Multiobjective Cost Function Based Digital Video Watermarking Technique

Amir M. U. Wagdarikar
Department of Electronics
KL University,Vijaywada,India
amirmuwagdarikar@gmail.com

Ranjan K. Senapati
Department of Electronics
KL University, Vijaywada, India

Abstract: Video watermarking is a key process for copyright protection of digital data. In the video sequence, there is a huge quantity of data and these data are more susceptible to attacks. Hence, this paper intends to develop a novel video watermarking technique with the aid of multi-objective cost function. Initially, the keyframes are extracted from the input video sequence and these extracted video frames are subjected to wavelet transform for achieving the wavelet coefficients. Then, a multi-objective cost function is proposed with the help of multiple criteria like edge, brightness, the intensity of pixel, coverage, energy and frequency coefficient. The bit plane technique is utilized here to develop multiple binary images for the secret message by means of partitioning the secret messages. Further, with the aid of the multi-objective cost function, the message bit is embedded in the frame (wavelet coefficient). The embedded image is transmitted from the sender to the receiver via a communication channel and on the receiver side, the retrieval process takes place. The original image is retrieved in the receiver side with the help of a multi-objective cost function. Finally, a comparative analysis is accomplished between the proposed wavelet+Cost (wavelet transform based on a multi-objective cost function) and the existing models like wavelet model, Linear Significant Bit model (LSB) and LSB + cost model in terms of peak-to-signal noise ratio (PSNR) and correlation coefficients.

Keywords: Video Watermarking; Wavelet Transform; Multiobjective Cost Function; Bit Plane Technique;wavelet+Cost model.

Nomenclature

Abbreviations	Descriptions
DFT	Discrete Fourier Transform
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
HVS	Human Visual System
CS-SCHT	Conjugate Symmetric Sequence- Complex Hadamard Transform
LABS	Low Amplitude Block Selection
AB	Amplitude Boost
HEVC	High-Efficiency Video Coding
PSNR	Peak Signal to Noise Ratio
MOS	Mean Opinion Score
BER	Bit Error Rate
DIBR	Depth-Image-Based Rendering
DT-CWT	Dual-Tree Complex Wavelet Transform
BWT	Bi-orthogonal Wavelet Transform
SVD	Singular value decomposition
BELM	Bi-directional Extreme Learning Machine
CEN	Contrast Enhancement Neighbourhood
Multi-BAM-FUZ	Multiple watermarks based on Bi-directional Associative Memory Neural Networks and Fuzzy Inference System

1.Introduction

In recent days, the world records a tremendous growth in the field of multimedia application particularly in the distribution of multimedia content across the globe via the internet [23]. However, this growth imposes more challenges for the content owner in terms of efficient copyright protection for intellectual video content as well as security in different applications like video-conferencing, digital television, and

video-on-demand and so on [8] [9]. One optimal solution for copyright protection of video frames enclosing audio, text, static and moving images is digital video watermarking [10]. This digital video watermarking technology includes three steps like embedding, broadcasting, and retrieval. In the embedding process, the original video frame is marked with secret data with two crucial questions 'when' and 'where' to embed the watermark [11]. This embedding process can take place either after or before the compression. The watermarked video is transmitted using the distribution channel in the broadcasting stage and here the attacks try to gather or revoke the contents. The recovery of watermarked video frames is achieved in the retrieval stage [14]. Different digital image watermarking technique is being well-established in video frames as it encloses images, videos, and audio. copyright protection to the video frames by hiding the copyright information against unauthorized copy and distribution is quite a complex task. Apart from this, there are still some challenges in the selection of the position to insert watermark and limitation in bandwidth [12] [13] [7].

Different digital video watermarking algorithms are being developed on the basis of their working domain either spatial domain or frequency domain. The spatial domain watermarking techniques is the most-straightforward method as it makes use of the entire cover image for watermarking. In this method, the pixel values of the original image frames are modified and hence it becomes robust to attacks like cropping, noise, lossy compression [15] [16]. In contrast, these pixel-based methods are vulnerable to multiple frame collusion and de-synchronization attack [20]. In frequency domain techniques, the transform coefficients of the frame are modified using transforms like DFT, DCT, and DWT. Here, the original frame image is converted into the frequency domain. Then, the watermarking information is stored in the image by altering the transformed domain coefficients. Then, the watermarked video is obtained by applying the inverse transform [17] [18].

In DWT based video watermarking technique, the video frame is decomposed into sub-images as lower resolutions as well as in higher resolutions and the approximation of the sub-images resembles the original image [19]. The major drawback of DWT is a change in the quality of the image due to high-frequency components. The DCT uses global DCT with the intention of embedding the watermark in HVS. This model is much prone to lossy compression attacks [21] [22]. Apart from these drawbacks, there are few other technical challenges in terms of robustness and imperceptibility. Therefore, there is a necessity to develop an apt video watermarking technique that can achieve both robustness and imperceptibility.

The major contribution of this research work is Initially, the keyframes that undergo watermarking are extracted from the huge video database. The wavelet coefficient is achieved from the original frame by applying the wavelet transform. The cost function that plays a major role in the embedding and retrieval process is computed for the extracted key frames (cover image) using multi-objective like the intensity of pixel, energy, edge, coverage, frequency coefficient, and brightness. The bit plane technique helps in partitioning the secret message into multiple binary images. In the embedding process, the message bit is embedded on the wavelet coefficients on the basis of the multi-objective cost function. In the receiver side, the secret message is retrieved using the same multi-objective cost function. Finally, the proposed wavelet+Cost model is compared with the existing models like wavelet model, LSB model, LSB+Cost model in terms of PSNR and Correlation coefficient.

The remaining sections of this research work are organized as: Section 2 portrays the literature works undergone in the field of video watermarking. Section 3 depicts the architecture of the proposed multi-objective cost function based video watermarking technique. The processing steps of video marketing are depicted in Section 4 and the results obtained and their corresponding discussion is shown in Section 5. Section 6 provides a strong conclusion to this research work.

2. Literature Review

2.1 Related Works

In 2016, Meenakshi *et al.* [1] formulated blind video watermarking with the aid of CS-SCHT for HD video. Both LABS and AB was efficient in making the proposed video watermark as robust and imperceptible. The attacks like HEVC compression, Rescaling and Cropping were evaluated through simulations. The degree of imperceptibility of the proposed video watermarking technique was computed in terms of PSNR. Further, the proposed CS-SCHT based video watermarking was compared with the existing DFT in terms of PSNR, MOS, and BER.

In 2016, Asikuzzaman *et al.* [2] proffered a blind DIBR 3D video watermarking scheme with the intention of preserving the left, right and the central views of the video frame. The dual-tree complex wavelet transform was utilized for embedding watermark of both the center view chrominance channels

(U and V). The depth image-based rendering technique was employed with the aim of generating the depth map as well as the watermarked center view from the left and right views. Further, from the depth map, the virtual left, as well as the virtual right views, were generated.

