Chapter

DWT-Based Data Hiding Technique for Videos Ownership Protection

Farhan Al-Enizi and Awad Al-Asmari

Abstract

This chapter proposes a wavelet data hiding scheme for video authentication and ownership protection. A watermark in the shape of a logo image will be hidden. In this research, a discrete wavelet transform (DWT) process is implemented using orthonormal filter banks, where the Y components of the YUV color space of the video frames are decomposed using DWT, and a watermark is inserted in one or more of the resulting sub-bands in a way that is fully controlled by the owner. Then, the watermarked video is reconstructed. The filters used for the DWT decompositions are randomly generated to increase the security of the algorithm. An enhanced detection technique is developed to increase the reliability of the system. The overall robustness of this scheme is measured when common attacks are applied to the test videos. Moreover, the proposed algorithm is used with the high-efficiency video coding (HEVC) technique to examine the whole performance. Furthermore, a selective denoising filter is built to eliminate the effect of the noise. The simulation results show that the proposed algorithm achieves well under both the visual and the metric tests. Moreover, it performed well against intentional and unintentional attacks. The average normalized correlation achieved is 97%, while the mean peak signal-to-noise ratio (PSNR) is 45 dB.

Keywords: watermarking, filter banks, data hiding, pyramid transform, video coding, video attacks

1. Introduction

The digital age brought new technologies and services to people, industry, and governments. The digitization process covered all types of information being exchanged especially images and videos. The exchange of these forms of multimedia became faster and easier with the new communication network capabilities. On the other hand, that made it easier to steal or use these forms of multimedia illegally. Those concerns opened new horizons in the field of multimedia security especially data hiding [1]. Data hiding has many types; a useful classification is to divide data hiding into watermarking and steganography. New techniques used in video processing brought new challenges and difficulties to the data hiding methods. Moreover, the new compression techniques especially the high-efficiency video coding (HEVC) or H.265 are added to these challenges and difficulties to have

reliable, secure, and robust data hiding methods [2, 3]. Various watermarking schemes that use different techniques have been proposed over the years [4–9]. To be effective, a watermark must be imperceptible within its host, extracted with ease by the owner, and robust in the face of both intentional and unintentional distortions [7, 10, 11]. In specific, discrete wavelet transform (DWT) has wide applications in the different areas of image and video processes such as compression, noise reduction, and watermarking [12]; this is attributed to its characteristics in space-frequency localization, multi-resolution representation, and superior human visual system (HVS) modeling [5]. The robustness is a very important aspect in data hiding or watermarking. To achieve the highest levels of robustness, new methods and techniques should be introduced and optimized at both the sender and receiver sides. Furthermore, the detection process should be enhanced to meet these requirements.

In this research, a video watermarking process that depends on the discrete wavelet decompositions will be developed. Moreover, the detection process will be enhanced through statistical derivations. The security will be maintained through the adoption of random filter banks, the study of the motion and motionless scenes in the video frames, and the spread spectrum generation of the watermarks. The overall technique has to meet the requirements of visual quality, security, robustness, and computational complexity.

2. Proposed watermarking technique

In this section, we introduce our digital video watermarking technique for the purpose of authentication and ownership protection. The proposed technique is aimed at achieving reasonable degrees of robustness, visual quality, and security. The embedding technique involves two stages: first, a decomposition process and then a hiding process. The watermark can be any binary sequence; normally a binary image of a specific size is used. The encoded videos can be in any color space; in our case, YUV space is used. It is possible to perform the hiding process in any of the three components: Y, U, and V. In this work, the luminance Y frames are used as host images for the data hiding process; that is, the hiding of the watermark will be performed in one or more of the sub-bands that result from the discrete wavelet analysis process. Choosing the wavelet filters is an important aspect in the efficiency of the reconstruction process; special types of filters are the randomly generated orthonormal filter banks [13]. These filter banks can be generated randomly depending on the generating polynomials; hence, by generating random numbers for the polynomial coefficients, it is possible to build multiple filter banks that are used for the different stages of our decomposition processes. The orthonormal analysis and synthesis filters can be generated in different ways; for our technique, having large side-lobes is preferred. This enables us to hide more energy in the medium frequencies of the image; in doing so, we construct a more robust way that can counteract the effects of different image processes, which take place intentionally or unintentionally over the course of the handling process. Each filter bank that is generated is used for one level of the DWT analysis and synthesis processes. Moreover, the number of the levels and the structure that is followed during the analysis process are controlled by the owner. It is well known to the image processing community that the medium-frequency bands are preferred for hiding. This will avoid hiding in the lower-frequency bands where most of the energy is concentrated and the higher-frequency bands where the possibility of losing the data is high due to compression processes. Furthermore, the possibility of

using more than one sub-band rather than a single sub-band is there; this method is useful in having a robust method against the nonlinear collusion attack.

There are many scenarios that can be followed for the embedding process; one of them is to embed the data which is our binary watermark using a generated pseudorandom sequence [14]. This method depends mainly on doing the watermarking process by converting the original binary watermark image Q to some sort of a binary sequence S of a specific length M; in this case, the data pixels are given the value +1, whereas the background pixels are given the value −1. Furthermore, a pseudorandom sequence P of the same length M as our watermark sequence is generated using a secret key; likewise, this sequence is represented by values that are either +1 or −1. The DWT coefficients of the decomposed sub-bands that will be used for the hiding process are represented as a matrix Q_1 of the same size as our watermark. Moreover, it can be written as a vector T of length M. The binary watermark is hidden into this vector T, and that will result in a new vector that is called T' according to the rule that is shown in this equation:

$$t_i' = t_i + \alpha * p_i * s_i, for \ \ i = 1, 2 \ldots M \tag{1}$$

where α is a numerical factor which represents a weighting constant that determines the strength of the processed watermark. This number is chosen in such a way to offer a trade-off between the required robustness and the acceptable visual quality. Moreover, choosing this weighting factor should take into consideration many elements in image processing techniques such as the compression standard that is used and its intensity, the smooth features or the textures that are there in the image, and the algorithm that is followed when doing the detection process. Furthermore, how much energy content is there in the wavelet sub-bands must be considered at the hiding stage. One way to get the numerical magnitude factor is to have a comparison process between the energy of the original coefficients of the host DWT sub-band Q_1 and energy content of the original watermark image Q elements according to this empirical formula:

$$\alpha = 2 * \sqrt{\frac{E(Q_1)}{E(Q)}} \tag{2}$$

where $E(Q_1)$ represents the energy content of the original wavelet coefficients, while $E(Q)$ represents the energy content of the watermark image Q; the energy was computed by taking the sum of the squared elements. The manipulated wavelet coefficients according to our hiding process are used then depending on their respective locations to reconstruct and build the watermarked image frame. The overall hiding process of a binary watermark for a Y frame is shown in **Figure 1**. It is clear from this figure, and this, in fact, depends on the decomposition structure that is followed that the low-low (LL) frequency area of the decomposed image is not used for our embedding process. This area or band is called the decimated image normally, and it results in both the pyramidal and DWT decompositions. It is clear that this band or image has most of the information or energy of the original image frame; the other images in other bands are normally called the error images, and they have lower energy content. In fact, they represent other bands depending on the analysis filters which are the low-high (LH), high-low (HL), and high-high (HH) bands. These bands offer better places for the hiding process.

