional detection region. The figure adapted from [24]

2.3 Watermark detection

The inverse of the watermark embedding process must be carried out for the watermark detection. As the first step, the desired watermark should be calculated. Then the watermark detection is performed thanks to the correlation between the extracted and desired watermarks. To this end, the mask $M_{y'}$ is extracted from the Y channel of the received video (denoted by Y') in the same manner as that explained in the watermark embedding process. Then the received U frame (denoted by U') is element-by-element divided by the mask $M_{y'}$, yielding U''. This reduces the effect of U frame elements in the detection process. Then the first and sixth sub-bands of the third level of a 3-level DTCWT applied on U'' are extracted. They are denoted by W_1' and W_6', respectively.

To detect the embedded watermark, W_1' and W_6' must be compared to the expected watermark with respect to the key K. For this sake, the watermark W is constructed with respect to the size of U', based on the previous section. Then the first and sixth sub-bands of W are extracted using a one-level DTCWT, and are called W_1'' and W_6'', respectively. At the end, it only suffices to calculate the correlation between W_1', W_6', W_1'', and W_6''.

Now the detection steps are described mathematically. After extracting the received Y and U frames (Y' and U') and generating the mask $M_{y'}$, the matrix U'' is constructed as below:

$$U'' = U'./M_y \tag{6}$$

where ./ represents the element-wise division. Then the correlation between the desired and extracted watermarks is evaluated based on 2D NCC criteria as below:

$$NC_f = \frac{1}{2} \times \left(W_1' * W_1'' + W_6' * W_6''\right) \tag{7}$$

Table 1 The threshold of the watermark detection system for each video size

Video Size	1920 × 1080	1280 × 720	352 × 288
Threshold	20847.51	7718.93	62.97

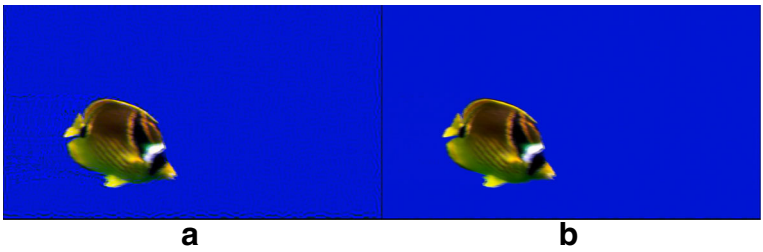

Fig. 6 The effect of the watermark strength on perceptual quality of video. **a** the method of [3] with recommended watermark strength **b** the proposed method with $\alpha = 3.5$

where * stands for 2D NCC calculated as below:

$$W * V = \frac{||W.V||}{||w|| \times ||v||} \tag{8}$$

f in NC_f represents that the correlation calculation is performed for each frame. To reduce the effect of the uncorrelated values on the non-watermarked frames, NC_f is averaged over t consecutive frames, resulting in decreasing the false positive rate. Therefore, the final correlation value will be:

$$NC = \frac{1}{t} \sum_{f=1}^{t} NC_f \tag{9}$$

Then NC must be compared to a certain threshold. The determining element of the threshold is set such that the probability of false detection (P_{fd}) lies below 10^{-6}. The detection threshold is calculated utilizing the model described in [24]. The watermark detection system is illustrated in Fig. 4. The detection value calculator is 2D NCC in our problem. NCC can also be described as below:

$$w * v = \frac{||W.V||}{||W|| \times ||V||} = \cos(\gamma) \tag{10}$$

where γ is the angle between w and v. More precisely, since w and v are two-dimensional, γ is the angle between the normal vectors of w and v planes. The threshold value can be determined with respect to the γ angle. All the possible watermarks lie on a hyper-circle. The ratio of the area of the hyper-circle located inside the detection region to the entire hyper-circle area determines the false alarm, as shown in Fig. 5. This ratio is calculated as below [24]:

$$p_{fd} = \frac{I_{n-2}(T)}{2 \times I_{n-2}(\pi/2)} \tag{11}$$

Table 2 The robustness of three methods against compression attack

Video Size	[3]	[13]	proposed
352x288	0.5418 ± 0.1601	0.5399±0.2063	0.5412 ± 0.0938
1280x720	0.4731 ± 0.1689	0.4880±0.2384	0.4913 ± 0.0856
1920x1080	0.4831 ± 0.1651	0.5109±0.2197	0.5544 ± 0.0832
total	0.4993 ± 0.1670	0.5129±0.2135	0.5290 ± 0.0928