In 2008, Coria *et al.* [3] developed a novel watermarking method for playback control that had the capability of eliminating the common geometric attacks like rotation, scaling, and cropping, accompanied during compression. Initially, to each individual frames of the video sequence, the watermark was inserted. Further, to each of the frames, a four-level DT-CWT was employed. The inverse DT-CWT performed the decoding process. Then, DT-CWT was utilized in order to neglect the shortcoming of the regular wavelets with the help of perfect reconstruction and better directional selectivity.

In 2018, Sake *et al.* [4] proposed BWT and SVD in video watermarking with the objective of preserving the copyright of images in the video frames. The watermark embedding process as well as the watermark extraction process was utilized for enhancing the efficiency of video watermarking. In the embedding process, the random frame was generated using the improved Artificial Bee Colony Algorithm (ABC). In the transformed 2D images, the motion vector was estimated using the motion estimation technique.

In 2019, Raj pal *et al.* [5] introduced an innovative watermarking scheme with the aid of BELM for Advanced Video Coding (MPEG-4 AVC) videos. On the basis of the threshold operation, the relevant frames to be watermarked were identified with the novel fuzzy-based frame selection algorithm. The enhancement in the security level of the video sequences was ensured by the proposed model. The transposition cipher encrypted the binary watermark in the video frames.

In 2016, Loganathan *et al.* [6] developed Multi-BAM-FUZ with the aim of designing a robust reversible adaptive video watermarking system. The frequency, as well as the spatial domain distributions in the Wavelet domain, was localized to inherently make use of the Wavelet domain. In DWT, the Fuzzy inference system fed the HVS characteristics as input to DWT in order to generate the adaptive embedding factor.

Table 1: Features and Challenges of State-of-art Video Watermarking Techniques

| Author [citation]	Methodology	Features	Challenges
Meenakshi et al.[1]	CS-SCHT	• Less computational cost • Less computational complexity	• Requires huge memory space • Prone to additive noise
Asikuzzaman et al.[2]	DIBR	• Secured against geometric attacks	• Tedious process • Loss of video information
Coria et al.[3]	DT CWT	• Good directional selectivity • Perfect reconstruction of video frames	• High cost • Not suitable for DVD players.
Sake et al.[4]	BWT-SVD	• Highly robust • Against Salt and Pepper noise attack, Gaussian attack, speckle attack, and Poisson attack	• Low PSNR • Low visual quality
Rajpal et al.[5]	BELM	• Suitable for real-time compressed videos • Good robustness	• High bit-error rate
Loganathan et al.[6]	Multi-BAM-FUZ	• Enhanced visual quality • High PSNR	• Low data payload • Low watermark • Capacity

2.2 Review

Table 1 represents the methodology, features, and challenges of the existing video watermarking techniques. All these challenges stimulate the researches to conduct further researches on video watermarking to formulate a good video watermarking technique. CS-SCHT in [1] has the advantages of low computational cost and low computational complexity. This technique suffers from the shortcomings of the huge requirement of memory space and highly prone to additive noise. DIBR is secured against the geometric attacks. But, this is a tedious process and information may get losses during compression [2]. Further, a better directional selectivity, as well as higher reconstruction rates, was achieved in DT-CWT [3]. This approach consumes high cost for implementation and it is not suitable to DVD players. BWT-SVD [4] is highly robust against the Salt and Pepper noise attack, Gaussian attack, speckle attack, and Poisson attack. Apart from this, the visual quality and PSNR of video are low. BELM technique utilized in [5] is sufficient to real-time compressed videos and has achieved high robustness. The major cons of this model are low BER. Multi-BAM-FUZ [6] has the capability of generating an enhanced video quality and high PSNR. The data payload and the capacity of watermarking are low.

3. Proposed Architecture of Video Watermarking Technique

3.1 Architectural Representation

The proposed video watermarking technique encloses five major phases (i) video acquisition phase (ii) key frame extraction phase (iii) evaluation of cost multi-objective function phase (iv) embedding phase and (v) retrieval phase within itself as per Fig. 1. Initially, the video acquisition is accomplished to obtain a particular video sequence from the huge database that undergoes watermarking. Then, from the chosen video sequence, the keyframes are exacted in an equal time interval. To each of the extracted key frames, the two-level decomposition of the wavelet transform is applied to achieve the wavelet coefficients. The benchmark for both the embedding and extraction phase is the cost function computation, which portrays the pixels for which the watermarking needs to be undergone. As a novelty, a multi-objective cost function is proposed with multiple criteria like the intensity of pixels, the energy of wavelet, edges, brightness and frequency coefficients. Further, with the aid of the evaluated multi-objective cost function, the watermarked frames are concealed into cover frames and the bit plane technique obtains the equivalent binary image of the secret image by portioning the secret image(video). On the basis of the cost function, the message bit is embedded into wavelet coefficients (frame) in the embedding phase. In the sender side, the embedded image is generated and it is transferred via the communication channel to reach the destination. On the receiver side, the retrieval operation is performed by the receiver to retrieve the message from the watermarked image. The retrieval operation is assisted by the original frame cost function.ame and the multi-objective cost function. Thus, at the end of the retrieval process, the original message is retrieved and robustness of the video marking technique is enhanced. The proposed model is simply referred to as a wavelet+Cost model.
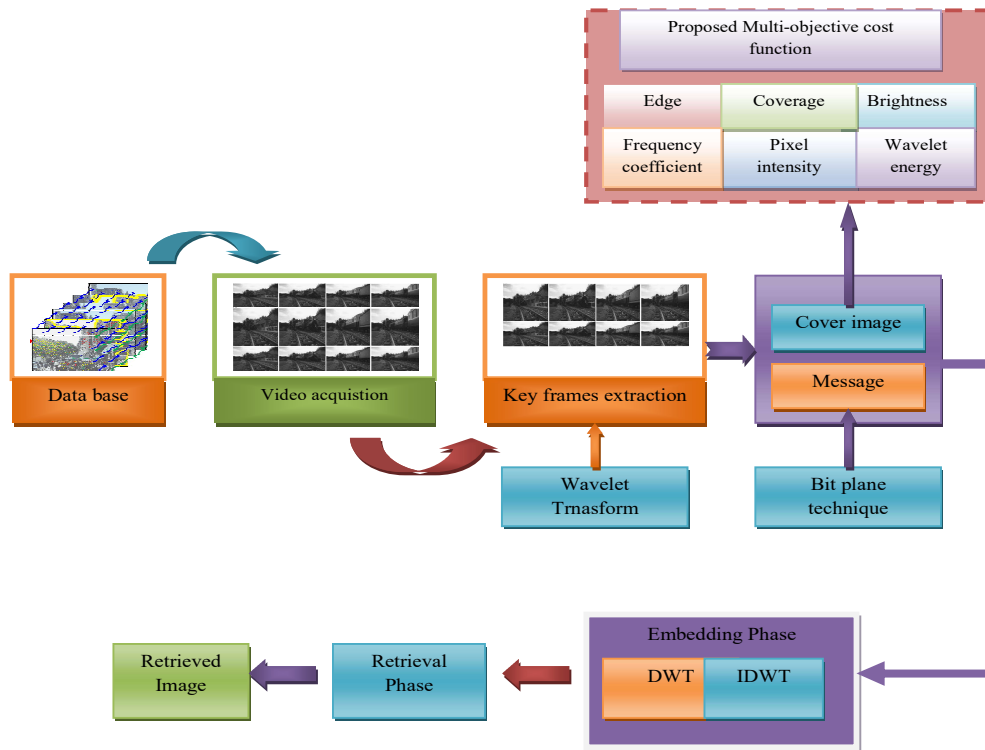
Fig. 1. *Proposed intelligence architecture of video watermarking technique*