The watermark, which is primarily a binary image, can be embedded in any of the frames of the host video; moreover, the frames can be chosen in a fully controlled selective way. The degree of randomness that is achieved is up to the user

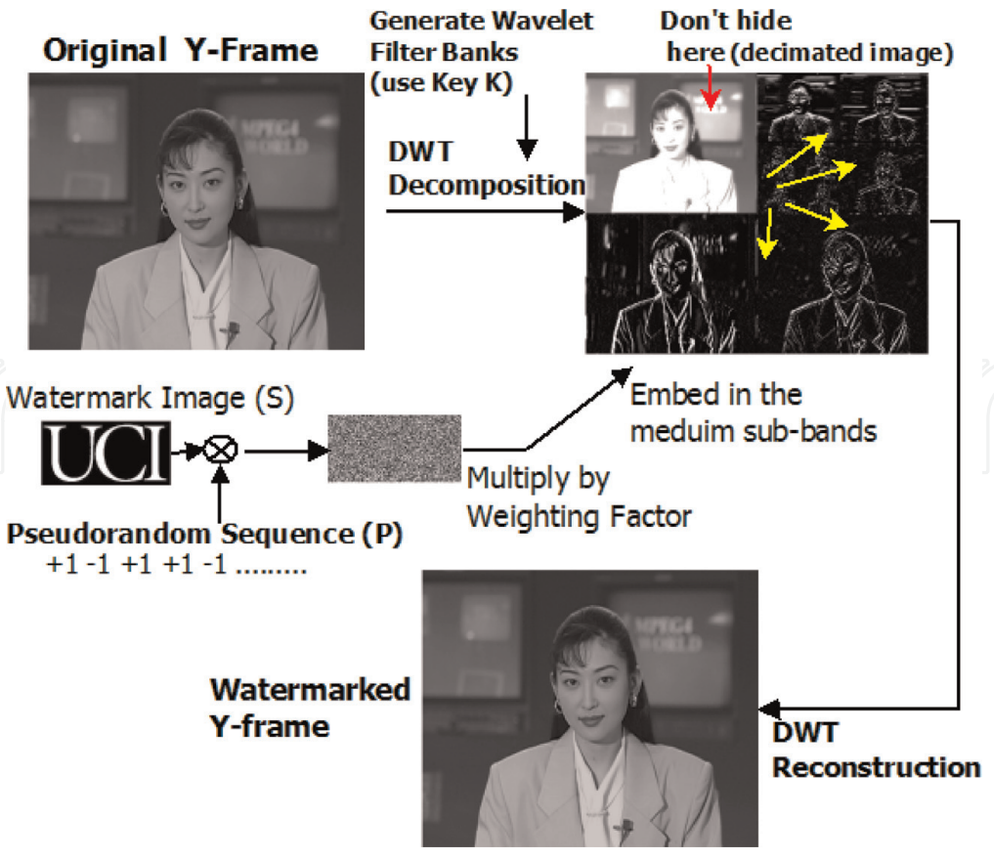

Figure 1.
The block diagram of the proposed watermarking method.

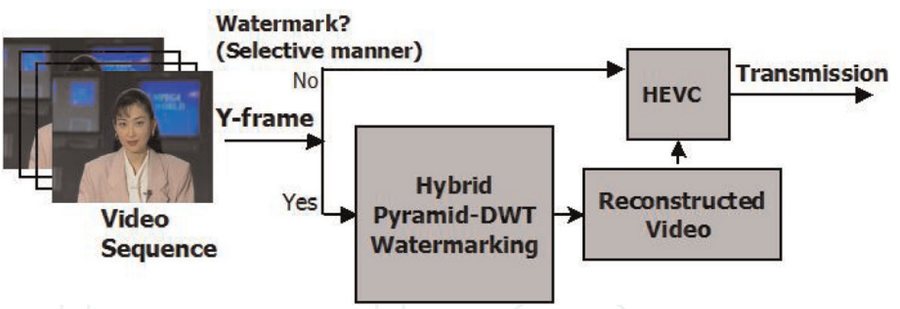

Figure 2.
The block diagram of the watermarking process with the application of HEVC process.

who is the sole owner. Furthermore, the security of the system depends partially as well on the degree of the randomness of the pseudorandom sequence that is used in the encoding process. On the other hand, the Y components of the color space were chosen intentionally because they have higher resolution and therefore higher hiding capacity, but we have to keep in mind the fact that the U and V components likewise can be used. As we mentioned in the introduction, our techniques will be used when the HEVC process is applied; **Figure 2** shows the proposed hiding process when the HEVC or H.265 process is applied to the video that is watermarked.

2.1 1D discrete Fourier transform

When designing a robust and dependable embedding system, security concerns always come to the forefront. Hiding the same watermark in a repetitive manner to each and every frame of the host video may cause a problem of maintaining the

statistical invisibility, which is an important condition of every security system [15]. Moreover, applying independent watermarks to each and every of these frames also causes a security problem if these frames have few or no motion areas inside them; these motionless regions in successive video frames may be statistically compared or averaged to remove independent watermarks. Attacks of such kinds are normally called collusion attacks. The inter-frame collusion attacks, for instance, exploit the repetition in the video frames and their scenes or in the watermarks themselves to produce a false copy of the video that does not have any watermarks; these attacks can be divided into watermark estimation remodulation (WER) attack and frame temporal filtering (FTF) attack [16]. Classifying the video frames according to the amount of motion in them is useful in this regard. The motion in videos is a relative one, since most of the videos have motion, but what interest us here are the amount of this motion, how fast this motion is, the relative motion with respect to the surroundings, and the distribution of this motion across the frames. Most of the video compression techniques use inter-frame motion estimations to encode the frames; however, there are other methods that can be used to detect static and dynamic scenes in videos. One method can be built depending on the *1D discrete Fourier transform* (DFT). The 1D DFT in temporal direction performs a transformation process of a group of pictures (GOPs) into a temporal frequency domain; in the resulting domain, both the video frames spatial and temporal frequency information exist in the same resulting frame. Higher frequencies are a reflection of the fast motion from one frame to other frames [17]. The 1D DFT of a video *f(x,y,t)* that has a specific size of MxNxT, in which MxN is the size of each of the video frames and T is the number of the video frames that are grouped in one GOP, is given by

$$F(u,v,\tau) = \sum_{t=0}^{T-1} f(x,y,t) e^{-j\,2\Pi(t\tau/T)} \tag{3}$$

where u and v represent the spatial domain of the video frames, while τ represents the temporal domain of these frames. Normally the GOPs are taken as five frames or a close number. Depending on that, a group of the so-called spatiotemporal frames can be constructed for the Foreman video. Twenty-five frames of the Foreman video were transformed using this method of the 1D DFT, and since the DFT is a symmetric process in one GOP, so it is logical to show only the first spatiotemporal frame of each of those groups of pictures. **Figure 3** shows the first frame of the Foreman video, while **Figure 4** shows the 5 temporal frames of this

Figure 3.
The first frame of the Foreman video.