The average and standard deviation of the normalized correlation are reported for all methods for different video qualities

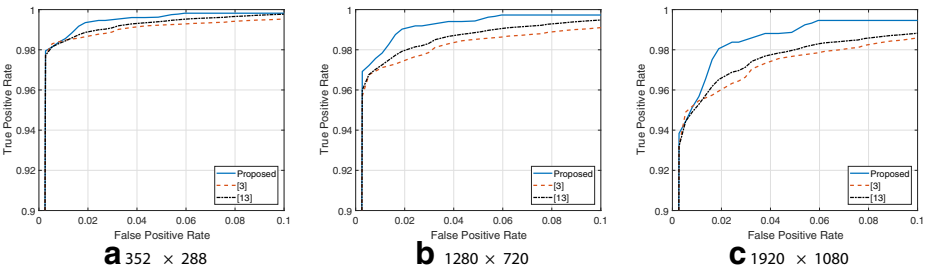

Fig. 7 ROC curve for compression attack

where n is the number of watermark elements and

$$I_d(\theta) = \int_0^{\theta} \sin^d(\theta) d\theta \tag{12}$$

In (11), T and d are the threshold value and the watermark dimensionality, respectively. Using this formula, the threshold value is determined numerically to achieve $P_fd = 10^{-6}$. The threshold values for three different video sizes are given in Table 1.

3 Experimental results

In order to evaluate the proposed method, 20 standard videos of resolution 288 × 352 including *akiyo*, *bowing*, *bridge_close*, *bridge_far*, *bus*, *coastguard*, *crew*, *deadline*, *foreman*, *highway*, *husky*, *mad*900, *silent*, *students*, *suzie*, *waterfall*, *city*, *container*, *flower*, and *football*(*b*), 10 HD-quality videos of resolution 1080 × 1920 including *blue_sky*, *crowd_run*, *dinner*, *ducks_take_off*, *factory*, *into_tree*, *old_town_cross*, *park_joy*, *pedestrian_area*, *riverbed* and lastly 8 videos of resolution 720 × 1280 including *vidyo*4, *vidyo*3, *vidyo*1, *Stockholm*, *sintel_trailer*, *shields*, *parkrun* and *mobcal* are applied, all of them containing 300 frames in 4:2:0 format recorded at the rate of 30 fps. The watermarking key changes once every 20 frames, that is v=20 in (1). The average correlation of all frames is applied at the receiver for the sake of detection, i. e. t=300. The performance of the proposed method is compared to [3] and [13]. Considering the different effects of the watermark power in each method, the watermark power factor (α) is chosen such that the videos watermarked by all methods have the same PSNR quality. This choice can be better understood through a simple example. Figure 6 shows a single video frame watermarked by the proposed method at $\alpha = 3.5$ and the same

Table 3 The robustness of the proposed method against scaling

Original Size	After Scaling	[3]	[13]	proposed
1920 × 1080	1280 × 720	0.3558±0.1125	0.3420±0.1933	0.3979±0.1771
1920 × 1080	352 × 288	0.2791±0.0969	0.2711±0.2547	0.3006±0.1525
1280 × 720	352 × 288	0.3498±0.1815	0.3726±0.1736	0.5418±0.0938
total		0.3282±0.1398	0.3285±0.2011	0.4131±0.1760

The average and standard deviation of the normalized correlation are reported for all methods for different scalings