4 Processed steps for Video Watermarking

4.1 Key Frames Extraction

The database holds the huge volume of the input video sequence, from which the sensitive frames for which the watermarking needs to be undergone, is selected. The key frame extraction plays a vital role in enhancing the robustness of the system by extracting the keyframes from massive video sequences. The

database with a vast amount of video (x) is represented using the term Di, where $Di = \{Vi_m : 1 \le m \le x\}$. The term Vi_m in the input database (Di) represents the mth video available within the database. Every video sequence encloses y count of videos frames (Fi), where $Vi = \{Fi_n : 1 \le n \le y\}$ and term n depicts the integral value of the equal interval keyframes. In the watermarking technique, the secret message (Mg), which is a grey scale image is a major concern and it is in the size kXl. For efficient watermarking on non-stationary real-time images, it is essential to choose the significant keyframes to aid the embedding as well as retrieval phase. Such that the equal interval keyframes are selected in the form y/n. The mathematical formula for the extraction of significant keyframes is shown in Eq. (1), in which the original frame utilized for embedding the secret image is represented as X. Once, the significant keyframes are extracted, the secret message in the frames need to be covered by identifying the suitable pixels with the help of cost function. The parameters a and b are the pixel.

$$Vi = \{X_b : 1 \le b \le a\} \tag{1}$$

4.2 Multi-Objective Cost Function

The cost function computes the cost value of each individual pixel in the original frame. The proposed multi-objective cost function computes the cost function with the help of multiple criteria like intensity, brightness, edge, frequency coefficients, and coverage. Fig. 2 illustrates the diagrammatic representation of the cost function evaluation.

The intensity of pixels: in the original frame, each and every individual pixel is deemed as a sample of a video frame that furnishes the original image accurately. The count of the pixels in the original frame ranges from 0 to 255 and each pixel holds an intensity value. This is one among the major criteria for cost function evaluation and this is assisted by neighbouring values of pixels in the current frame. In general, the size of the original frame is 288×352, the original frame is resized to 72×88 using the resizing operation and this resized image is represented as X_{re}. The mathematical formula for pixel intensity is expressed in Eq. (2), in which the current pixel value and k^{th} neighborhood pixel value are represented using the term $X_{re}(p,q)$ and $X_{re}^k(p,q)$, respectively. The value of k range from 1 to 8 for the eight band sub-pixels used here. The term p,q is the position of the pixel.

$$Im_{re}^{in}(p,q) = \frac{1}{255*8} \sum_{k=1}^{8} X_{re}(p,q) \quad X_{re}^k(p,q) \tag{2}$$

Edge detection: The Sobel edge detector [27] is employed for estimating the edge-based cost function. Typically, the edge pixels are obtained from the resized frame with the help of the Sobel edge detector. The vertical and the horizontal edges in the image are detected and the high spatial frequency regions corresponding to the edges in the image are emphasized by the Sobel operator using 2-D spatial gradient measurements. The blurred regions in the frame are mitigated and the sharp lines are perceived effectively. The magnitude, strength and the angle of orientation of edges of the resized frame are computed using the mathematical formulas showcased in Eq. (3), Eq. (4) and Eq. (5), respectively. The gradient components in the horizontal and vertical orientation are represented as Gi_x and Gi_y, respectively. The overall mathematical formula for edge detector is manifested in Eq. (6), in which the edge based video frame and input frame is represented as S_{edge} and X. The edge-based cost function for the video frame is determined by using Eq. (7) and here the parameters $S_{edge}^s(p,q)$ and $S_{edge}^z(p,q)$ denotes the edge-based cost value and neighbourhood edge pixel. The term $Sk_{edge}(p,q)$ tells that the pre-defined pixel is an edge.

$$|Gi| = \sqrt{Gi_x^2 + Gi_y^2} \tag{3}$$

$$|Gi| = |Gi_x| + |Gi_y| \tag{4}$$

$$\theta = \arctan\left(\frac{Gi_y}{Gi_x}\right) \quad \frac{3\pi}{4} \tag{5}$$

$$S_{edge} = edge(X_{re}) \tag{6}$$

$$S_{edge}^s(p,q) = \frac{1}{8} \sum_{k=1}^{8} Sk_{edge}(p,q) \tag{7}$$

Brightness: it is one of the major criteria for enhancing the visual perception of the frame. The brightness of the image in the frame is determined using the CEN technique. This CEN technique rejects

the uncertainty in the frames, thereby improves the significant features of the frame. Typically, in the resized image X_{re}, CEN modifies the intensity of the pixels and provides a massive count of bins. Further, with the aid of local maxima as well as the local minima of the pixel, the contrast enhancement is achieved. The mathematical formula for computing the contrast enhancement is shown in Eq. (8).The contrast enhancement frame image is denoted using the term $X_{re}^{N}(p,q)$ and the maximum gray scale value is represented as Ti_g (here, $Ti_g = 255$). The minima and the maxima are depicted as Ti_1 and Ti_2, respectively. Further, the obtained contrast-enhanced image $Im^B(p,q)$ is the size 72×88 and the mathematical formula for the brightness of the resized image is shown in Eq. (9).

$$X_r^N(p,q) = \frac{X_r(p,q) \quad Ti_1}{Ti_2 \quad Ti_1} \times Ti_g \tag{8}$$

$$Im^B(p,q) = \frac{1}{255*8} \sum_{k=1}^{B} X_{re}^N(p,q) - X_{re}^N(p,q) \tag{9}$$

Wavelet energy: The energy of the pixel is computed here by the conventional wavelet transform [28], which extracts the wavelet coefficients (Wa) in terms of position and scale features. Typically, the original image frame is decomposed by the wavelet transform into four sub-bands namely LL, LH, HH, and HL. The mathematical formula for computing the cost function of the wavelet transform is represented as per Eq. (10). This equal decomposes the frame image into horizontal, diagonal, approximation and vertical. The approximation band is LL band and it is equivalent to the original frame image. Further, to achieve the 72×88 sized LL band, the second-level decomposition is undergone. Then, the achieved 72×88 size LL band is portioned into c blocks and the energy is computed. The evaluated energy value is allocated to their corresponding pixel values. The mathematical formula for block energy of the wavelet coefficient is exhibited in Eq. (11), in which the probability measure and the total count of pixels in each of the block is represented using the term Pb_p and N, respectively. In Eq. (12), the normalized factor that aid in evaluating the cost function c from the range 0 to 1 is denoted as Z_n while calculating the overall pixels energy using wavelet coefficient $(Im^w(p,q))$.