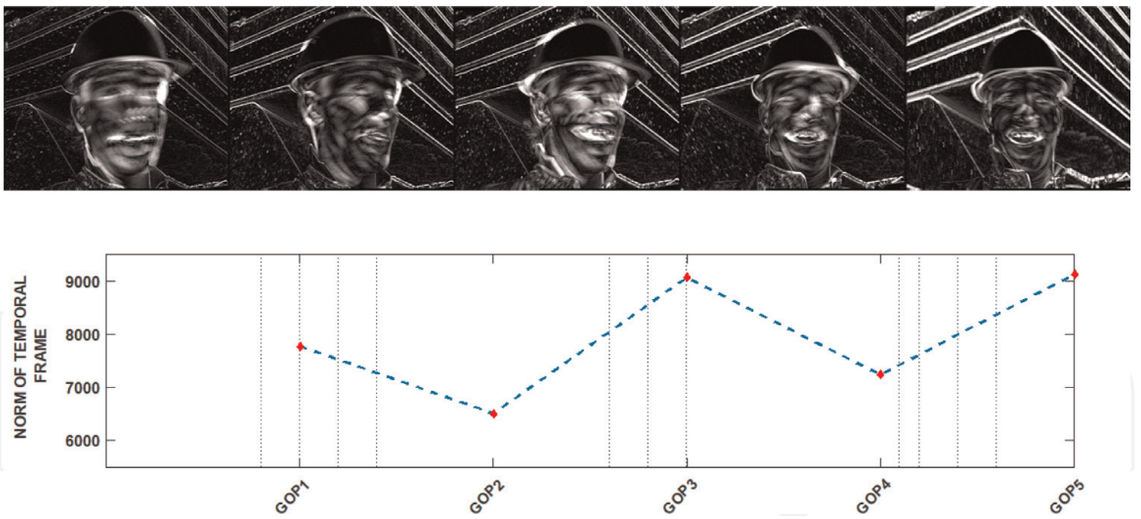

Figure 4.
The 1D DFT of 5 GOPs of 25 frames of Foreman video and their corresponding norms.

video that were evaluated from the original 25 frames; their norms are shown as well. The edges that are seen in these frames correspond to high frequencies which reflect the motion in temporal domain and how this motion in each frame is distributed; furthermore, the values of the evaluated norms reflect how much and how fast the motion in each GOP is. For instance, the intensity of the edges shows the movement of the head and the relative motion with respect to the building; it can be seen that the background has some motion that corresponds to a moving camera which is exactly the case here.

Depending on the previous analysis of the videos using the 1D DFT and the classification of the video frames into dynamic and static frames, a significant enhancement can be added to the hiding process in terms of both security and reliability. Using this analysis, different binary watermarks will be embedded in motion frames, and the same binary watermark will be embedded in motionless ones. In fact, since we need to have some repetition of the watermark to enhance the detection process, this method helps us without weakening our algorithm due to statistical estimation methods that are used in steganalysis, for instance; moreover repeating the watermark in motionless frames increases the cohesion of the watermarked video sequence. Furthermore, the bands being used are not confined to the high-frequency ones; the effect of averaging and collusion attacks is reduced as well. Using 1D DFT to establish motion information is not the only way that can be used; 3D DWT, for instance, can be used to construct spatiotemporal components of videos frames. Choosing the proper method to determine motion in frames depends in the first place on the application and other elements such as computational complexity. Since we are only looking for a method to estimate motion but not in a strict and precise way, using 1D DFT meets our needs at this stage.

3. Watermark detection process

The extraction process depends mainly on the hiding process, and so we are performing a reverse process. This is a blind watermarking method; hence, knowing the original watermark image is not a requirement, but, still, knowing the reconstruction synthesis filter banks and the generated pseudorandom sequence is required to extract our hidden watermark. To get the hidden watermark, a prediction and estimation process of the original values of the pixels is required [14]. This process should also take into account that different types of processing will take

place such as the lossy compression, the additive noise, and the geometrical operations; this, in turn, renders the detection process a challenging one. Furthermore, security concerns arise as a critical point; this is reflected in the attempts of attackers to know or destroy the hidden watermark. To cope with these difficulties, an enhanced detection and estimation process is developed.

A noise elimination method can be developed to estimate the original pixels; to do that, the extracted coefficients can be smoothened using a spatial convolution mask of size 5x5. In fact the 5x5 mask gave higher performance than the 3x3 mask when our videos were subjected to noise and compression. Moreover, the selective denoising filter which is presented in Section 4 gave good results in removing the noise and smoothening the extracted image. Using a subtraction operation as opposed to the additional one in Section 2 and setting the positive and negative values to +1 and −1, respectively, a coarse version of the watermark can be extracted. The enhanced detection process is then set to use multiple extracted watermarks, which were embedded randomly in different video frames in the first place, for our final estimation process. It was shown in the previous section that either the same watermark or multiple watermarks can be used depending on the changes in the scenes; moreover it was shown that the 1D DFT is helpful in determining these changes. This means that we are not sacrificing the security when using the same watermark in a random manner; on the contrary we are increasing robustness against detection or manipulating attempts. Let us assume that the extracted watermarks are grouped in a set $W = w_1, w_2, \ldots w_n$. To choose the set of watermarks that can be used in the final estimation process, cross-correlation test can be performed between every two extracted watermarks w_i and w_j. The normalized cross-correlation coefficient between two matrices A and B is given according to the following equation:

$$R = \frac{\sum_m \sum_n \left(A_{mn} - \overline{A}\right)\ \left(B_{mn} - \overline{B}\right)}{\sqrt{\left(\sum_m \sum_n \left(A_{mn} - \overline{A}\right)^2\ \right)\left(\sum_m \sum_n \left(B_{mn} - \overline{B}\right)^2\right)}} \tag{4}$$

where $\overline{A}$ and $\overline{B}$ are the means of A and B, respectively. The attacks that the videos are subjected to are of different natures and scopes; they can be divided into geometrical, statistical, additive noise, etc. Hence, the watermarks that are extracted can be talked about as noisy versions of the originally hidden ones or, equivalently, noisy signals. The cross-correlation test gives good indication of the similarity between two signals, and this can be applied to our extracted watermarks which are expected to have some sort of similarity. Depending on this statistical analysis, it is possible to establish a new set of extracted watermarks W_1.

Depending on the resulting cross-correlation value, it is possible to get the set of coefficients that can be used in our final decision-making process. Hence, the final watermark set W_1 can be established. On the other side, if the correlation value was low, that means that the coefficients are so corrupted, and therefore they will be excluded from our final set. This cross-correlation process can be seen in **Figure 5** where **Figure 5(a)** shows a plot of the cross-correlation matrix between two sets of coefficients that are highly correlated, and that means they can be included in our final set, while **Figure 5(b)** shows the opposite of that, where these coefficients are corrupted. To establish a good estimation process, a threshold value should be defined for the cross-correlation value, and the decision can be done accordingly.

Since the cross-correlation between binary images is a measure of similarity between these images, this tells us that flipping the value of any pixel will reduce this similarity. If $w_i \in W$, then a cross-correlation process is performed between w_i

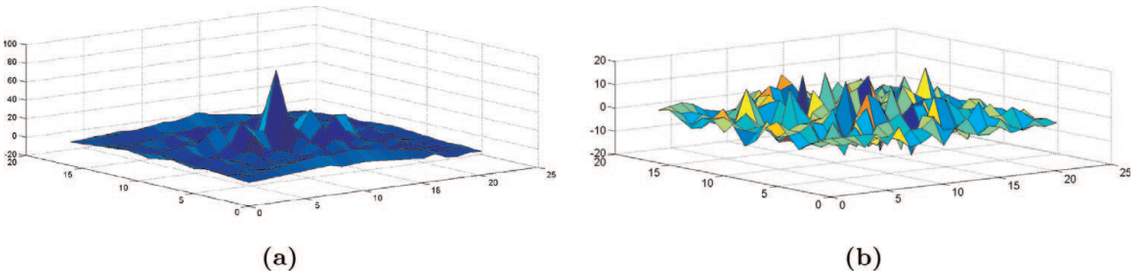

Figure 5.
3D plots of the cross-correlation matrices of two extracted watermarks.

