IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 May 2022, accepted 27 June 2022, date of publication 18 July 2022, date of current version 21 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3191667

RESEARCH ARTICLE

Toward Authentication of Videos: Integer Transform Based Motion Vector Watermarking

RAFI ULLAH[1], **SULTAN DAUD KHAN**[1], **MOHIB ULLAH**[2], **(Member, IEEE), FADI AL-MACHOT**[3], **AND HABIB ULLAH**[3]

[1]Department of Computer Science, National University of Technology, Islamabad 44000, Pakistan
[2]Department of Computer Science (IDI), Norwegian University of Science and Technology (NTNU), 2815 Gjøvik, Norway
[3]Faculty of Science and Technology (REALTEK), Norwegian University of Life Sciences (NMBU), 1430 Ås, Norway

Corresponding author: Habib Ullah (habib.ullah@nmbu.no)

ABSTRACT Nowadays, digital content like videos, audio and images are widely used as evidence in criminal courts and forensic laboratories. Due to the advanced low-cost and easily available multimedia/communication tools and softwares, manipulation of the content is a no-brain task. Thus, the protection of digital content originality is a challenge for the content owners and researchers before it can be produced in court or used for some other purpose. We proposed a motion vector watermarking technique that validate and authenticate videos. We are embedding the correlated watermark in the integer wavelet transform domain. In our method, the selection of embedding areas is based on the variation of motion vectors. The video frames are fully protected in both spatial and transform domains since the watermark is correlated with the approximation subbands of wavelet transform before embedding. The proposed technique can concisely determine the attacked regions. The results validate the performance of the proposed approach in terms of quality metrics like peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), normalized coefficients (NC) and bit error rate (BER).

INDEX TERMS Watermarking, authentication, motion estimation, integer wavelet transform, embedding, extraction.

I. INTRODUCTION

Regular use of the Internet has increased the transmission and sharing of digital content. Trillions of bits are created and shared in seconds and hence many issues like proof of ownership, copyright protection, authentication, controlling illegal distribution, broadcast monitoring, transaction tracking etc. become more and more significant [1], [2]. Digital watermarking, which is the process of embedding secret information in the cover work e.g. videos, is one of the solutions to the above-mentioned issues [3]. In the last couple of decades, many researchers have introduced watermarking techniques to protect multimedia content and these watermarking techniques were elaborated with different criteria like watermarks should be imperceptible, blind, robust to different attacks, not removable, and not accessible to unauthorized users, etc. [4].

A watermarking model has three important functions: watermark embedding, watermark extraction and watermark detection. We usually embed digital signatures, logos, symbols, etc. in the most appropriate areas of the content and its suitability is determined by many factors e.g. robustness is focused on copyright protection and fragility is focused on authentication. The watermarking algorithms are designed according to the input data e.g. still images, audio, videos and other documents [5].

In this paper, we introduce a blind semi-fragile motion vector watermarking algorithm that authenticates the video content and can survive against different attacks up to some extent. The motion vectors represent the displacement from the matching block to the current block. The frames are extracted from the input video and the motion vectors are generated by using one of the block-based motion estimation techniques. There are several block-based motion estimation techniques i.e. full/exhaustive search (FS) [6], three-step

The associate editor coordinating the review of this manuscript and approving it for publication was Zijian Zhang.

search (TSS) [7], new three-step search (NTSS) [8], four-step search (FSS) [9], diamond search (DS) [10], two-dimensional logarithmic search (TDLS) [11], orthogonal search algorithm (OSA) [12] to name a few. The suitable vectors are selected for embedding purposes. The integer wavelet transform is applied to the target frame where the watermark bits are embedded in the suitable coefficients. Before embedding, we correlate the watermark with the coefficients other than those coefficients that are used for embedding i.e. the features selected for correlation must not belong to the features to be watermarked. Thus, there is no conflict while choosing the coefficients for embedding and correlation [13]. Due to this correlation, the proposed technique can increase content security as well as deal with collage/counterfeiting attacks.

II. RELATED WORK

Initially, digital watermarking has been used for the protection of still images. However, in the last few years, numerous watermarking algorithms have been designed for the protection of video content. In the literature, we can find several watermarking techniques that protect videos for different applications like medical, military, surveillance etc. Due to the temporal dimension in videos, watermarking algorithms for videos are complex as compared to the techniques that are used for still images [14]. Many researchers extend digital image watermarking techniques to video watermarking. The watermarks are embedded in all frames of the video which make it robust against different attacks but in return, it becomes time-consuming and degrades the imperceptibility of the watermark [7], [15]–[17].

Usually, watermarking algorithms are designed according to possible attacks. Cox *et al.* [2] introduced four types of attacks. a) The attacker has no watermark detector and also knows nothing about the watermark algorithm. b) Attacker is experienced in watermarking that provokes the collusion attacks even without knowing about the algorithms. c) Attacker is assumed to know the algorithms but has no information about the secret key. d) Attacker has a watermark detector even though nothing knows about the algorithms. Spatial domain video watermarking methods are also proposed [18], [19]. The authors are embedding watermarks in spatial domain components of the video frames having low complexity and being easy to implement but fragile to various attacks. Furthermore, the frequency domain video watermarking is proposed, where the magnitudes of the coefficients are modulated in transform domain such as discrete cosine transform (DCT), discrete wavelet transform (DWT), integer wavelet transforms (IWT), dual-tree complex wavelet transform (DTCWT), singular value decomposition (SVD) [20]–[28]. The robustness has been increased but is more complex as compared to the spatial domain algorithms. Qingliang *et al.* [29] proposed DWT and singular value decomposition (SVD) based video watermarking to enhance the visual perception and robustness of the watermark. They divide the input video into different scenes based on the histogram of the adjacent frames and extract the keyframes. The authors claim how to optimally find the target wavelet coefficients for embedding. Similarly, video watermarking in the DWT-SVD domain is proposed where the Fibonacci-based keyframes selection is introduced [30] and also scene change detection is identified. Block-based undecimated discrete wavelets transform (UDWT) based video watermarking has been proposed in [31] where 8 × 8 blocks are used for embedding and the two AC coefficients are used for embedding a watermark bit. The masking propertied of HVS of the UDWT makes the watermarking scheme oblivious. Margarita and Alexandr [32] proposed a high-resolution video watermarking based on code-128 barcoding and DWT domain. The authors proposed a low-cost textual video watermarking using barcoding which allows the full restoration of the watermark under some internet attacks.