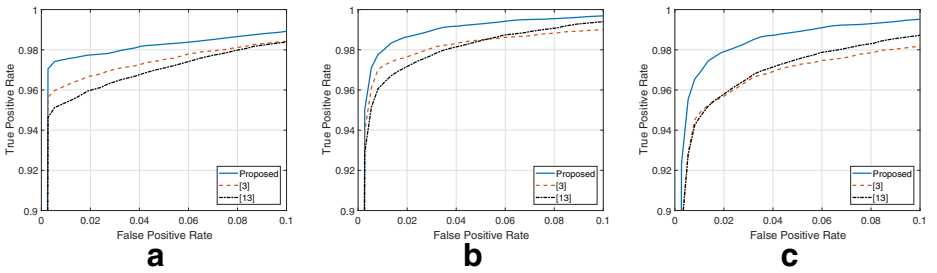

Fig. 8 ROC curve for the scaling attack **(a)** 1920 × 1080 → 1280 × 720 **(b)** 1920 × 1080 → 352 × 288 **(c)** 1280 × 720 → 352 × 288

frame watermarked by the method proposed in [3] at $\alpha = 36$ (authors' recommendation). As can be observed, the proposed method does not impose any distortion perceivable to HVS at all, while [3] with the recommended watermark power, impose perceivable distortion that can be perceived according to HVS. Therefore, to better compare the results, it is preferred to apply the PSNR in videos to adjust the α values.

3.1 Robustness against compression

The first experiment investigates the robustness of the proposed method against the lossy compression. All three video sets mentioned above are tested in this experiment. Every video is watermarked 30 times and then is compressed using an H.264/AVC encoder and decoder with the quantization parameter set to 30. All methods perfectly detect all the watermarked videos. Therefore, for a better comparison, the average and standard deviation of the normalized correlation for the watermarked videos are shown in Table 2.

Although the performance of all methods is too close to each other in smaller videos, the superiority of the proposed method in higher qualities is evident. Also the standard deviation of normalized correlation is always smaller than the other methods which shows that the proposed method is more robust against compression attack. The receiver operation characteristic (ROC) curves of three methods are plotted for better comparison. Figure 7 illustrates that the proposed method achieves a higher detection rate for a certain false alarm ratio; hence it is more robust to the compression attack.

3.2 Robustness against scaling

The robustness of the proposed method against rescaling of the watermarked video is investigated in the next experiment. To this end, two video sets of sizes 1920 × 1080 and

Table 4 The robustness of three methods against the rotation attack

Rotation Degree	[3]	[13]	proposed
1	0.4098 ± 0.1714	0.3947 ± 0.2690	0.4352 ± 0.1509
2	0.3853 ± 0.1744	0.3709 ± 0.2713	0.4233 ± 0.1289
3	0.3688 ± 0.0540	0.3473 ± 0.2814	0.4221 ± 0.0724
4	0.3369 ± 0.579	0.3405 ± 0.2687	0.4009 ± 0.0511
5	0.3271 ± 0.0153	0.3386 ± 0.1914	0.3896 ± 0.0152

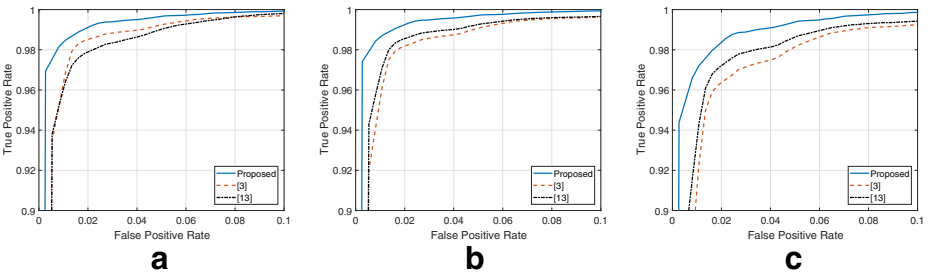

Fig. 9 ROC curve for the rotation attack. **(a)**, **(b)**, and **(c)** correspond to 3, 4, and 5 degree of rotation, respectively

1280×720 are applied. Videos of size 1920×1080 are rescaled to 1280×720 and 352×288, after being watermarked and compressed. Furthermore, 1280×720 videos are also rescaled to 352×288, after the compression. The average and standard deviation of the normalized correlation for all methods are reported in Table 3.