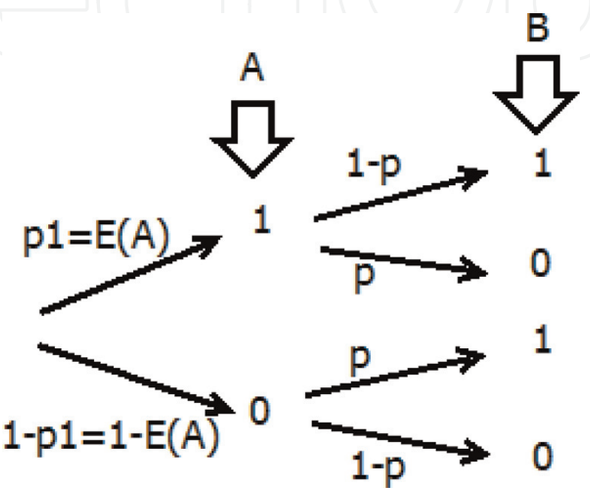

Figure 6.
Expected values of the input and output binary images.

and all the other extracted watermarks in the set; then the average cross-correlation parameter is evaluated. The same process is done for all the watermarks in the set. A set that includes each extracted watermark and its corresponding average correlation value is established. Then, by establishing a threshold value h for the average cross-correlations, the extracted watermarks that do not achieve the threshold test are excluded from the new set W_1. The final extracted watermark w_e can be evaluated by performing an averaging process on the watermarks in the set W_1, where

$$w_e = Ave\{W_1\} \tag{5}$$

Doing an averaging process is attributed to the fact that binary sets follow specific statistical pattern. The correlation coefficient R between any two arbitrary matrices A and B is given in Equation 4; the mean value of a binary image A is at the same time the expected value of A or $\mathrm{E}(A)$. Assuming that at the input, the probability of 1 is p_1 and that the probability of flipping of the value is p as shown in **Figure 6**, then

$$\overline{A} = E(A) = p_1 \tag{6}$$

Moreover, the probability of having 1 at the output $\overline{B} = p_1 * (1-p) + (1-p_1) * p$ and by taking Equation 6 into consideration, this equation can be rewritten as

$$\overline{B} = E(B) = E(A) + (1 - 2 * E(A)) * p \tag{7}$$

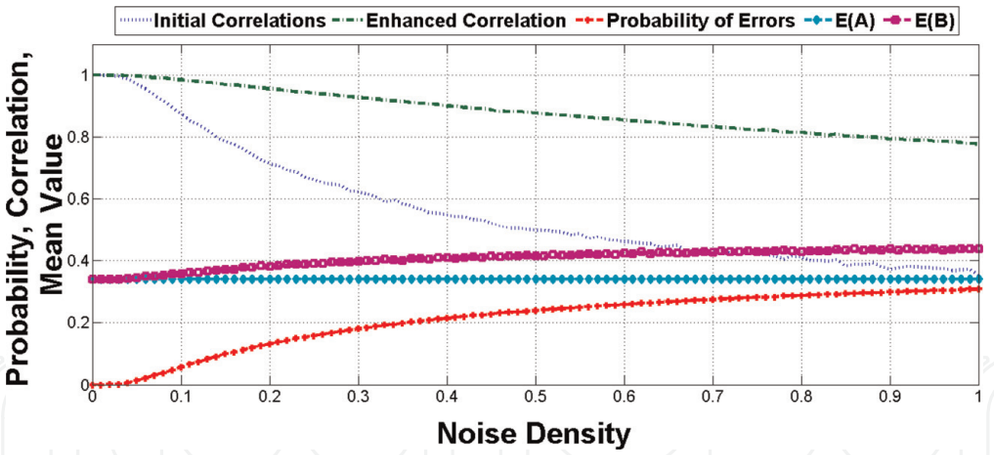

Figure 7.
The enhanced correlations vs. noise density.

Assuming that we are using a specific binary watermark, then the input matrix A is constant during our watermarking process. This means that in the above equation, the flipping probability of the pixels p is the sole variable. Moreover, by having a comparison between Eqs. (4) and (7), it can be seen that the correlation between the two matrices A and B is dependent on the flipping probability of the pixels; hence, the flipping effect is reduced somehow by doing the averaging process of the extracted watermarks. An enhanced version can be built as far as p is not equal to the value 0.5 which corresponds to a unity entropy value. To demonstrate this analogy, **Figure 7** shows the changes in these parameters when a random binary watermark is subjected to Gaussian noise with zero mean and different variances; in this figure, the variance of noise is represented by the term density for the illustration and clarification purposes.

4. Noise removal selective filter

One of the challenging aspects in video encoding and watermarking is the additive noise that results in distorted video streams. The nature of the additive noise depends primarily on the source of this noise. Not only the additive noise tends to distort the visual quality of the video in question, but it also has its noticeable impacts on the watermarking process. One type of noises that is common in video processing techniques is the salt-and-pepper (S&P) noise. This type of noise could be added to the video frames during the transmission process when the communication channels, in a sense, are noisy, or it could be a result of the hardware-generated errors during the encoding and decoding processes. Removing the noise without disturbing the watermarking process on the one hand and preserving the visual qualities on the other hand is a challenging process. As far as the watermarking process is concerned, it is useful to check the effects of both the additive noise and the removal process on our data hiding process. Many methods were proposed to eliminate the noise or enhance the visual appearance of the images [18, 19]; these methods depend mainly on the idea of median filters. The normal median filters, for example, which are used to eliminate the salt-and-pepper noise in images, do in fact filter the whole image regardless of the presence or absence of the noise in a certain area. This process reduces the original resolution of the image to a great extent in such a way that the qualities of high-definition (HD) videos are lost. This means that our watermarking process would not achieve the visual quality

condition. In this research, a noise detection process that depends on the absolute differences between a pixel aij and its surrounding pixels is proposed. In order to enhance the detection process, the variance of the pixels in the surrounding window is calculated. This step is important because of false detections, especially at edged and textured details of the image where the absolute difference value could be high, while the region is noise-free. This method takes into account the fact that such variances are dramatically high at these locations. However, this is not the case around noisy pixels in general where some sort of consistency is there. The proposed method for noise detection and elimination process involves the following steps:

1. For each pixel a_{ij}, a sub-window of size 3x3 around this pixel is taken.

2. The absolute differences between the pixel a_{ij} and the surrounding pixels are calculated.

3. The arithmetic mean (AM) of the calculated differences for a given pixel a_{ij} is computed. The AM is then compared with a threshold value t to detect whether the pixel a_{ij} is informative or corruptive.

4. The 3x3 pixel window is converted to an array, and then it will be arranged in an ascending order. The largest and the smallest values will be eliminated. This will help in removing other noisy pixels in the surrounding window. The resulting array will be denoted L. The variance of the pixels in the array L is computed and denoted as V.