Motion vector video watermarking techniques have been proposed by many researchers in the last few years. Imen *et al.* [33] exploited the human visual system (HVS) to select the frames to embed the watermark effectively. The authors proposed blind and robust motion vector video watermarking based on SVD. The authors have selected the fast motion frames in each shot for watermark embedding consequently reducing the watermark payload and improving imperceptibility. Jie and Songbin [34] proposed a motion vector watermarking for HEVC (high-efficiency video coding) videos and they claimed that this is the first information hiding technique for securing HEVC videos. An efficient video watermarking is proposed in [35], where the block-based watermark embedding and extraction are considered in the high-resolution videos. It is difficult to comply with the latest high-resolution video where compression up to 50 % is allowed during transmission on a communication channel. The luminance of a certain block of the input video is modified rather than the (high-efficiency video encoder) HEVC encoder.

An uncompressed motion vector watermarking has been proposed by Anurag and Monika [36]. They embedded a logo watermark in the frequency domain of an uncompressed AVI video. For embedding purposes, the authors are using a back-propagation neural network to select the target blue channel video frames. Three different types of attacks namely: scaling, compression and Gaussian noise were applied to examine the robustness of the scheme. In [37], the authors reviewed a brief survey on robust video watermarking for copyright protection based on original and compressed videos. The authors have discussed many challenges in robust video watermarking e.g. tradeoff in watermark capacity and imperceptibility, computational complexity, video coding standards, detection of the watermark in a small video segment, and real-time performance to ensure video smoothness. Similarly, in [26], the authors have proposed a robust video watermarking algorithm, where the complexity of the video watermarking system is decreased and the imperceptibility and robustness are improved under potential attacks.

The rest of the paper is organized as follows. The proposed method (watermark generation, embedding and extraction) is discussed in Section 3. Experimental results are provided in Section 4. In Section 5, a detailed analysis of the proposed algorithm is given. In Section 6, we conclude our work.

III. PROPOSED METHOD

The block diagrams of the proposed system (watermark embedding and extraction) are presented in Fig. 1 and Fig. 2. The system contains four consecutive parts: Obtaining motion vectors, generation and processing, watermark embedding and watermark extraction. Two standard video sequences, Suzie and Football are used for our experiments [38]. The Suzie video is a five seconds video (150 frames @30 frames/sec) with a frame size of 144 × 192, while the Football video is an eight seconds video 260 frames @30 frames/sec) with frame size 288 × 352. Both the input videos are coloured uncompressed AVI videos. The algorithm is implemented in MATLAB 2015(a).

A. MOTION VECTORS SELECTION

The motion vectors are obtained by using the full search motion estimation technique. The selection of target frames/areas is based on the magnitudes of motion vectors (MVs). The (MVs) having the magnitudes of the defined threshold are selected i.e. the areas having such (MVs) are selected for watermark embedding.

B. WATERMARK GENERATION AND PRE-PROCESSING

A binary random sequence of watermark bits is generated based on a private key. Before embedding, the watermark bits are correlated with the coefficients other than to-be-modified coefficients. The size (resolution) of the watermark varies according to the embedding capacity. Due to correlation, the watermarked video will be robust against counterfeiting/collage attacks [13], [39], [40]. Two different types of collage attacks can be applied by the attackers. The first attack copies part of the watermarked image to an arbitrary position in another authenticated image where the same watermark is used. Since both images are watermarked by the same watermark, the newly forged image will pass the security check. The second type combines parts of the watermarked images of the same owner and forges a new image by combining those parts and keeping their relative positions in the original images [41].

C. WATERMARK EMBEDDING

As can be seen in Fig. 1, the original video is converted into a frame/image sequence. The motion vectors are generated by applying the full search block-based motion estimation algorithm. The purpose of the full search BMA is to get efficient motion vectors. Based on the magnitude, the motion vectors are selected and hence the keyframes are selected accordingly. The randomly generated watermark is XORed with the approximation of the frame. The keyframes are then passed through the IWT using the lifting scheme which is the fast approach of DWT [42]. The LL1 subband is used to correlate with the randomly generated bit to make the bits to-be-ready for embedding. The watermark bits are embedded in the suitable coefficients which cause fewer perceptual artefacts. The summary of the watermark generation and embedding procedure is provided in Algorithm 1.

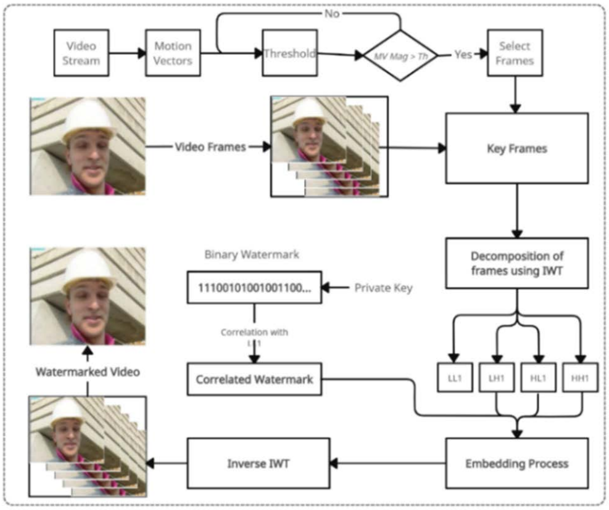

FIGURE 1. Block diagram of the watermark generation and embedding.

TABLE 1. Symbols and definition of parameters used in algorithms 1 and 2.