Table 3 indicates that for the all attacks, the normalized correlation of the proposed method lies above that of [3] and [13]. The ROC curves are presented in Fig. 8 to better comparison of three methods, where the curve of the proposed method always lies above that of other methods.

3.3 Robustness against rotation

To investigate the robustness of the proposed method against the rotation, every video is rotated for one to five degrees after being watermarked. This attack can be carried out in two scenarios. In the first scenario, the rotated video is cropped to scale back to the original size, while it is not cropped in the second scenario, resulting in the increase in the video size and reduction of the lost information. The first scenario is applied here to examine the robustness against the rotation attack, that is, the video is cropped and scaled back to the original size, after being rotated. The results are demonstrated in Table 4.

It can be inferred from Table 4 that the proposed method reaches higher peaks compared to the other ones for the all rotation cases. The ROC curves for three out of five rotation cases

Table 5 The robustness of the methods against combined attacks which includes rotation, scaling, cropping, and compression in terms of False Negative Rate

Video Size	Down Scale to	Method	1%	3%	5%	7%	10%	15%	20%
1920 × 1080	1280 × 720	Proposed	0.00	0.00	0.00	0.00	0.00	0.00	1.83
		[3]	0.00	0.00	0.00	0.56	0.74	2.68	3.22
		[13]	0.00	0.00	0.00	0.00	0.49	1.92	2.43
1920x1080	352x288	Proposed	0.00	0.10	0.53	1.07	1.68	2.98	4.06
		[3]	0.00	0.26	1.30	2.16	2.84	4.18	6.35
		[13]	0.00	0.21	1.19	1.88	2.44	3.25	5.76
1280x720	352x288	Proposed	0.00	0.00	0.02	0.29	0.91	1.44	2.41
		[3]	0.00	0.05	0.69	0.95	1.27	2.63	5.14
		[13]	0.00	0.02	0.38	0.77	1.01	1.84	4.26

Table 6 Comparison of methods in terms of watermarking computational complexity

Video Size	1920x1080	1280x720	352x266
Proposed	11.20	29.33	143.88
[3]	3.84	10.53	48.09
[13]	1.69	4.13	28.65

The values are given in FPS

are given in Fig. 9, where the proposed method exhibits a better performance compared to [3] and [13]. This advantage is better observable when the degree of the rotation being increased.

3.4 Combined attacks

In this section the robustness against the combined attacks is investigated. For this sake, videos of size 1920×1080 and 1280×720 are applied. The videos are rescaled after being watermarked, and then they are rotated three degrees in the same manner as stated in the previous section. They are cropped from both sides by 1, 3, 5, 7, 10, 15, and 20 percent of the original size. Finally, videos are compressed as stated in Section 3.1. The results of this experiment in terms of false negative ratio (FNR) are given in Table 5. The FNR values are given in percent.

3.5 Computational complexity

To compare the embedding complexity of the methods, all standard videos are watermarked three times and the watermarking speed is recorded in terms of fps. Results are derived running MATLAB 2017a on a system with Intel Core i7 960@3.2GHz processor and 16GB of memory, and are reported in Table 6 based on the resolution of the watermarked video. Results show that the proposed method is much faster than [3]. The values in the table shows the number of frames watermarked by the corresponding algorithm in one second. This computational gain is achieved due to shifting the watermark addition from the transformed domain to the spatial domain that removes the requirement of applying the forward and inverse transforms.

4 Conclusion

In this paper, a video watermarking algorithm is proposed to tackle the problem of video piracy.The proposed method uses the low frequency components of the U channel of the video frames for watermark embedding. To embed watermark, the coefficients of the first and sixth sub-bands of DTCWT has been employed. The proposed method is evaluated using different kinds of attacks. Experimental results show that the proposed method is highly robust against different attacks which includes compression, scaling, rotation, and the combination of them. To ensure the high perceptual quality of the proposed method, the watermark power is chosen using a subjective test. Also, thanks to the linear attribute of DTCWT, the mask of embedding is designed to be applied in the spatial domain. This reduces the timing process of our watermark embedding system significantly.