5. A comparison will be performed between AM and V on one side and their respective thresholds on the other side:

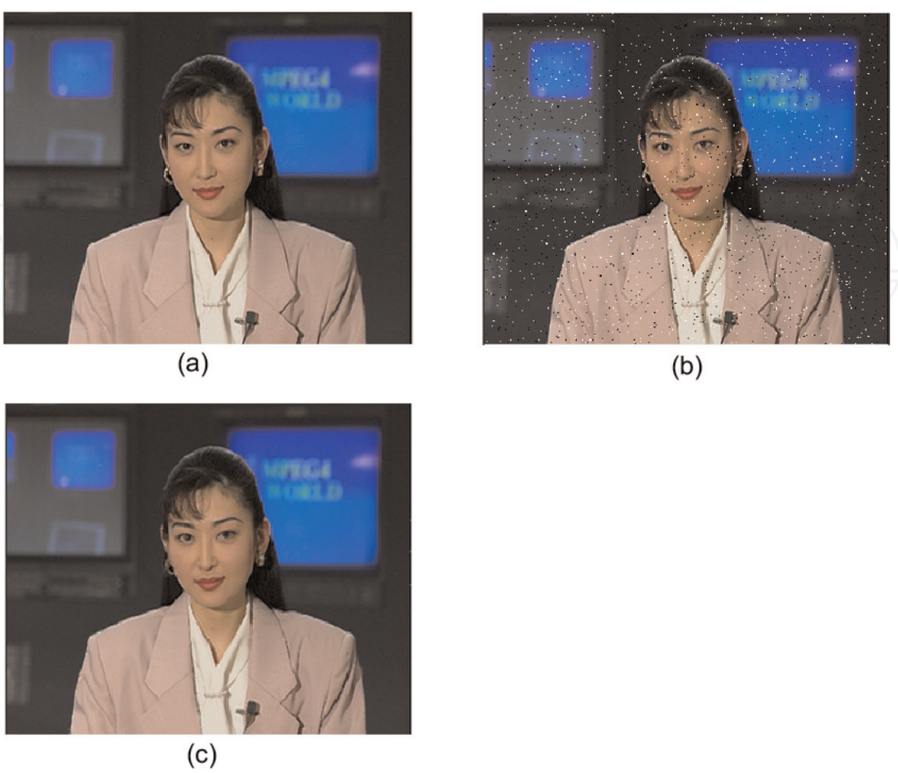

Figure 8.
(a) Original Akiyo frame; (b) 2% S&P noisy Akiyo frame; and (c) the denoised frame.

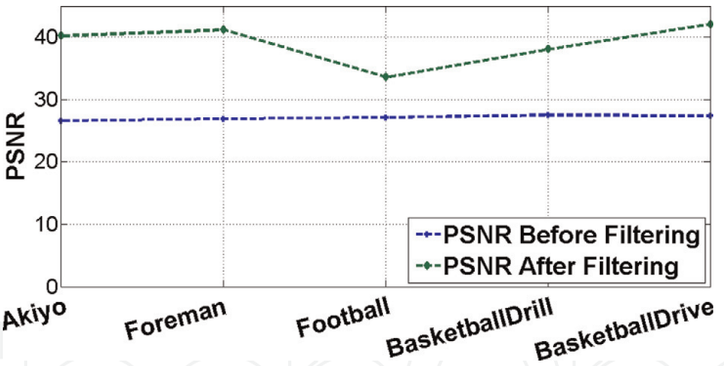

Figure 9.
PSNRs of standard videos before and after denoising process.

- If *AM* is greater than *t* and *V* is less than the variance threshold, then do the elimination process by replacing the noisy pixel by the median of the surrounding pixels in the window.

- Otherwise, do nothing. In this case, either there is no noise, or the pixel in question is on one of the edges of the image, and nothing should be done accordingly. Smooth and textured images perform differently with respect to noise; it is easier to remove noise, specifically salt-and-pepper noise from smooth images.

The arithmetic mean (AM) threshold is a user-defined value between the minimum and maximum pixel values (0.255) which are used to distinguish an informative pixel from a noisy one. On the other hand, the variance *V* can take larger values, and its threshold value can be determined accordingly. In fact, its value depends on the images themselves whether they were textured or smooth ones. The original Akiyo frame, a noisy version of this frame with salt-and-pepper noise of 2% density, and the same frame after the denoising process are shown in **Figure 8**. **Figure 9** shows the peak signal-to-noise ratio (PSNR) values of the noisy and denoised versions of the standard videos: Foreman, Akiyo, Football, BasketballDrill, and BasketballDrive.

5. Experimental results

In this section we demonstrate the performance of our algorithm using our proposed method on different standard videos with and without HEVC process, under different attacks. Furthermore, it will be compared with the method in [20]. Watermarked and unwatermarked versions of a frame of BasketballDrill video (832 × 480 pixels) are shown in **Figure 10**. The embedded and extracted watermarks of size 15 × 26 are shown in **Figure 11**; in fact, they are enlarged for illustration purposes.

Our algorithm performance will be evaluated in terms of PSNR between the original and the watermarked videos and the normalized correlation (NC) between the original and the extracted watermarks for the standard videos: Foreman, Akiyo, Football, BasketballDrill, and BasketballDrive. For the CIF (352 × 288) videos, a 9 × 11 watermark was used, while for the other two videos, the watermark in **Figure 11(a)** was used. In these tests, 100 frames were watermarked. **Figure 12** shows the NC of the extraction process; moreover, **Figure 13** shows the enhanced

Figure 10.
The first frame of (a) original BasketballDrill frame and (b) watermarked BasketballDrill frame.

Figure 11.
(a) Original watermark and (b) recovered watermark.

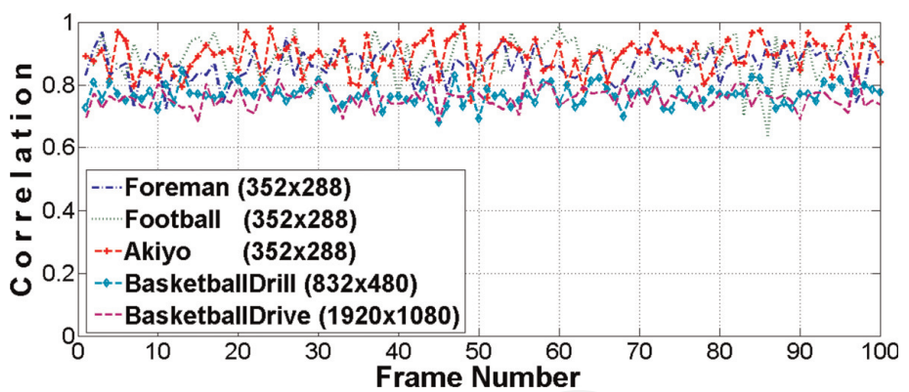

Figure 12.
Normalized correlations of the proposed watermarking process.

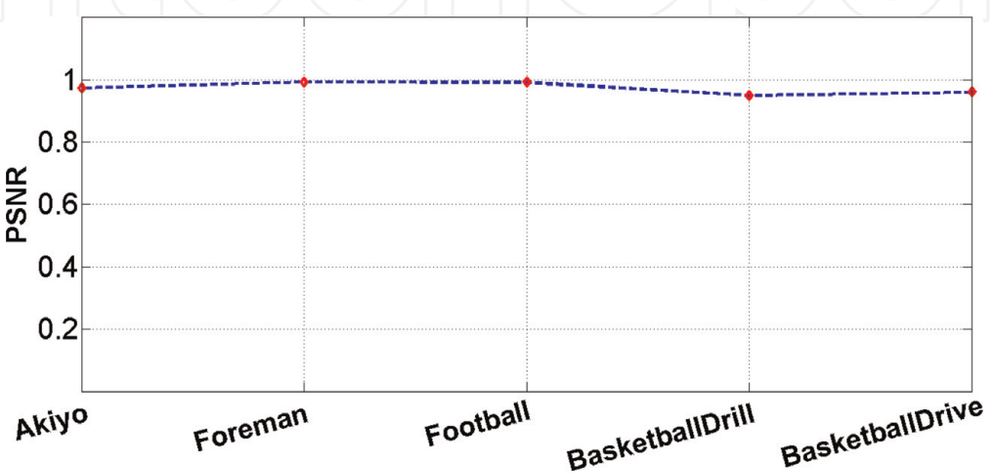

Figure 13.
Correlation values of the extracted watermarks using our detection process.

correlation values using our detection algorithm, while **Figure 14** shows the PSNRs of the reconstructed frames. **Figure 15** shows the PSNRs for the proposed watermarking method and the method of [20] when different numbers of frames of Football video were used. The method of [20] gives a maximum PSNR of 30 dB for the Football video, while our method gives an average of 42 dB; moreover, the performance in terms of NCs for the Football video stream using our method and the method in [20] was evaluated. The NC of method in [20] has an average value of 0.73, while our method gave a smooth performance with an average value of 0.99; this was shown as well in **Figure 13**.