S.no	Symbol	Definition
1	$\mathbb{V}_i$	Input video
2	τ	Random bit. The length of r varies
3	η	Secret key initially used in water mark generation
4	f_r	Video frame
5	mv	motion vectors
6	Fs	Full search (blocked based motion estimation)
7	f_s	Selected frame
8	Ω	Magnitude of motion vector
9	Γ	Threshold
10	f_{wd}	Frame in wavelet domain
11	ν	Vector created by concatenating the sub bands of fwd excluding LL_1
12	α	Group of vectors. ν is divided into group according to the watermark strength
13	α_w	Weighted mean of α
14	$\mathbb{Q}$	Quantized value of the α_w and calculate as
15	α_w'	Modified α_w
16	ν_m	Modified vector obtained from the modified group
17	LH_1, HL_1, HH_1	Sub-bands obtained from modified vector
18	λ_{inv}	Inverse Integer Wavelet Transform
19	λ	Integer Wavelet Transform
20	f_{rem}	Remaining frames not selected for watermark embedding
21	$\mathbb{V}_o$	Watermarked video
22	ω	Extracted watermark
23	ω_f	Final watermark to be embedded

The symbols used in Algorithm 1 and Algorithm 2 are given in Table 1.

D. WATERMARK EXTRACTION

The block diagram of the watermark extraction is presented in Fig. 2. The reverse procedure of watermark embedding is followed to extract the embedded watermark and check the integrity of the watermarked video. All of the keys, group size, wavelet type, quantization factor, etc. are supposed to be available on the receiving side (send through a private communication channel). We take the watermarked video as an input and convert it into several frames. Full-search BMA is applied to obtain the motion vectors. The keyframes are selected based on the magnitude of the motion vectors.

One-level decomposition is applied to the selected frames and the areas are searched where the watermark bits are embedded. The embedded bots are extracted from the to-be-checked areas. The extracted bits are then XORed with the approximation of the frame.

Algorithm 1 Algorithm for Watermark Embedding

```
Input: Video sequence $\mathbb{V}_i$
Output: Watermarked video sequence

$\mathbb{V}_o, \omega$
function Watermark embedding($\mathbb{V}_i$)
    Generate initial water mark $\omega_i = \tau \oplus \eta$
    Split $\mathbb{V}_i$, into $n$ number of frames.
    Let $= \mathbb{V}_i\{f_1, f_2, \ldots, f_n\}$.
    for all frames $f_j \in \mathbb{V}_i$ do
        Compute $mv_j \leftarrow Fs(f_j)$
        if $\Omega_j \leq \Gamma$ then
            insert $f_j$ in $f_s$
        end if
    end for
    Decompose $f_s : f_{wd} \leftarrow \lambda(f_s)$
    $\omega_f \leftarrow \omega_i \oplus LL_1$
    $\nu \leftarrow \{LH_1, HL_1, HH_1\}$
    Divide $\nu$ into group of vectors $\alpha$
    $\alpha_w \leftarrow (\alpha)$; weighted mean of $\alpha$
    for all frames $f_j \in \mathbb{V}_i$ do
        if $\mathbb{Q}(\alpha_w) = \omega_f$ then
            $\alpha'_\omega = [\alpha_\omega + \frac{\mathbb{Q}}{2}].\mathbb{Q} + \frac{\mathbb{Q}}{2}$
        else
            $\alpha'_\omega = [\alpha_\omega - \frac{\mathbb{Q}}{2}].\mathbb{Q} - \frac{\mathbb{Q}}{2}$
        end if
    end for
    $\nu_m \leftarrow \alpha'_\omega$; Groups are converted into vector
    $\{LH'_1 HL'_1 HH'_1\} \leftarrow \nu_m$; Convert to subbands
    $f'_j \leftarrow \lambda_{inv}(\{LH'_1 HL'_1 HH'_1\})$
    Create video from $f'_j$ and remaining frames as
    $\mathbb{V}_o \leftarrow [f'_s, f_{rem}]$
    return $\mathbb{V}_o, \omega$
end function
```

The watermark is generated on the extraction side using the same key as was used on the embedding side. Finally, the extracted bits are compared to the generated watermark. If both are the same, then the video is authentic, otherwise it is a tampered one. The summary of watermark extraction is provided in Algorithm 2.

IV. EXPERIMENTAL RESULTS

Frames, keyframes and scenes extraction of Suzie and Scene Change videos are shown in Fig. 3. In Fig. 3 (a), the video is considered a single scene video and in Fig. 3 (b), the video has abrupt changes and has several different scenes. Some of the frames from five different scenes are shown in Fig. 3 (b). We input a five seconds Suzie video and divide it into frames (150 frames @ 30 frames/sec). Scene change has a very important role in motion estimation. Scene changes have been detected in [43] and [44] using statistical features and intra-mode statistical constraints. Similarly, the abrupt and gradual scene change has been detected in [45] where the prediction mode, motion vectors and the DC component of inter-prediction residual are used. In *suzie* video, there is no abrupt change in the scene. Fig. 4 shows the scene change in Suzie, Football and SceneChange videos. The changes in Suzie and Football video are below the pre-defined threshold which means there is no scene change in the video. However, in the SceneChange video, the scenes are changed abruptly. We tested our algorithm considering Football and Suzie videos. In the Football video, we can see more changes in the mean values of frames (45-140) as compared to the mean values in the Suzie video (86-125). The videos i.e. Suzie and Football can also be divided into different scenes based on the differences in the histogram of adjacent frames [29]. In case, if the input videos have scene changes, then the video will be divided into different scenes and then in each scene, the target frames are selected based on the magnitudes of the motion vectors. The histogram difference between the consecutive frames of different videos is shown in Fig. 4. The frequent changes in the scenes have several peaks. The Suzie and Football videos have almost the same variation in the histograms (Fig. 5).