References

1. Amirjan P, Mansouri A (2017) Video watermarking using quantized coefficients and motion vectors analysis. In: 2017 Iranian conference on electrical engineering (ICEE). IEEE, pp 1560–1563
2. Asikuzzaman M, Alam MJ, Lambert AJ, Pickering MR (2012) A blind digital video watermarking scheme with enhanced robustness to geometric distortion. In: 2012 International conference on digital image computing techniques and applications (DICTA). IEEE, pp 1–8
3. Asikuzzaman M, Alam MJ, Lambert AJ, Pickering MR (2014) Imperceptible and robust blind video watermarking using chrominance embedding: a set of approaches in the dt cwt domain. IEEE Trans Inf Forensics Secur 9(9):1502–1517
4. Asikuzzaman M, Alam MJ, Pickering MR (2015) A blind and robust video watermarking scheme in the dt cwt and svd domain. In: Picture coding symposium (PCS), 2015. IEEE, pp 277–281
5. Bhardwaj A, Verma VS, Jha RK (2017) Robust video watermarking using significant frame selection based on coefficient difference of lifting wavelet transform. Multimedia Tools and Applications 77:1–20
6. Cedillo-Hernandez A, Cedillo-Hernandez M, Miyatake MN, Meana HP (2018) A spatiotemporal saliency-modulated jnd profile applied to video watermarking. J Vis Commun Image Represent 52:106–117
7. Chan PW, Lyu MR, Chin RT (2005) A novel scheme for hybrid digital video watermarking: approach, evaluation and experimentation. IEEE Trans Circuits Syst Video Technol 15(12):1638–1649
8. Chen W, Li X, Zhan S, Niu D (2018) Multimedia video watermarking algorithm using svd and secret sharing. In: 2018 2nd IEEE advanced information management, communicates, electronic and automation control conference (IMCEC). IEEE, pp 1682-1686
9. Coria LE, Pickering MR, Nasiopoulos P, Ward RK (2008) A video watermarking scheme based on the dual-tree complex wavelet transform. IEEE Trans Inf Forensics Secur 3(3):466–474
10. Das S, Banerjee M, Chaudhuri A (2018) An improved video key-frame extraction algorithm leads to video watermarking. Int J Inf Technol 10(1):21–34
11. Dawei Z, Guanrong C, Wenbo L (2004) A chaos-based robust wavelet-domain watermarking algorithm. Chaos, Solitons & Fractals 22(1):47–54
12. Jing L (2009) A novel scheme of robust and blind video watermarking. In: International forum on information technology and applications, 2009. IFITA'09. IEEE, vol 1, pp 430-434
13. Joshi AM, Gupta S, Girdhar M, Agarwal P, Sarker R (2017) Combined dwt–dct-based video watermarking algorithm using arnold transform technique. In: Proceedings of the international conference on data engineering and communication technology. Springer, pp 455–463
14. Kingsbury N (1998) The dual-tree complex wavelet transform: a new efficient tool for image restoration and enhancement. In: 9th European signal processing conference (EUSIPCO 1998). IEEE, pp 1–4
15. Kingsbury N (1999) Shift invariant properties of the dual-tree complex wavelet transform. In: 1999 IEEE international conference on acoustics, speech, and signal processing, 1999. Proceedings. IEEE, vol 3, pp 1221-1224
16. Kingsbury NG (1998) The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters. In: Proceedings of the 8th IEEE DSP workshop. Utah, vol 8, pp 86
17. Kingsbury NG (2005) Human vision [openstax-cnx web site]. http://cnx.org/content/m11084/2.5/
18. Kwitt R, Meerwald P, Uhl A (2009) Blind dt-cwt domain additive spread-spectrum watermark detection. In: 2009 16th international conference on digital signal processing. IEEE, pp 1–8
19. Lee MJ, Im DH, Lee HY, Kim KS, Lee HK (2012) Real-time video watermarking system on the compressed domain for high-definition video contents: practical issues. Digital Signal Processing 22(1):190–198
20. Leng L, Zhang J, Xu J, Khan MK, Alghathbar K (2010) Dynamic weighted discrimination power analysis in dct domain for face and palmprint recognition. In: 2010 international conference on information and communication technology convergence (ICTC). IEEE, pp 467–471
21. Lin ET, Delp EJ (2004) Temporal synchronization in video watermarking. IEEE Trans Signal Process 52(10):3007–3022
22. Liu J, She K (2010) Robust image watermarking using dual tree complex wavelet transform based on human visual system. In: 2010 international conference on image analysis and signal processing (IASP). IEEE, pp 675–679
23. Mabtoul S, Hassan E, Elhaj I, Aboutajdine D (2007) Robust color image watermarking based on singular value decomposition and dual tree complex wavelet transform. In: 14th IEEE international conference on electronics, circuits and systems, 2007. ICECS 2007. IEEE, pp 534–537
24. Miller ML, Bloom JA (1999) Computing the probability of false watermark detection. In: International workshop on information hiding. Springer, pp 146–158