For further investigation and evaluation of the robustness of our technique, the test videos will be subjected to some familiar attacks. These include additive noise, cropping, sharpening, rotating, frame averaging process, and HEVC compression. The attacks have the following characteristics:

- The additive noise will be Gaussian with a mean of 0 and a variance of 0.01.
- The salt-and-pepper noise has 1% noise density.
- Twenty-five percent of the even frames will be cropped.
- All the frames will be sharpened.
- The even frames will be rotated 1 degree counterclockwise.
- Ten random frames will be averaged with their respective successors.

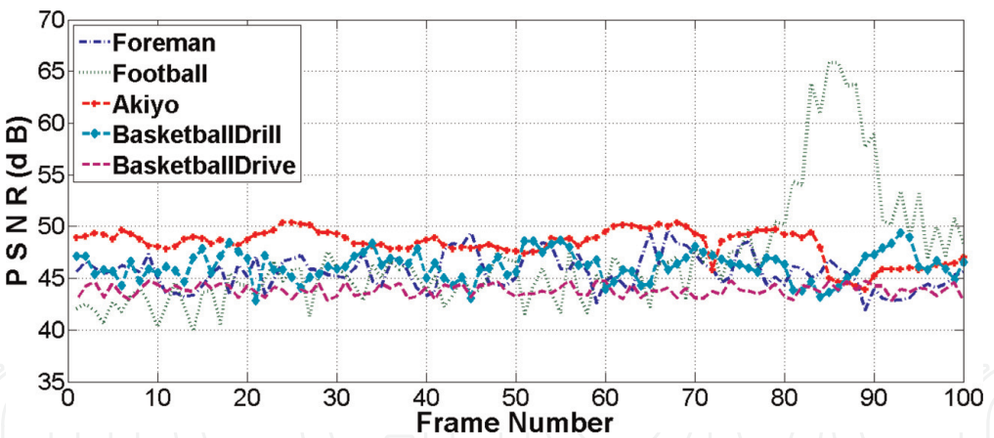

Figure 14.
PSNRs of the test videos when the proposed watermarking process is used.

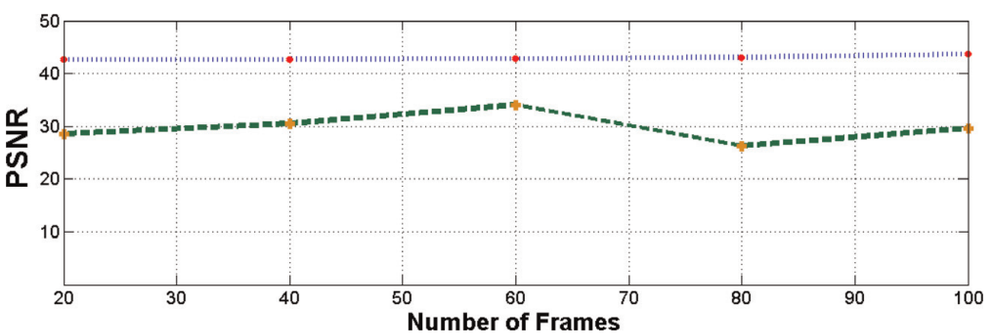

Figure 15.
PSNRs of the proposed method (the blue line) and method of [20] (the green line) when different numbers of frames of the video sequence Football are used.

The results for each attack for the different standard videos are shown in **Figure 16**.

To ensure the robustness of the algorithm, it was tested with the application of HEVC process. HEVC process with a quantization parameter (QP) value of 20 was applied to 100 frames of the test videos. Different compression ratios will result depending on each input video when this quantization factor is used. First, the watermarking process was applied to the test video frames and without applying the enhancement process. The PSNRs are shown in **Figure 17**, and the NCs at the watermarked frames are shown in **Figure 18**. A significant observation here is that lower values will result in the frames between 80 and 90 for the Football video, and

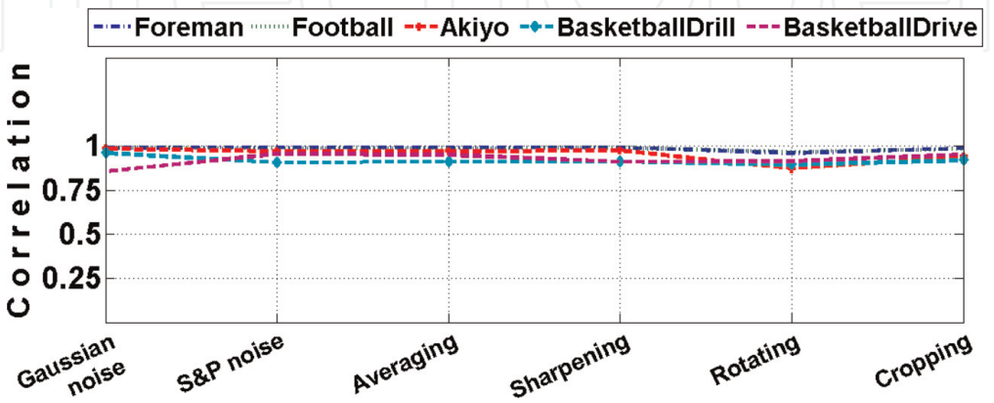

Figure 16.
Performance of the watermarking process under common aggressive attacks.

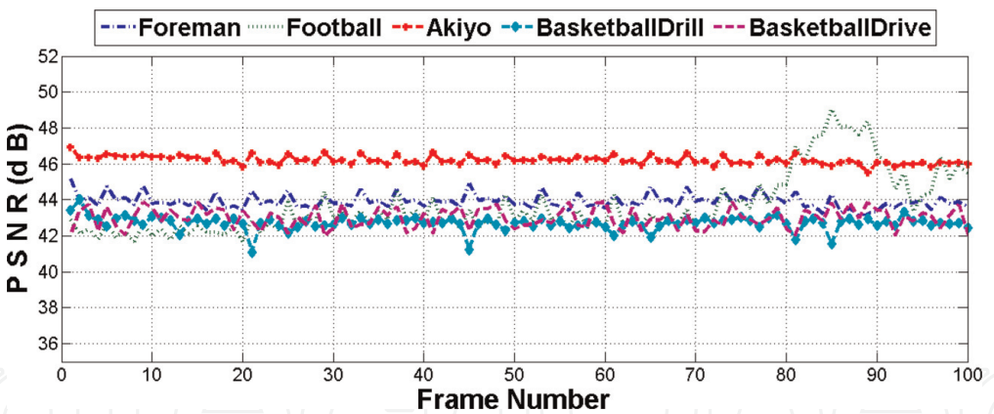

Figure 17.
PSNRs of the proposed watermarking method when HEVC process is applied.

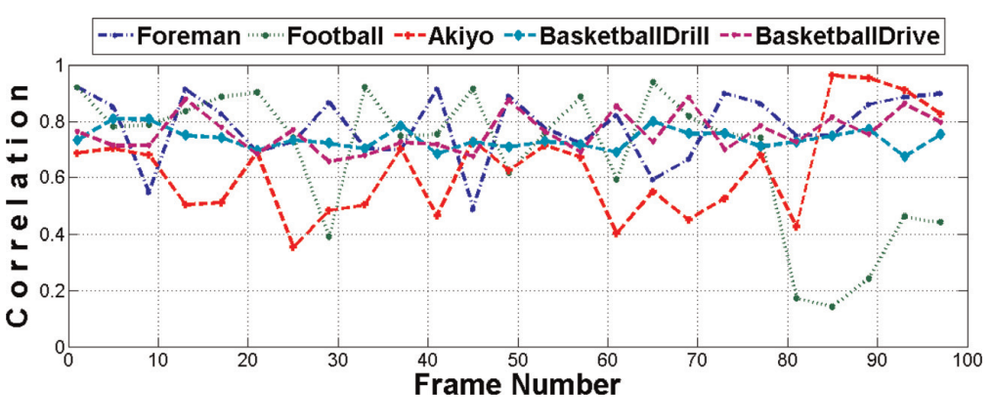

Figure 18.
Normalized correlations of the proposed watermarking method when HEVC process is applied.

this can be attributed to the blurring effects in these frames as a result of the fast panning of the camera. This, in turn, would result in lower energy in the DWT coefficients, and it is possible to overcome this phenomenon by our enhanced detection process as we will see later.