Algorithm 2 Algorithm for Watermark Extraction

```
Input: Video sequence $\mathbb{V}_o$
Output: Label $\mathbb{L}$
function Watermark Extraction($\mathbb{V}_o$)
    Use Algorithm 1 to generate watermark $\omega_f$ and $\omega$
    if $\omega_f = \omega$ then
        $\mathbb{L} =$ "Video is authentic"
    else
        $\mathbb{L} =$ "Video is tampered"
    end if
    return $\mathbb{L}$
end function
```

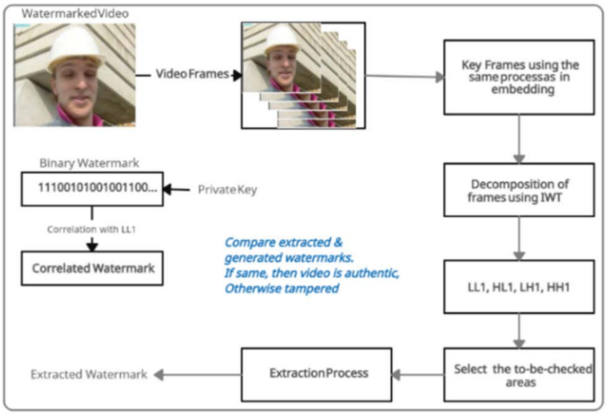
WatermarkedVideo
VideoFrames
Key Frames using the same processas in embedding
Decomposition of frames using IWT
LL1, HL1, LH1, HH1
Select the to-be-checked areas
ExtractionProcess
Extracted Watermark
Binary Watermark
11100101001001100...
PrivateKey
Correlation with LL1
Correlated Watermark
Compare extracted & generated watermarks.
If same, then video is authentic,
Otherwise tampered

FIGURE 2. Block diagram of the watermark extraction.

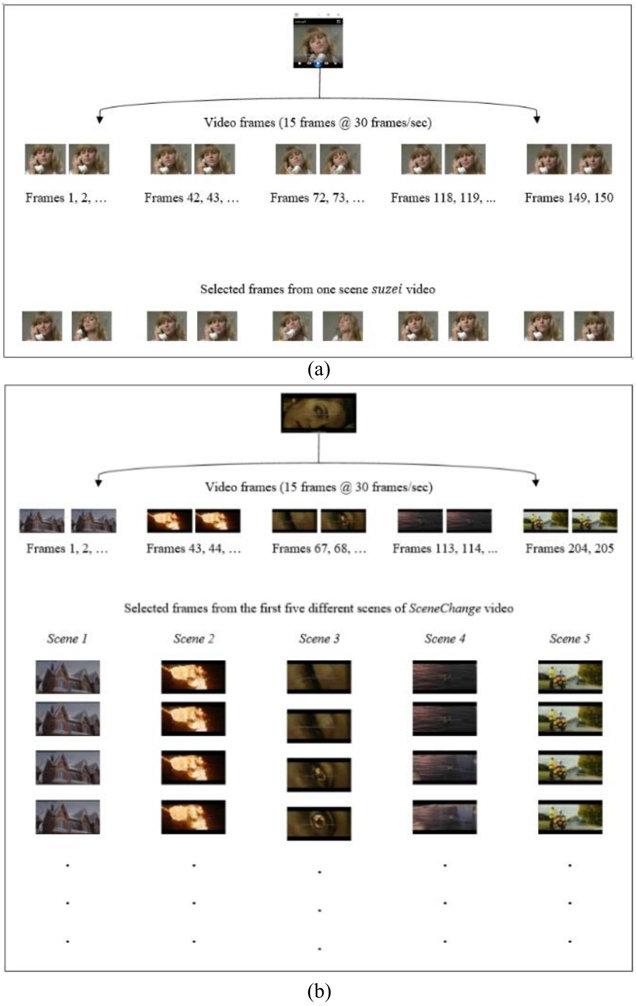

FIGURE 3. (a). Suzie video frames. (b). SceneChange video frames and scene division.

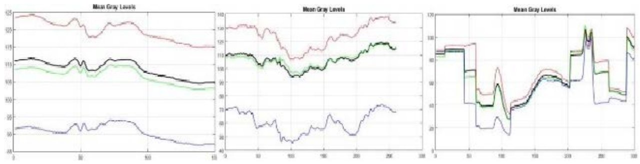

FIGURE 4. Changes in mean values of the Suzie, Football and SceneChange video frames (for R, G, B and its gray version).

In Fig. 5, we can see the histogram difference in consecutive frames of Suzie, Football and SceneChange videos. The scene changes are obvious in the SceneChange video. Thus, if we input a video stream having frequent scene changes, then we will divide the video first into scenes and calculate the motion vectors of every scene separately. This is because; the motion estimation cannot be traced when there is an abrupt change in the scene. In the first two videos, only one scene is considered as the motion vectors trace the movement easily. However, in the last video, there is a disconnectivity between two different scenes. The number of scenes are calculated based on the histogram difference. We have nine different scenes in the last video and we have to deal with every scene separately when we embed and extract the watermark bits.

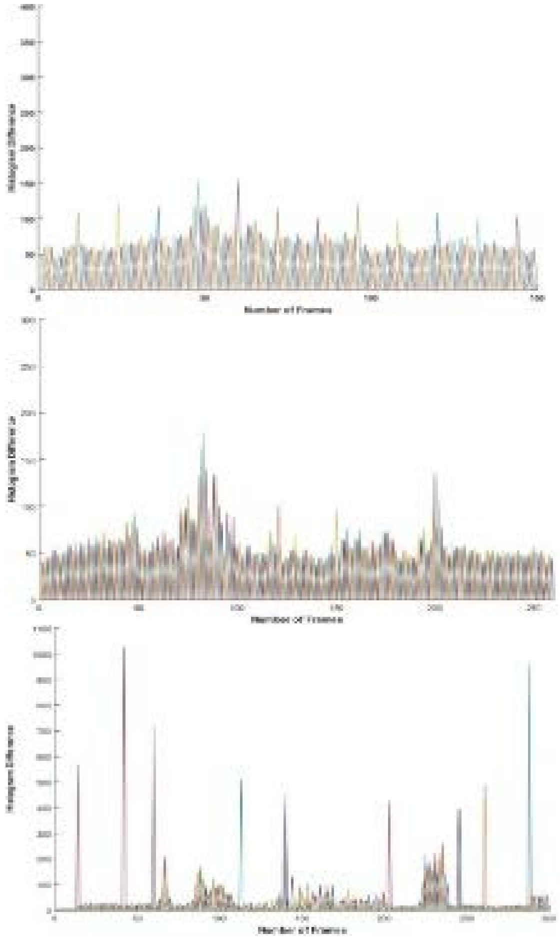

FIGURE 5. Histogram difference in consecutive frames of Suzie, Football and SceneChange videos.