25. Párraga C, Brelstaff G, Troscianko T, Moorehead I (1998) Color and luminance information in natural scenes. JOSA A 15(3):563–569
26. Pickering M, Coria LE, Nasiopoulos P (2007) A novel blind video watermarking scheme for access control using complex wavelets. In: 2007 Digest of Technical Papers. International Conference on Consumer electronics, 2007. ICCE. IEEE, pp 1–2
27. Rana S, Sur A (2015) 3d video watermarking using dt-dwt to resist synthesis view attack. In: 2015 23rd European signal processing conference (EUSIPCO). IEEE, pp 46–50
28. Rasti P, Samiei S, Agoyi M, Escalera S, Anbarjafari G (2016) Robust non-blind color video watermarking using qr decomposition and entropy analysis. J Vis Commun Image Represent 38:838–847
29. Sake A, Tirumala R (2018) Bi-orthogonal wavelet transform based video watermarking using optimization techniques. Materials Today: Proceedings 5(1):1470–1477
30. Stelter B, Stone B (2009) Digital pirates winning battle with studios. The New York Times, New York
31. Terzija N, Geisselhardt W (2004) Digital image watermarking using complex wavelet transform. In: Proceedings of the 2004 workshop on multimedia and security. ACM, pp 193–198
32. Wang Y, Pearmain A (2006) Blind mpeg-2 video watermarking robust against geometric attacks: a set of approaches in dct domain. IEEE Trans Image Process 15(6):1536–1543

Reza Esfahani received the B.Sc. and M.Sc. degree in Electrical Eng. Now he is a PhD. candidate at Imam Hossein University. His PH.D topic is multimedia security and watermarking. He has published a few papers and participated in many scientific conferences.

Mohammad Ali Akhaee (S'07–M'07) received the B.Sc. degree in Electronics and Communications Eng. from the Amirkabir University of Technology, Tehran, Iran, and the M.Sc. and Ph.D. degrees from the Sharif University of Technology, Tehran, Iran, in 2005 and 2009, respectively. He is currently an Assistant Professor with the College of Engineering and the Director of the Secure Communication Laboratory, University of Tehran, Iran. He has authored or coauthored more than 60 papers, and holds one Iranian patent. His research interests include the area of signal processing, in particular multimedia security, data hiding, and machine learning. Prof. Akhaee was the Technical Program Chair of EUSIPCO'11 and the Executive Chair of ISCISC'14 and Financial Chair of RTEST'18. He received the Governmental Endeavour Research Fellowship from Australia in 2010 and the Governmental award from Ministry of Information and Communication Technology (ITC) from Iran in 2017.

Zynolabedin Norouzi received his BS and MS degrees in applied Mathematics in 2004 and 2007 from University of Tehran, respectively. He has Ph.D degree in applied mathematics - cryptography in 2012 from Kharazmi University. His research interest includes cryptography and steganography. He has published several technical papers and participated in many scientific conferences.