The correlations, moreover, were evaluated when the proposed detection process was applied. The watermarks were embedded randomly in multiple frames; then they were extracted and processed according to the method in Section 3. A number of frames being used for embedding process are 20%, 50%, and ultimately 100% of the 100 test frames of the proposed standard videos. **Figure 19** shows the performance under these circumstances. It can be shown that the detection process was enhanced dramatically when comparing with **Figure 18**. The system can perform well even with only 20% of the frames being watermarked. To evaluate our algorithm under different compression ratios, the standard test videos were watermarked and compressed using HEVC with different QPs: 15, 20, and 25. The value of 20 is a typical value for compression. **Figure 20** shows the performance of our system under these compression values. As QP becomes larger than 25, the video qualities go through noticeable degradation in terms of resolutions; in fact, the system performs well for QP values of 20 or less, and as QP values reach 25, the detection process starts to lose its efficiency for some videos. This is due to the aggressive quantization process of the discrete cosine transform (DCT) coefficients in the HEVC process.

Our selective denoising filter which was introduced in Section 4 was tested for the watermarked videos. The standard videos were watermarked according to the proposed embedding process; then they were subjected to salt-and-pepper noise

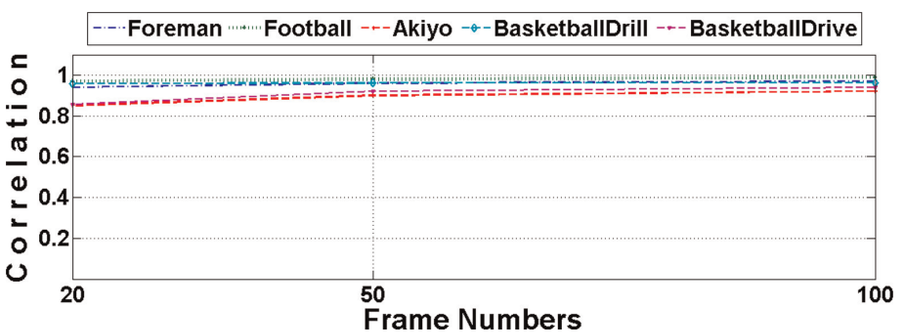

Figure 19.
Correlations vs. number of frames being watermarked when our detection method is used.

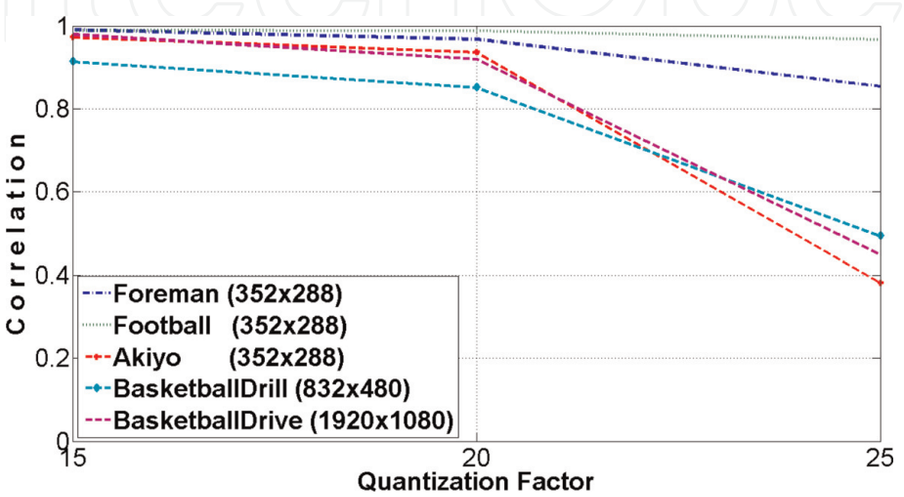

Figure 20.
Watermarking process performance under different quantization parameters.

with 1% density. At the receiving side, the videos were denoised using the selective filter; then the watermarks were extracted. **Figure 21** shows the normalized correlations and the PSNRs under these conditions. It is clear that the selective filtering scheme enhanced the visual appearance by eliminating the noise without significant effects on the efficiency of the watermarking process. Furthermore, **Figure 22** shows the results when 2D median filter is used. It can be seen that our selective denoising filter outperformed the 2D median filter in terms of the correlation

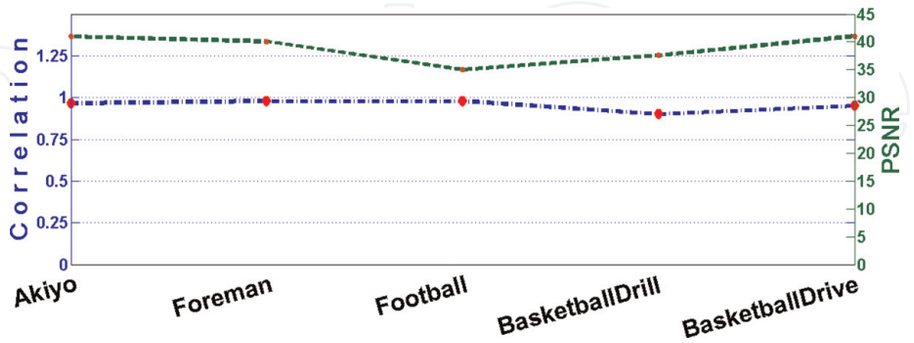

Figure 21.
The performance of the watermarking process with the use of the proposed selective filter where the blue line is the correlation and the green line is the PSNR.

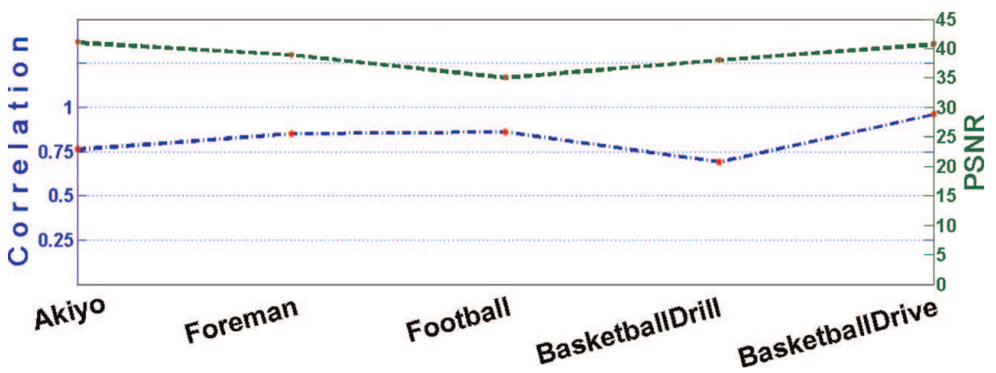

Figure 22.
The performance of the watermarking process with the use of a 2D median filter where the blue line is the correlation and the green line is the PSNR.