The histogram of the 35th frame of the Football video and its corresponding watermarked frame is shown in Fig. 6. From histograms of the original and watermarked frames, one can see that the imperceptibility which is the most important property of watermarking, the proposed algorithm performs well in terms of visual similarity of the original and watermarked frames.

Similarly, the entropy of the original and watermarked frames of the Football video is calculated. The average entropy values of the original frames and watermarked frames are 7.1481 and 7.1620 respectively which show the low disturbance in the video after embedding the watermark bits.

The motion vectors of some of the frames of Suzie, Football and SceneChange videos are shown in Fig. 7(first row) and Fig. 7(second row). The motion vectors are extracted by using a full search block-based motion estimation algorithm. In Suzie and Football videos, we see that all of the frames have connectivity in terms of motion. However, in Fig. 6(third row), we see that the SceneChange video has complete disconnectivity in motion when the scene changes i.e. we can see the abrupt changes in motion vectors. Every scene is considered separately while embedding the watermark bits in the selected frames.

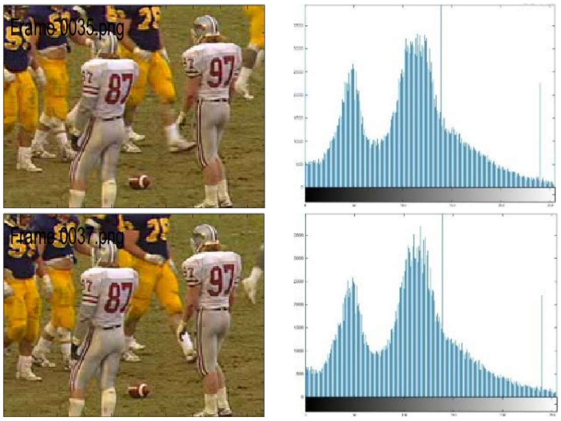

FIGURE 6. Column 1, the original and watermarked frame of the Football video. Column 2, the corresponding histograms of the frames.

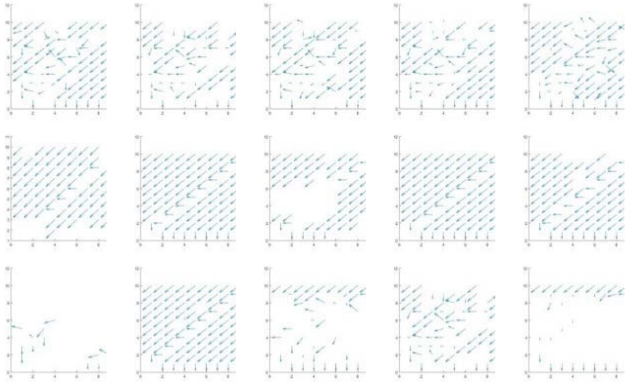

FIGURE 7. (a-c). Motion vectors of some of the frames of Suzie, Football and SceneChange videos.

In Fig. 8, the visual similarity (PSNR and SSIM) of some of the original and watermarked frames for Suzie, Football and SceneChange videos are shown. The PSNR and SSIM of the video (average PSNR and SSIM of the selected frames) are also given in Fig. 8 (first row). All of the frames are selected for watermark embedding (keyframes) as the threshold is low. When we keep the threshold high, then most of the frames are skipped. The size of the watermark is adjusted according to the embedding capacity. The group size is 12 i.e. 12 coefficients in a group, where one watermark bit is embedded in each group.

When we reduce the group size, the watermark payload increases and vice versa. In Fig. 8 (second row), the PSNR and SSIM for the first 100 frames of Suzie, SceneChange and Football videos are depicted. The watermark payload increases when the group size is low i.e. 6 coefficients in each group and hence it will reduce the imperceptibility and will improve the robustness. Similarly, by increasing the quantization factor, while embedding the watermark bits, the imperceptibility decreases and robustness increases and vice versa.

In Fig. 9 (first row), PSNR and SSIM between the original and watermarked images are given with group size 6 and quantization factor 20. Similarly in Fig. 9 (second row), the PSNR and SSIM are given with group size 12 and quantization factor 10. By changing the group size and quantization factor, the imperceptibility and robustness are affected. In Fig. 8 (first row) we see that the PSNR is around 33∼34, because of high values of quantization factor and low group size. The high value of the quantization factor i.e. 20 in this case, increases the robustness of the watermark against high compression. In the second row of Fig. 9, the quantization factor is 10 and it will reasonably survive compression up to some extent. The detail about compression is given in Section 4.4.

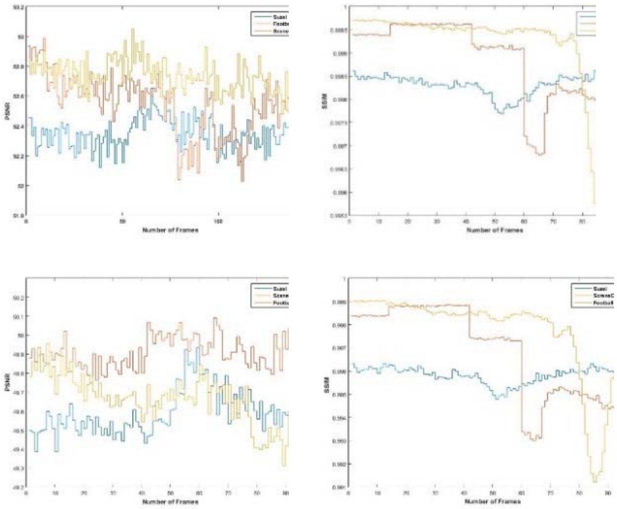

FIGURE 8. PSNR and SSIM of Suzie, Football and SceneChange videos.