$$Wa = DWT(X) \tag{10}$$

$$Wa_B(c) = \sum_{p=1}^{N} \left[Pb_p \log(pb_p) + 1 \right] \tag{11}$$

$$Im^w(p,q) = \frac{1}{Z_n} [Wa_B(c)] \; ; c \in (p,q) \tag{12}$$

Coverage: The neighbouring pixels play a major role in computing the coverage value of the wavelet coefficients. The coverage of a pixel is the ratio of the count of similar k neighborhood pixel denoted as NP_k to the total count of pixels available in the frame TC_d. The mathematical formula to determine the coverage cost of the wavelet coefficient is specified in Eq. (13).

$$Im^d(p,q) = NP_k / TC_d \tag{13}$$

Frequency coefficients: The frequency coefficients are computed from the original video frames by taking the magnitude of the coefficients. The DFT here performs the transformation of the continuous time function into its corresponding frequency elements. The frequency coefficients provide a high scale of robustness against certain attacks like rotation, scaling, cropping, etc. Typically, the energy of the image is mostly concentrated in LL region (i.e., lower frequency coefficient) and when the watermarking is carried out in this region, there is a significant degradation in the quality of the image, which can be revoked. On the other hand, the higher frequency coefficient sets (i.e. HH) encloses the image edges as well as texture. In most cases, the watermarking is accomplished in the middle frequency coefficient (LH and HL) in order to achieve better robustness in watermarking. But, here the noise is more. The mathematical formula for frequency coefficient extraction is shown in Eq. (14) and Eq. (15). Most of the technique makes use of HH and LL for water making.

$$Im^{FC} = DFT(X) \tag{14}$$

$$DFT(X) = \sum_{n=o}^{N\ 1} X_{re} e^{\frac{i2\Pi N}{N}} \tag{15}$$

The multi-objective cost function (Cost) of the frame is computed as per Eq. (16) in terms of intensity, the energy of wavelet, frequency coefficient, coverage, and edge. The most objective cost function is the key for embedding and retrieval process. The secrete message will we well-established in the frame when the cost value of the pixel is 1 and hence the size of the cost matrix is denoted as 72×88.

$$\text{Cost} = \frac{\left(1 \quad \text{Im}^s\right) + \left(1 \quad \text{E}^s\right) + 1 \quad \text{Im}^B + \text{Im}^w + \text{Im}^d + \text{Im}^{FC}}{4} \qquad (16)$$
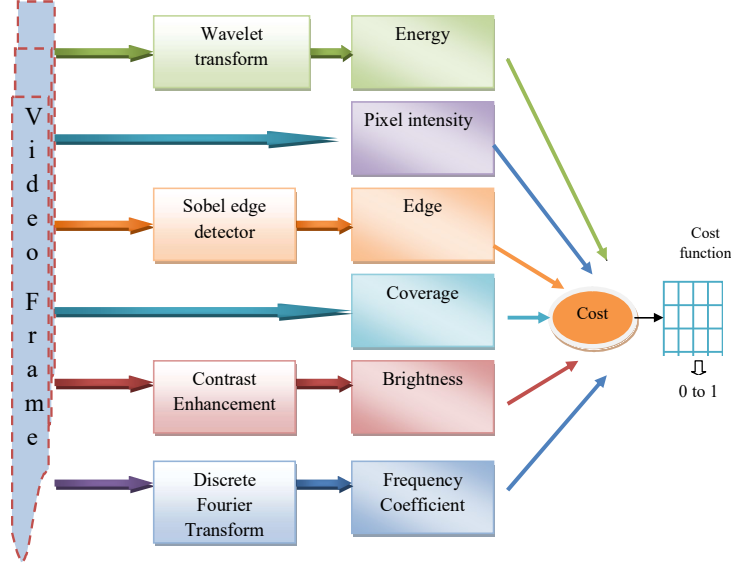
Fig. 2. *Cost function computation*

4.3 Bit Plane Technique

The message partition is accomplished using Bit plane technique. In general, the bit plane technique splits the original frame into smaller eight-bit planes and each of the bit planes here can belong to any of the eight different bit positions from MSB to LSB [25]. The secret message is transmitted from the source to the destination via the internet. The secret message Mg takes the size as 36×44. The count of secret message to cover the keyframes in the embedding process is a. The image frame has 8-bit grey scale value ranging from0 to 255. The bit plain technique has the advantage of granting noteworthy relation to each individual pixel in the frame. The bit plane technique utilized here acquires 8 binary images for the secret image and for every pixel, LSB generates the first-bit plane image. The last bit plane image is generated by all MSB in the pixel. The eight binary secrete images can be given in the form as specified in Eq. (17), in which the bit plane image and the last bit plane image is denoted as Mg_1 and Mg_8, respectively.

$$\text{Mg} = \left\{\text{Mg}_1, \text{Mg}_{2,\cdots,} \text{Mg}_8\right\} \qquad (17)$$

4.4 Embedding Process

Once the secrete message is generated based on the cost function, they are embedded into the frame by exploiting the wavelet transform. The two-level decomposition is applied to the input image in order to decompose the image into eight sub-bands. Thus, the secrete massages get well- established into the wavelet coefficients. The binary cost value is evaluated to find the position in which the message needs to be embedded. Typically, the cost function varies from 0 to 1 and here the size of the cost function is 72×88, which is equivalent to cover image. A huge count of cost values are given as the binary cot value and these cost values are identical to the size of the secrete message Mg. Thus, the size of the binary cot matrix becomes 36×44. Once, the secrete message and the binary cost value of the image is evaluated, they are embedded in the input image frame (cover image). When the cost value of the image is 1, the embedding of the message bit into the corresponding pixel takes place. The size of the cover image (colour image) is 288×352, which is converted from colour image to RGB using YUV spacing technique. The YUV image comprises of 3 components namely one luminance component (Y equivalent to grey scale) and two chrominance component (U and V as blue and red projection, respectively). In this proposed model, the one luminance component i.e. Y is considered as input.