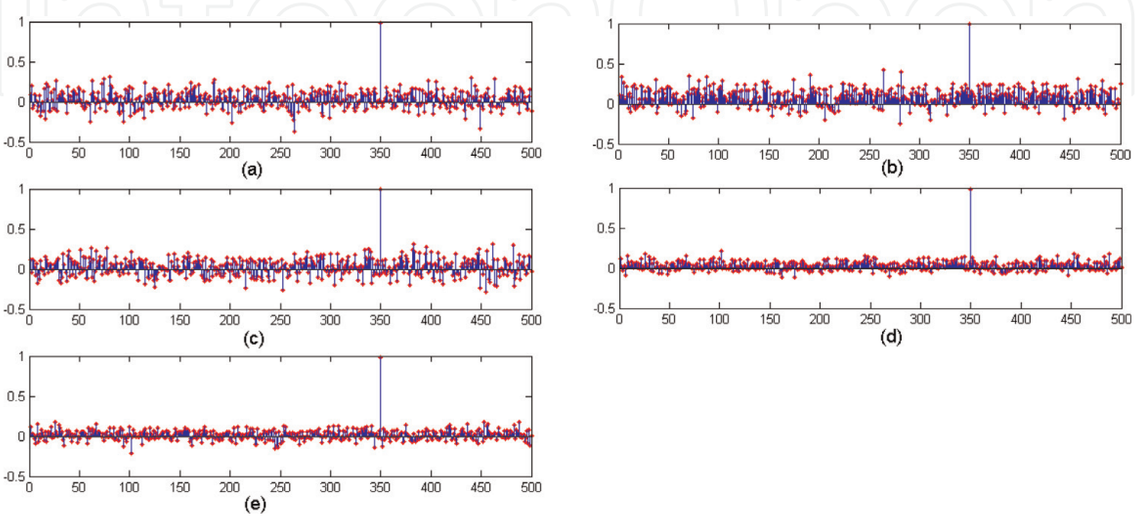

Figure 23.
The watermarking process response to false alarm test, the right watermark is the 350th with different videos: (a) Akiyo, (b) Foreman, (c) Football, (d) BasketballDrill, and (e) BasketballDrive.

values. The PSNR values are almost comparable, even though when assessing the quality of the images, the metric parameters are not the only factor that should be taken into consideration. The perceptual quality is another factor which was better with our own filter due to the selectivity process.

Since the security is an important aspect for our algorithm and any data hiding technique, false alarm attacks were studied for our test videos. That happens when no hiding process was done or a false watermark was hidden and still the system indicates the existence of our watermark. To do that, we generate 500 different random watermarks and hide them in the test videos according to our proposed algorithm, and the right watermark was one of them and it was set to be the 350th one. **Figure 23** shows the results. It can be seen that the response of our system was low to the false watermarks, and only the right watermark resulted in high response. This is an indication of good reliability of our system.

6. Conclusions and future work

This work proposes a DWT-based watermarking process using randomly generated orthonormal filter banks. An enhanced detection process was proposed to add to the robustness of the system. Moreover, a selective filtering process was developed to eliminate the noise. A good deal of the security of the system was achieved by the randomness in the filter banks, the pseudorandom sequence that was used to encode the watermark, and the regions of hidings. It was shown that the proposed technique performs well with and without HEVC. The compression ratio that was used is typical. Further investigation of the efficiency of the watermarking process under other aggressive attacks will be discussed and researched in future work. Moreover, an integration process of the data hiding process inside videos and the HEVC process will be studied and investigated.

Conflict of interest

The authors whose names are listed above certify that there are no conflicts of interests of any sort or nature between any of them and any institution or organization (public or private) that have special interest in the research that is the topic of this work.

Author details

Farhan Al-Enizi[1*] and Awad Al-Asmari[2]

1 PSAU University, Al-Kharj, Saudi Arabia

2 Shaqra University, Saudi Arabia

*Address all correspondence to: farhan414@gmail.com

References

[1] Langelaar GC, Setyawan I, Lagendijk RL. Watermarking digital image and video data. A state-of-the-art overview. IEEE Signal Processing Magazine. 2000; **17**(5)

[2] Sullivan GJ, Boyce JM, Chen Y, Ohm J-R, Segall CA, Vetro A. Standardized extensions of high efficiency video coding (HEVC). Selected Topics in Signal Processing, IEEE Journal of. 2013; **7**(6):1001-1016

[3] Sullivan GJ, Ohm J-R, Han W-J, Wiegand T. Overview of the high efficiency video coding (hevc) standard. IEEE Transactions on Circuits and Systems for Video Technology. 2012; **22**(12):1649-1668

[4] Panyavaraporn J. Multiple video watermarking algorithm based on wavelet transform. In: 2013 13th International Symposium on Communications and Information Technologies (ISCIT), IEEE. 2013. pp. 397-401

[5] Meerwald P, Uhl A. Survey of wavelet-domain watermarking algorithms. In: Photonics West 2001-Electronic Imaging. International Society for Optics and Photonics; 2001. pp. 505-516

[6] Lee M-S. Image compression and watermarking by wavelet localization. International Journal of Computer Mathematics. 2003;**80**(4):401-412

[7] Kundur D, Hatzinakos D. Digital watermarking using multiresolution wavelet decomposition. In: Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing; 1998; vol. 5; IEEE. 1998. pp. 2969-2972

[8] Guzmán VVH, Miyatake MN, Meana HMP. Analysis of a wavelet-based watermarking algorithm. In: 14th International Conference on Electronics, Communications and Computers, 2004. CONIELECOMP 2004; IEEE. 2004. pp. 283-287

[9] Wang S-H, Lin Y-P. Wavelet tree quantization for copyright protection watermarking. IEEE Transactions on Image Processing. 2004;**13**(2):154-165

[10] Bhattacharya S, Chattopadhyay T, Pal A. A survey on different video watermarking techniques and comparative analysis with reference to h. 264/avc. In: 2006 IEEE Tenth International Symposium on Consumer Electronics, 2006. ISCE'06. IEEE. 2006. pp. 1-6

[11] Martinez R, Reyes R, Cruz C, Nakano M, Perez H. A dwt-based video watermarking scheme resilient to mpeg-2 compression and collusion attacks. In: International Symposium on Information Theory and Its Applications 'ISITA 2008', IEEE. 2008. pp. 1-5

[12] Abdallah EE, Hamza AB, Bhattacharya P. Video watermarking using wavelet transform and tensor algebra. Signal, Image and Video Processing. 2010;**4**(2):233-245

[13] Vetterli M, Kovacevic J. Wavelets and Subband Coding; 1995

[14] Panyavaraporn J. Wavelet based video watermarking scheme for h. 264/avc. In: 2011 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS); IEEE. 2011. pp. 1-5

[15] Xu D, Wang R, Wang J. Video watermarking based on spatio-temporal jnd profile. In: International Workshop on Digital Watermarking; Springer. 2008. pp. 327-341

[16] Vinod P, Bora P. A new inter-frame collusion attack and a countermeasure.

In: International Workshop on Digital Watermarking; Springer. 2005. pp. 147-157

[17] Liu Y, Zhao J. A new video watermarking algorithm based on 1d dft and radon transform. Signal Processing. 2010;**90**(2):626-639

[18] Matsubara T, Moshnyaga VG, Hashimoto K. A low-complexity and low power median filter design. In: 2010 International Symposium on Intelligent Signal Processing and Communication Systems. 2010

[19] Deivalakshmi S, Sarath S, Palanisamy P. Detection and removal of salt and pepper noise in images by improved median filter. In: Recent Advances in Intelligent Computational Systems (RAICS), 2011 IEEE; IEEE. 2011. pp. 363-368

[20] Sundararajan M, Yamuna G. Dwt based scheme for video watermarking. In: 2013 International Conference on Communications and Signal Processing (ICCSP); IEEE. 2013. pp. 460-464