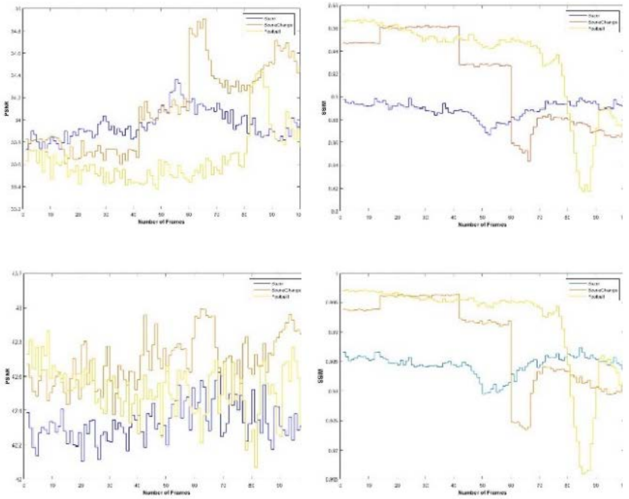

FIGURE 9. PSNR and SSIM between one hundred original and watermarked frames.

The number of erroneous bits of the one hundred original and watermarked frames of Suzie, Football and SceneChange are given in Fig. 10 (first row). The unsigned binary representations of original and watermarked frames are compared. The sizes of frames are 256 × 256 unsigned integers. Each frame contains $256 \times 256 \times 8 = 524,288$ bits. In Fig. 10

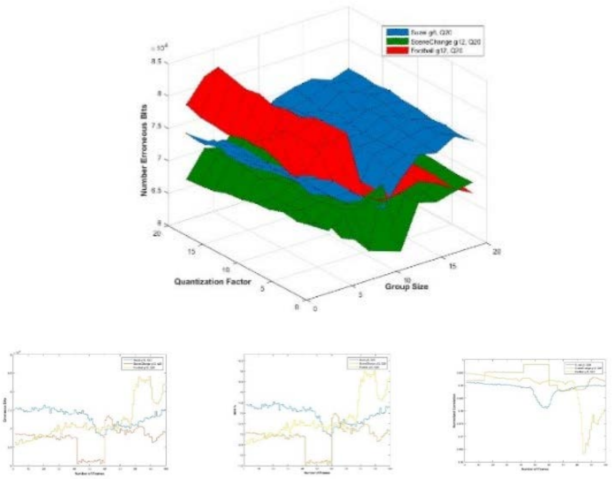

FIGURE 10. Number of erroneous bits & their percentage is a normalized correlation of original and watermarked frames.

(second row), the number of erroneous bits (differences in binary representation), the percentage of erroneous bits with original bits and also the normalized correlation of watermarked frames are given.

V. ANALYSIS OF THE PROPOSED SYSTEM

A. ANALYSIS OF EMBEDDED WATERMARK

The authenticity of the proposed system is based on the number of embedded bits in the video. The watermark payload depends on the group size where one bit is embedded in each group. When the group size increases, the payload decreases and vice versa. The payload directly affects the visual perception of the watermarked video and also, affects the authenticity/fragility of the watermark. Low payload increases fragility and imperceptibility and vice versa.

The quality metrics i.e. PSNR and SSIM have been used to measure visual similarity. Higher PSNR and SSIM values indicate good imperceptibility and vice versa. The PSNR and SSIM are calculated using (1) and (2).

$$PSNR = 20log_{10}(255^2/MSE) \tag{1}$$

where MSE is the mean square error between the original and watermarked images.

$$SSIM\,(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\mu_x^2 + \mu_y^2 + c_2)} \tag{2}$$

where x and y are two windows of size N × N, μ_x and μ_y are the averages of x and y, and σ_x^2 and σ_y^2 are the variances of x and y respectively. $c_1 = (K_1L)^2$ and $c_2 = (K_2L)^2$ where L is the maximum value i.e.$2^{\#bites}$ and $k_1 = 0.01$ and $k_2 = 0.03$ by default.

The robustness of any watermarking system is very important. Different types of attacks are applied to the proposed watermarking system and then we use NC and BER to test their robustness against the attacks. NC and BER are used to compare the similarities in the original and extracted watermarks [33]. The NC and BER are calculated in (3) and (4) respectively.

$$NC\left(W, W'\right) = \sum\sum \frac{W\,(i,j)\cdot W'(i,j)}{|W(i,j)|^2} \tag{3}$$

where W and W', are the original and extracted watermarks respectively.

$$BER = \frac{1}{s}\sum_{j=1}^{s}\left|W'\,(i,j)\,,W(i,j)\right| \tag{4}$$

where s is the size of the watermark, W and W' are the original and extracted watermark bits respectively. The NC value near 1 means the correlation between the embedded and extracted watermarks is high.

B. TAMPER DETECTION AND ITS LOCALIZATION

In the authentication process, first, we check the similarity of the extracted and original watermarks. If they are the same then, we conclude that the video is authentic, otherwise tampered with. In case, when the video tampers with any communication channel, then the strength of the attack is checked and also locates the tampered regions. All those frames of the video are checked which were selected for the watermark embedding. For all other frames, we can just conclude that the frames are verified not.

As the watermark capacity is associated with the group size but it does not affect the accuracy of tamper detection and localization. For every erroneous bit, the corresponding group is considered an unverified group. Further, all the groups are mapped back to the original positions, where the unverified coefficients belonging to the tampered regions converge together and hence, have more density. All other unverified groups that do not belong to the tampered regions are sparsely scattered [46]. This verification is done for all the frames of the video.

C. CLASSIFICATION OF ATTACKS

The proposed algorithm can classify malicious and friendly manipulation. When the frames(s) of the watermarked video tampers, the corresponding groups cannot be verified and hence are considered to be unauthentic. Further, we classify the group in two ways. If a pixel is erroneous and one of its neighbouring pixels is also erroneous, then it will be considered illegal tampering, otherwise, it is legal tampering like the compression of the frame up to some extent [47]. In this paper, we are not dealing with the temporal part of the compression, where the unwanted frames are removed. We can only deal with the compression on the individual frames i.e. if the frames are extracted from the watermarked video and then compressed.