Further, the wavelet component is exploited for extracting the wavelet coefficients used in both the embedded as well as the retrieval phase. The major reason behind the preference of wavelet transform is its faster computing behaviour and finer discrimination of fine details from the image. Initially, LL, HH, HL and LH components are extracted from the image by exploiting one-level decomposition. The size of each of the one-level decomposed sub-band is 144×176 and it can be mathematically given as per Eq. (18). The parameter X is the cover image and wavelet coefficients in the sub-bands are LL, HH, HL, LH. Further, 16 sub bands can be achieved by two-level decomposition of the bands. The two-level decomposed bands are mathematically given as per Eq. (19). After the second-level decomposition, the size of each sub-band becomes 72×88. Here, in this research work, only wavelet coefficients LL and HH are considered as the cover image for the purpose of embedding as well as retrieval. The representation of the coverage image for the proposed video marking technique is shown in Eq. (20). The overall mathematical formula for the embedding process is exhibited in Eq. (21). The count of wavelet bands is i, where $1 \le i \le 8$ and the count of message bit plane is j, where $1 \le j \le 8$. The watermarked image and the cover image is denoted using the term $EM_i(p,q)$ and $EM_i(p,q)$, respectively. Then, j^{th} bit plane message and strength of embedding is depicted using the term $Mb_j(p,q)$ and s, respectively. Fig. 3 illustrates the block diagram of the embedding process. Further, the embedded image gets transferred via the communication channel and reaches the destination. The embedded water marked image is subjected to two-level decomposition via inverse wavelet transform (inverse discrete wavelet transform (IDWT)) in order to obtain the watermarked image. Initially, the first-level decomposition is applied to the wavelet embedded images with the help of an inverse wavelet transform to form four sub-bands as per Eq. (22). Then, the two-level decomposition of the image is accomplished by using Eq. (23). The inverse wavelet transform of the embedded image and the watermarked image is represented using the term $Wt(p,q)$ and $\hat{Wt}(p,q)$, respectively in Eq. (24).

$$\{LL, LH, HH, HL\} = DWT(X) \tag{18}$$

$$\{LL_1, LH_1, HH_1, HL_1\} = DWT(LL) \quad \{LL_2, HH_2, HL_2, LH_2\} = DWT(LH)$$
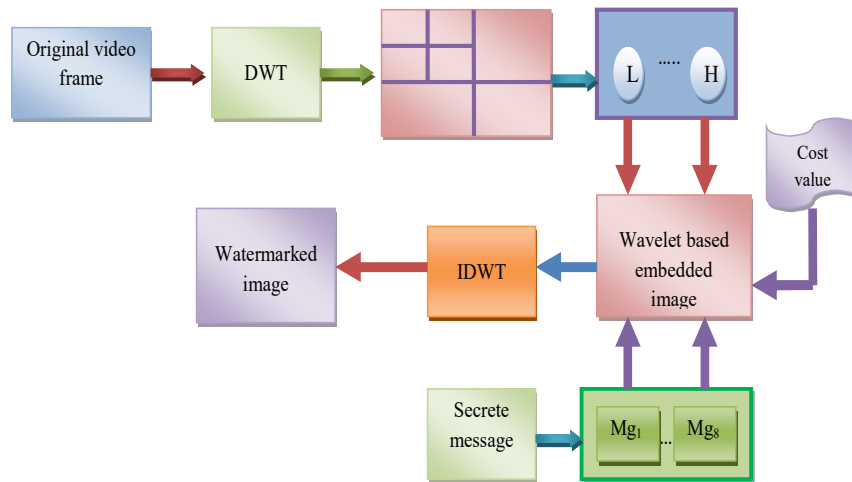$$\{LL_3, HH_3, LH_3, HL_3\} = DWT(HL) \quad \{LL_4, HH_4, HL_4, LH_4\} = DWT(HH) \tag{19}$$

$$Im = LL_1, HH_1, \cdots LL_4, HH_4 \tag{20}$$

$$EM_i(p,q) = Im_i(p,q) + s \quad Mb_j(p,q) \; ; \; if \; Cost_i(p,q) = 1 \tag{21}$$

$$EM_i(p,q) = LL_1^e, HH_1^c, \cdots, LL_4^e, HH_4^c \tag{22}$$

$$IDWT \{EM_i(p,q)\} \Rightarrow Wt(p,q) = \{LL_1^e, HH_1^c, \cdots, LL_4^e, HH_4^c\} \tag{23}$$

$$IDWT \{Wt(p,q)\} \Rightarrow \hat{Wt}(p,q) \tag{24}$$

Fig. 3. Embedding process

4.5 Secret Message Retrieval

The embedded watermarked image is transmitted to the receiver and the retrieval (extraction) process take place in the receiver side. The watermarked image, cost function, and the original frame image are the basis of the extraction process. The block diagram of the extraction process is manifested in Fig. 4. In

the receiver side, the keyframes are extracted from the received embedded video. The wavelet coefficients in the video frame are acquired here by exploiting the wavelet transform. The retrieval of the image takes places on the basis of the cost function. If the value of the cost function is 1, then the receiver extracts the message from the equivalent pixel. To the wavelet coefficients in the receiver side, the first-level decomposition is performed and four sub-bands are obtained as per Eq. (25). Then, to each of the sub-band, the second-level decomposition is performed as per Eq. (26). The embedded vide frame has the wavelet sub-bands Im* in Eq. (27) are used to retrieve the hidden message from the secrete massage on the basis of cost value.

$$\{LL, LH, HH, HL\} = DWT(\hat{Wt}) \tag{25}$$

$$\{LL_1^e, LH_1^e, HH_1^e, HL_1^e\} = DWT(LL*) \quad \{LL_2^e, HH_2^e, HL_2^e, LH_2^e\} = DWT(LH*)$$

$$\{LL_3^e, HH_3^e, LH_3^e, HL_3^e\} = DWT(HL*) \quad \{LL_4^e, HH_4^e, HL_4^e, LH_4^e\} = DWT(HH*) \tag{26}$$

$$Im* = LL_1^c, HH_1^c, \cdots LL_4^c, HH_4^c \tag{27}$$

In the retrieval phase, the retrieval of the extracted embedded image is accomplished by using Eq. (28). The retrieved message $Re(p, q)$ is of the size 36×44. The proposed video watermarking technique has enhanced the quality of the frame and has protected the video frame

$$Re(p, q) = Im_i^*(p, q) - Im_i(p, q) ; \qquad if \quad Cost_i(p, q) = 1 \tag{28}$$
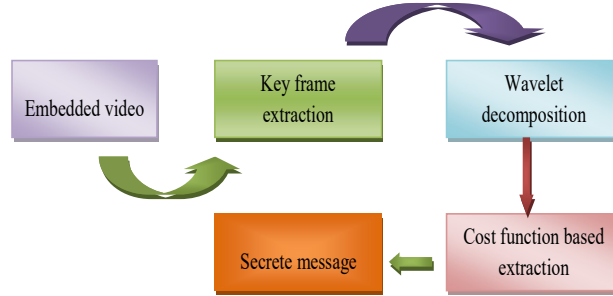
Fig. 4. *Diagrammatic representation of the retrieval process*

5 Result and Discussion

5.1 Experimental Setup

The proposed automatic video watermarking model was carried out in MATLAB, and the results related to the corresponding simulation were observed. In the current research work, the dataset is acquired from public available sources like YouTube at https://media.xiph.org/video/derf/ [access date: 2019-04-26] [26]Then, from, YouTube two input datasets namely video 1 and video 2 are gathered and the secret message is embedded into the video sequence in a secure manner. The proposed model is simply referred as the wavelet+Cost model is compared with the traditional techniques like wavelet-based video watermarking, LSB based video watermarking, and LSB + Cost based video watermarking in terms of PSNR and correlation coefficient. The mathematical formula for PSNR and correlation coefficient is showcased in Eq. (29) and Eq. (30).The height and the width of the coverage image are represented as A and B, respectively. The original, embedded and the maximum pixel value is denoted using the term $Im(p, q)$, $E(p, q)$ and Im_g (Im_g =255), respectively. SD_i And SD_y are the standard deviation and the sample cover is denoted as SD_{ij}.

$$PSNR = 20 \log_{10} \frac{Im_g \times A \times B}{\sum \sum (Im(p, q) \quad E(p, q))^2} \tag{29}$$

$$r_{pq} = \frac{SD_{ij}}{SD_i SD_y} \tag{30}$$

Fig.5 represents the overall results acquired with video watermarking. The cover image of the video frame and the secret message to be transmitted over the communication channel are shown in Fig.5 (a) and Fig.5 (b), respectively. When cost function $Cost = 1$, the secret message gets embedded into the cover image pixel and the watermarked image is exhibited in Fig. 5(c). On the receiver side, during the

retrieval phase, the original message is extracted only when $Cost = 1$ and the retrieved message is deliberated in Fig.5 (d).

(a)

(b)

(c)

(d)
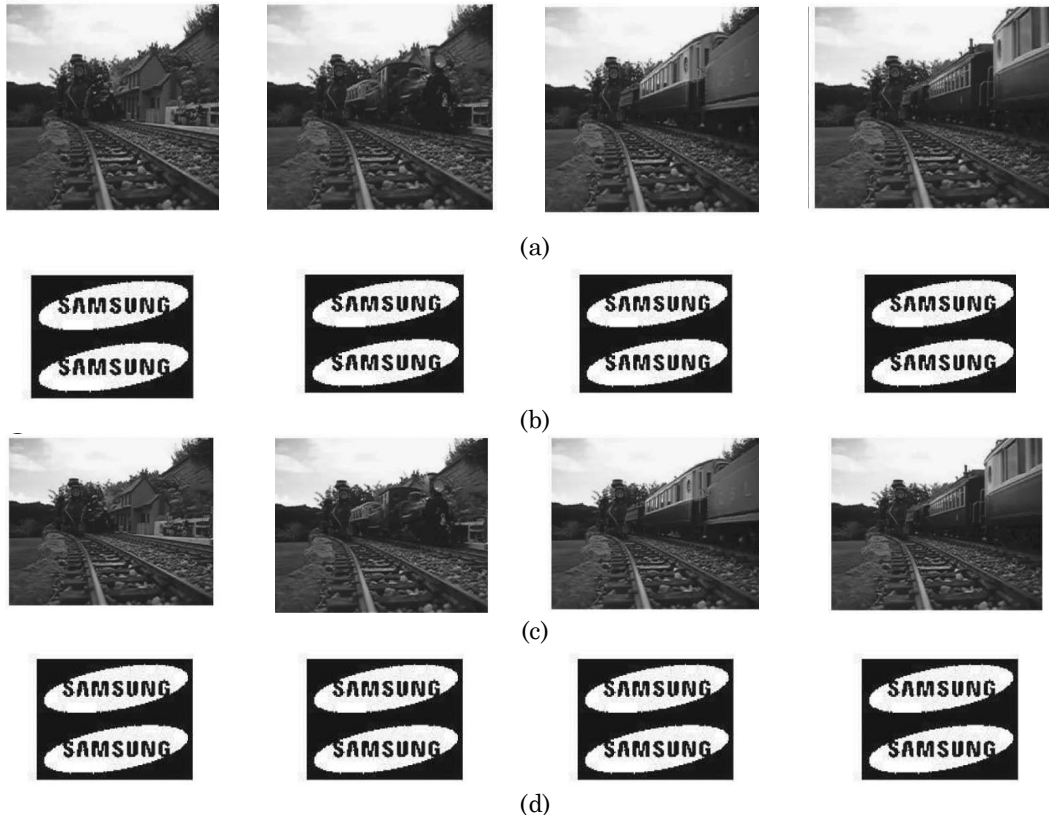
Fig. 5. *Diagrammatic representation of the resultant of the proposed model (a) original frame (b)secret message (c) watermarked embedded image (d) retrieved secret image*

5.2 Performance Analysis of Correlation and PSNR

The performance of the proposed WT +Cost model is evaluated in terms of the count of frames. For the acquired input video (video 1 and video 2), the performance parameters like PSNR and correlation coefficients are computed. Fig. 6 shows the overall performance analysis of video 1 and video 2 in terms of PSNR and correlation. Fig.6 (a) deliberates the correlation analysis and here the correlation is measured by varying the count of frames. When the count video frames are 4(F4) in video 1, the correlation of the proposed wavelet+Cost model is 20% better than wavelet model and 50 % and 60% better than LSB model and LSB+Cost model, respectively. In Fig.6(b) for video 2, the proposed wavelet+Cost model at F8 (count of frame=8) is 20%, 40% and 4% superior to the existing model like wavelet model, LSB model, and LSB+Cost, respectively in terms of correlation. Fig. 6(c) represents the PSNR analysis for video 1 and here the evaluation is carried out with 6 frames (F6) and the results exhibit a superiority of 7.14% by wavelet model, 14.2% by LSB model and 28.5% by LSB+Cost model over the proposed wavelet+Cost model. Then, for the video 2, the proposed WT+Cost model at F8 is 35%, 40% and 50% better than the traditional model like wavelet model, LSB model and LSB+Cost model, respectively in terms of PSNR. From the analysis, it is vivid that the correlation, as well as PSNR of the proposed model, is high and hence the quality of the embedded image is high.

(a)   (b)