If the watermarked video (some and/or all of the frames) have tampered with a malicious attack i.e. geometric attack e.g. cut/copy past, scaling, rotation, translation etc., it will be considered as a malicious attack. We do not differentiate the incidental and malicious attacks by the number of erroneous pixels, but by how erroneous pixels appear. For example, if someone removes a very small area maliciously from one and/or some of the video frames, the algorithm checks the sparsity/density of erroneous pixels. The lossy compression with an 80% quality factor is considered a friendly attack. Although, due to compression the watermarked frames have more erroneous pixels all of them are sparse, therefore it will be considered a friendly/legal attack.

D. ISSUES THAT AFFECT THE PERFORMANCE OF THE SYSTEM

When the group size increases or decreases, it will affect the performance of the proposed algorithms, in terms of time consumption and imperceptibility. Also, the quantization factor will affect the survival of the proposed approach against the compression of the frames i.e. the quantization factor controls the robustness of the system against compression. Also, the wavelet type, secret keys etc. must be provided to the receiver as it is the main issue with any of the watermarking algorithms. We can use public-key crypto watermarking, where the receiver can use his/her key. RSA encryption is one of the encryption, where the receiver can use his/her key where the private key is used for encrypting the watermark bits and the public key is used for its decryption [48]. So, before embedding, if we encrypt the correlated watermark using RSA encryption, then the secret keys and other necessary data related to embedding can be provided without sending it on the private/communication channel. In this case, the watermark must contain the keys and other necessary embedding data rather than random bits.

E. SURVIVAL AGAINST LOSSY COMPRESSION

As we increase the quantization factor, the survival against JPEG lossy compression of the frame increases. The proposed approach differentiates the acceptable (incidental) and unacceptable (malicious) compression of video frames. One of the frames from the SceneChange video is given in Fig. 11 (first row). In second row of Fig. 10, the compressed watermarked frames are given with quality factors 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, and 100%. In the third, fourth and fifth rows of Fig. 11, the frames show the erroneous pixels for different group sizes and quantization factors.

The visual quality of the watermarked frames and robustness of the watermark varies according to the group size and quantization factor. As we are increasing the compression ratio, the sparse pixels increase and become dense pixels when the compression strength crosses the acceptable threshold. The acceptability of the compression is application dependent. For sensitive applications i.e. military and medical applications, the acceptable threshold is very low where an average lossy compression is considered a malicious attack. In Fig. 11 (last row), the graph show number of erroneous pixels for different group sizes and quantization factors.

The difference in the original and extracted watermark for the uncompressed and compressed watermarked images is shown in Fig. 12. One of the frames from the SceneChange video is used where the embedded bits are extracted when the watermarked image is uncompressed and compressed with JPEG with 90% quality. There are no erroneous bits in the retrieved watermark (difference in the original and extracted watermarks) for the uncompressed watermarked frame and 301 erroneous bits when the watermarked frame is compressed.

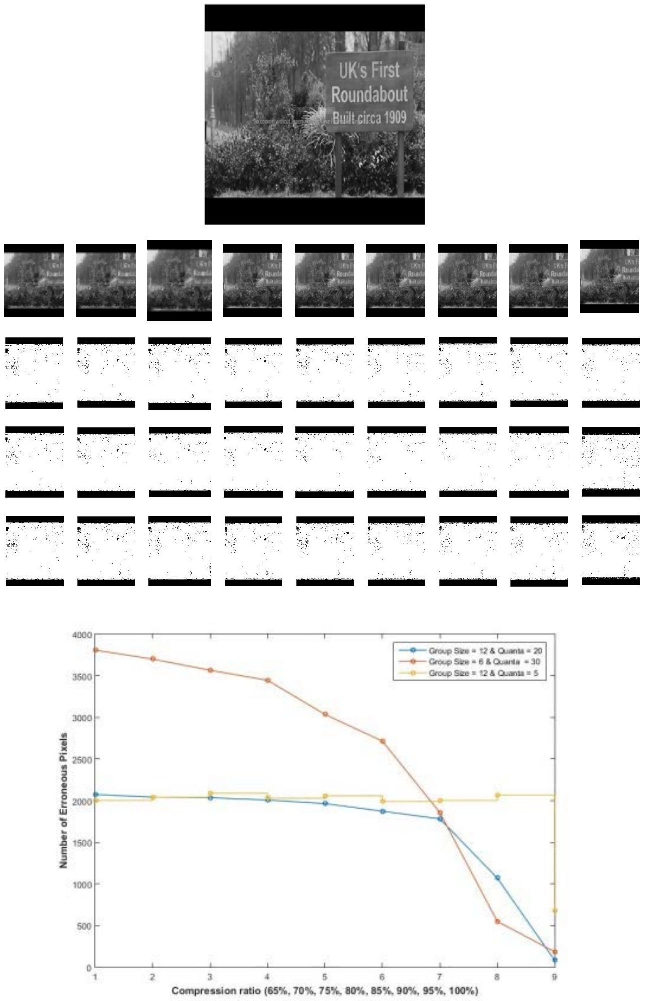

FIGURE 11. One of the extracted gray frame (irst row), the compressed version of a frame (second row) with compression quality 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, erroneous pixels in third, fourth and fifth row. The last row shows the graph of several erroneous pixels with different group sizes and quantization factors.

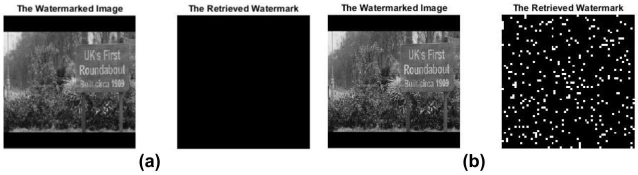

FIGURE 12. (a – b). Uncompressed and compressed watermarked video frame and retrieved watermarks.