(c)   (d)
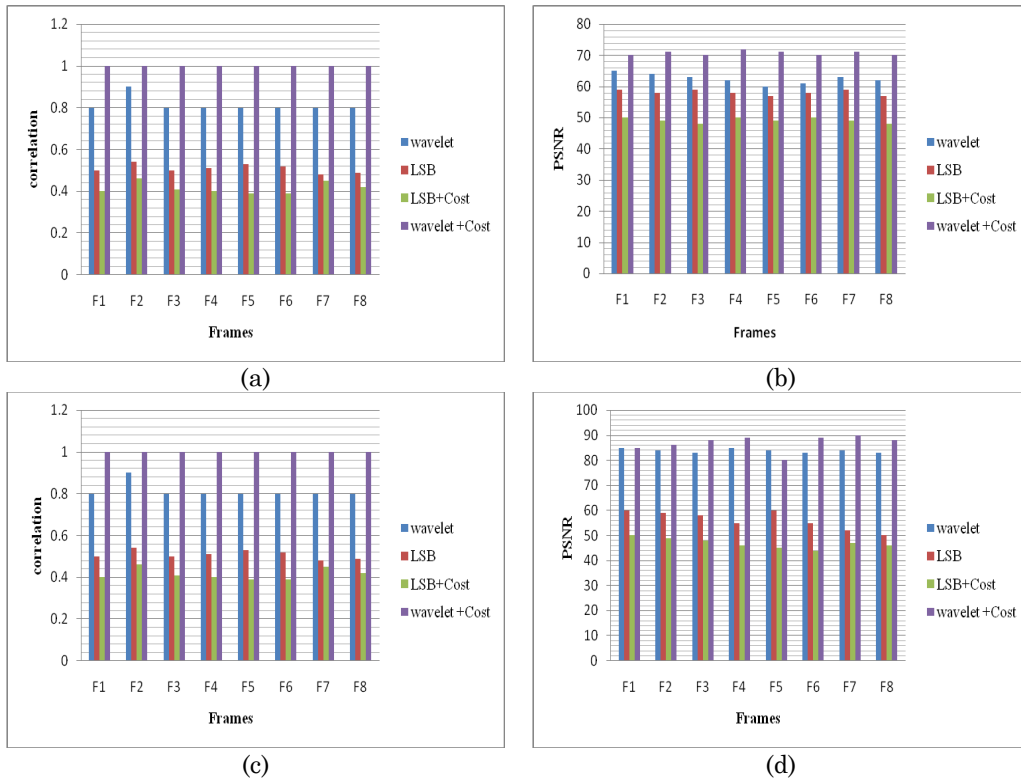
Fig. 6. *Performance analysis of the proposed wavelet + cost model over the existing wavelet, LSB and LSB+cost model for video frame 1 focusing on (a) correlation coefficient (b) PSNR and for video 2 focusing on (c) correlation coefficient (d) PSNR*

5.3 Robustness Analysis based Compression, Cropping and Frame Drop

Table 2 represents the performance analysis of the proposed model over the existing model in terms of compression, frame drop as well as cropping for video 1. In terms of compression, the proposed wavelet +Cost model is 16.03%, 5.15%, 4.97% better than the existing models like wavelet model, LSB model, and LSB+Cost model, respectively. In terms of 5% frame drop, the proposed wavelet +Cost model is 18.75%, 6.25% and 4.14% superior to the state-of-art model like wavelet model, LSB model and LSB+Cost model, respectively. An enhancement of 18%, 6.3% and 2.12% is recorded by the proposed wavelet +Cost model over the existing model like wavelet model, LSB model and LSB+Cost model, respectively in terms of cropping at 10×10.

Table 3 depicts the performance evaluation of the proposed model to the existing model for video 2 in terms of compression, frame drop as well as cropping. The proposed wavelet +Cost model exhibits enhancement of 22.6% by wavelet model, 1.03% by LSB model and 2.03% by LSB+Cost model in terms of compression. At the frame drop rate of 10%, the proposed wavelet +Cost model is 20.8% better than wavelet model, 6.25% better than LSB model and 4.16% better than LSB+Cost model. The cropping at 15×15 in the proposed wavelet +Cost model shows the superiority of 21.7%, 5.4% and 4.3% over the existing models like wavelet model, LSB model, and LSB+Cost model, respectively. This clearly reveals that the proposed model is superior to the existing model in terms of compression, frame drop as well as cropping.

Table 4 depicts the performance analysis of the proposed model over the existing model in terms of Scaling and cropping, without any distortions as well as Rotation and H.264 compression for suzie video collected from [25]. The Scaling and cropping at15% for the proposed wavelet +Cost model is 20.4%, 72%, and 1.07% better than the existing models like DT-CWT, DTW1 and DTW2, respectively. In terms of Rotation at 9°, the proposed wavelet +Cost model is 23% better than DT-CWT, 60% better than DTW1 and 56% better than DTW2 model. Thus, from the analysis, it is clear that the proposed wavelet +Cost model is superior to DT-CWT, DTW1 and DTW2 in terms of Scaling and cropping as well as rotation.

Table 2: *Performance Analysis of the proposed Wavelet +Cost model to the existing model for Video 1*

Metrics	LSB LSB + cost	Wavelet	Wavelet +Cost
Cropping			
10 × 10	0.8880 0.9275	0.7634	0.9423
15 × 15	0.8493 0.8718	0.7523	0.9366
20 × 20	0.8074 0.8100	0.7021	0.8825
Compression	0.9249 0.9275	0.8196	0.9761
Frame drop			
5%	0.9079 0.9275	0.7833	0.9622
10%	0.8693 0.8917	0.7722	0.9565
15%	0.7875 0.7901	0.6822	0.8626

Table 3: *Performance Analysis of the proposed Wavelet +Cost model to the existing model for Video 2*

Metrics	LSB	LSB + cost	Wavelet	Wavelet +Cost
Compression	0.9698	0.9594	0.7586	0.9787
Cropping				
10 × 10	0.8937	0.9127	0.7508	0.9496
15 × 15	0.8713	0.8898	0.7296	0.9260
20 × 20	0.8457	0.8637	0.6836	0.8987
Frame drop				
5%	0.9096	0.9290	0.7677	0.9666
10%	0.8873	0.9062	0.7464	0.9430
15%	0.8298	0.8474	0.6668	0.8817

Table 4: *Performance Analysis of the proposed Wavelet +Cost model to the existing model for Suzie Video*

	DT-CWT[24]	DTW1[24]	DTW2[24]	Wavelet +Cost
H.264 compression	1	1	1	1
Rotation				
3°	1	1	0.68	1
6°	0.86	0.30	0.64	0.88
9°	0.50	0.26	0.28	0.65
Rotation				
Without any distortions	1	1	1	1
Scaling and cropping				
5%	1	0.4	1	1
10%	0.98	0.32	1	1
15%	0.74	0.26	0.92	0.93

6 Conclusion

This paper focused on design as well as the development of a multi-objective cost function based video watermarking technique. Initially, the keyframes that undergo watermarking were extracted from the database. The wavelet transform was applied to the keyframes to achieve a wavelet coefficient. The multi-objective cost function was proposed for embedding and retrieval process on the basis of multiple criteria like pixel, energy, edge, coverage, frequency coefficient and brightness. Then, the bit plane technique was employed to partition the secret message into multiple binary images. In the embedding process, the message bit was embedded on the wavelet coefficients on the basis of the developed multi-objective cost function. This embedded watermarked message was transmitted from the sender to the receiver via a suitable communication channel and on the receiver side, the retrieval process takes place using the same multi-objective cost function. Finally, the proposed wavelet+Cost model was compared with the existing models like wavelet model, LSB model, LSB+Cost model in terms of PSNR and Correlation coefficient. At the random noise level of $9× 10^{-3}$, the proposed wavelet+Cost model is evaluated to be 59.5%, 54.5%, and 49.4% better than the traditional models like wavelet model, LSB model and LSB+Cost model, respectively for video 1. The proposed wavelet+Cost model is 55% superior to the wavelet model, 40% superior to LSB model and 42% superior to LSB+Cost model at histogram equalization 254 for video 2.

Compliance with Ethical Standards

Conflicts of interest: Authors declared that they have no conflict of interest.

Human participants: The conducted research follows the ethical standards and the authors ensured that they have not conducted any studies with human participants or animals.

References

[1] K. Meenakshi, K. Satya Prasad and C. Srinivasa Rao, "Development of Low-Complexity Video Watermarking With Conjugate Symmetric Sequency–Complex Hadamard Transform," in IEEE Communications Letters, vol. 21, no. 8, pp. 1779-1782, Aug. 2017.

[2]]M. Asikuzzaman, M. J. Alam, A. J. Lambert and M. R. Pickering, "Robust DT CWT-Based DIBR 3D Video Watermarking Using Chrominance Embedding," in IEEE Transactions on Multimedia, vol. 18, no. 9, pp. 1733-1748, Sept. 2016.

[3] L. E. Coria, M. R. Pickering, P. Nasiopoulos and R. K. Ward, "A Video Watermarking Scheme Based on the Dual-Tree Complex Wavelet Transform," in IEEE Transactions on Information Forensics and Security, vol. 3, no. 3, pp. 466-474, Sept. 2008.

[4] Anjaneyulu Sake, Ramashri Tirumala,"Bi-orthogonal Wavelet Transform Based Video Watermarking Using Optimization Techniques",Materials Today,vol.5, no.1,pp.1470-1477,2018.

[5] Ankit Rajpal, Anurag Mishra, Rajni Bala,"A Novel fuzzy frame selection based watermarking scheme for MPEG-4 videos using Bi-directional extreme learning machine",Applied Soft Computing, Vol.74,pp.603-620,January 2019.

[6] Agilandeeswari Loganathan, Ganesan Kaliyaperumal,"An adaptive HVS based video watermarking scheme for multiple watermarks using BAM neural networks and fuzzy inference system",Expert Systems with Applications, vol.63, pp.412. -434,November 2016.

[7] Antonio Cedillo-Hernandez, Manuel Cedillo-Hernandez, Mireya Garcia-Vazquez, Mariko Nakano-Miyatake, Alejandro Ramirez-Acosta,"Transcoding resilient video watermarking scheme based on spatio-temporal HVS and DCT",Signal Processing, vol.97, pp. 40-54,April 2014.

[8] Minghua Chen, Yun He and R. L. Lagendijk, "A fragile watermark error detection scheme for wireless video communications," in IEEE Transactions on Multimedia, vol. 7, no. 2, pp. 201-211, April 2005.

[9] H. Tian, Y. Xiao, G. Cao, J. Ding and B. Ou, "Robust watermarking of mobile video resistant against barrel distortion," in China Communications, vol. 13, no. 9, pp. 131-138, Sept. 2016.

[10] S. Lian, Z. Liu, Z. Ren and H. Wang, "Commutative Encryption and Watermarking in Video Compression," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 6, pp. 774-778, June 2007.

[11] A. Koz and A. A. Alatan, "Oblivious Spatio-Temporal Watermarking of Digital Video by Exploiting the Human Visual System," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, no. 3, pp. 326-337, March 2008.

[12] S. Chen and H. Leung, "Chaotic Watermarking for Video Authentication in Surveillance Applications," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, no. 5, pp. 704-709, May 2008.

[13]]M. Fallahpour, S. Shirmohammadi, M. Semsarzadeh and J. Zhao, "Tampering Detection in Compressed Digital Video Using Watermarking," in IEEE Transactions on Instrumentation and Measurement, vol. 63, no. 5, pp. 1057-1072, May 2014.

[14] J. Zhang, A. T. S. Ho, G. Qiu and P. Marziliano, "Robust Video Watermarking of H.264/AVC," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 54, no. 2, pp. 205-209, Feb. 2007.

[15] M. Noorkami and R. M. Mersereau, "Digital Video Watermarking in P-Frames With Controlled Video Bit-Rate Increase," in IEEE Transactions on Information Forensics and Security, vol. 3, no. 3, pp. 441-455, Sept. 2008.

[16] M. Barni, F. Bartolini and N. Checcacci, "Watermarking of MPEG-4 video objects," in IEEE Transactions on Multimedia, vol. 7, no. 1, pp. 23-32, Feb. 2005.

[17] H. Mareen, J. De Praeter, G. Van Wallendael and P. Lambert, "A Novel Video Watermarking Approach Based on Implicit Distortions," in IEEE Transactions on Consumer Electronics, vol. 64, no. 3, pp. 250-258, Aug. 2018.

[18] H. Mareen, J. De Praeter, G. Van Wallendael and P. Lambert, "A Scalable Architecture for Uncompressed-Domain Watermarked Videos," in IEEE Transactions on Information Forensics and Security, vol. 14, no. 6, pp. 1432-1444, June 2019.

[19] Wenwu Zhu, Zixiang Xiong and Ya-Qin Zhang, "Multiresolution watermarking for images and video," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 9, no. 4, pp. 545-550, June 1999.

[20] A. Mansouri, A. M. Aznaveh, F. Torkamani-Azar and F. Kurugollu, "A Low Complexity Video Watermarking in H.264 Compressed Domain," in IEEE Transactions on Information Forensics and Security, vol. 5, no. 4, pp. 649-657, Dec. 2010.

[21] K. Su, D. Kundur and D. Hatzinakos, "Statistical invisibility for collusion-resistant digital video watermarking," in IEEE Transactions on Multimedia, vol. 7, no. 1, pp. 43-51, Feb. 2005.

[22] A. M. Alattar, E. T. Lin and M. U. Celik, "Digital watermarking of low bit-rate advanced simple profile MPEG-4 compressed video," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 8, pp. 787-800, Aug. 2003.

[23] Yulin Wang and A. Pearmain, "Blind MPEG-2 video watermarking robust against geometric attacks: a set of approaches in DCT domain," in IEEE Transactions on Image Processing, vol. 15, no. 6, pp. 1536-1543, June 2006.

[24] L. E. Coria, M. R. Pickering, P. Nasiopoulos and R. K. Ward, A video watermarking scheme based on the dual-tree complex wavelet transform, IEEE Trans. Inf. Forensics Security vol.3, pp.466–474,2008.

[25] Xiph.org Video Test Media [derf's collection]. Available at: https://media.xiph.org/video/derf/, Accessed on 3 September, 2016.

[26] P. Patiala and C. Biradar, Robust scheme of digital video watermarking, Int. J. Sci. Res., vol.4,pp. 460–465, 2015.

[27] H. Zhang, Q. Zhu and X. Guan, "Probe into Image Segmentation Based on Sobel Operator and Maximum Entropy Algorithm," 2012 International Conference on Computer Science and Service System, Nanjing, 2012, pp. 238-241.

[28] S. D. Thepade and P. Bidwai, "Iris recognition using fractional coefficients of transforms, Wavelet Transforms and Hybrid Wavelet Transforms," 2013 International Conference on Control, Computing, Communication and Materials (ICCCCM), Allahabad, 2013, pp. 1-5.