F. ALGORITHM PERFORMANCE UNDER MALICIOUS ATTACKS

In Fig. 13, the frame from the SceneChange video is tested for different malicious attacks i.e. Poisson filter, Gaussian noise

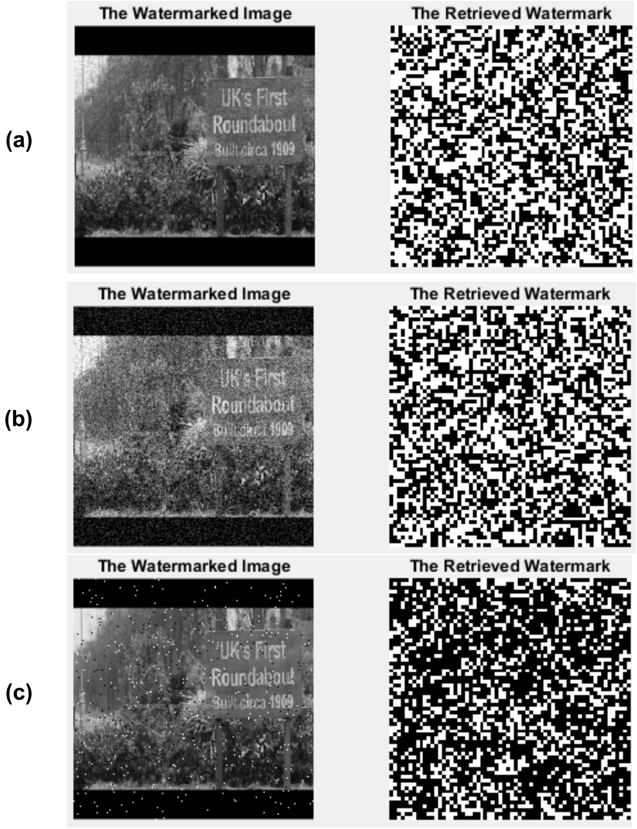

FIGURE 13. Watermarked frames were attacked with (a – c) Poisson filter, Gaussian noise (0.05) and salt and pepper noise (0.02) and the corresponding difference in original and extracted watermarks.

and salt and pepper noise and the number of erroneous bits for these noises are 2097, 2005 and 1295 respectively. The parameter values can adjust the severity of the noise. Most of the erroneous bits are dense and hence considered malicious attacks. However, after compressing the watermarked frame with 75% quality, there are many erroneous bits i.e. more than two thousand, still considered a friendly attack as all of the erroneous bits are sparse.

G. COMPARISON WITH OTHER APPROACHES

The proposed algorithm is compared with other approaches, [5], [29], [32] in terms of visual quality and robustness against different attacks, especially collage attack and lossy compression in Table 2, Table 3 and Table 4. The PSNR and SSIM of the original and watermarked frames of the video depend on the watermark payload. The watermark payload varied according to the group size (number of coefficients in one group). The group size increases, the watermark payload decreases and vice versa. Similarly, several erroneous bits (BER %) after modifying the video by embedding the watermark bits also depend on the watermark payload. The BER increases when the watermark strength increases and vice versa. Also, the proposed algorithm is compared for detecting, differentiating different attacks and survival against lossy compression in Table 4. Increasing the quantization factor will increase the survival of the watermark against compression and vice versa. By this comparison, we see that the proposed approach performs well in terms of imperceptibility, robustness and semi-fragility. The counterfeiting/collage attack can also be detected because the watermark bits are correlated with the frames and regions that are not used in the embedding.

TABLE 2. Performance comparison of the proposed algorithm with [5], [29], [32] in terms of different attacks.

Method	Salt & Pepper Noise	Poisson Filter	Gaussian Filter	Frame Dropping
[5]	0.965	0.98	0.933	0.966
[29]	0.99	0.98	0.992	0.97
[32]	-	-	-	-
Proposed	0.97	0.94	0.95	0.95

TABLE 3. Performance comparison of the proposed algorithm [5], [29], [32] in terms of imperceptibility and bit error rate.

Method	PSNR	SSIM	BER (%age)
[5]	39.40 dB	-	-
[29]	35 – 37 dB	-	5% - 35%
[32]	60 dB	1.00	-
Proposed	51 – 54 dB	0.94 – 0.98	12% - 14%

TABLE 4. Performance comparison of the proposed algorithm [5], [29], [32] in terms of different features. the key features include differentiating individual and malicious attacks, survival against lossy compression and detecting counter feting attacks.

Key Features	Ref [5]	Ref [29]	Ref [32]	Proposed
Normalized coefficients (NC)	0.998	-	-	0.998
Differentiating incidental and malicious attacks	-	-	-	✓
Survival against JPEG compression	-	✓		✓
Detecting counterfeiting/collage attacks	-	-	-	✓

The results given for the proposed algorithm in Table 2 have group size = 12 coefficients, quantization factor = 20 and the algorithm applied to three videos Suzie, SceneChange and Football.

VI. CONCLUSION

In this paper, semi-fragile blind motion vector video watermarking has been proposed. The video is converted into frames and the embedding procedure is done in the transform domain (IWT domain) of the selected frames. The frame selection is done based on the motion vectors where the full-search block-based motion estimation algorithm is used to get the motion vectors. Before embedding the watermark bits, first, we correlate them with the approximation of not only the particular frame but also with the correlate with the approximation of the non-selected frames. By this correlation, all of the video frames are secured and any tampering in any frame of the video can be detected. PSNR and SSIM are used to check the visual similarity of the original and watermarked videos. BER and NC are used to test the robustness of the embedded watermark. The watermark extraction is the reverse procedure of the embedding procedure and the extracted watermark is then compared with the original watermark. If there is any tampering, then not only we can localize it but also check the strength of the attack that it is either legitimate or illegitimate. The results are compared with some of the previous approaches which show the superiority of the proposed algorithm.

REFERENCES

[1] A. Piva, F. Bartolini, and M. Barni, "Managing copyright in open networks," *IEEE Internet Comput.*, vol. 6, no. 3, pp. 18–26, May/Jun. 2002.

[2] I. J. Cox, M. Miller, J. A. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*. Burlington, VT, USA: Elsevier, 2008.

[3] D. Rajani and P. R. Kumar, "An optimized blind watermarking scheme based on principal component analysis in redundant discrete wavelet domain," *Signal Process.*, vol. 172, pp. 1–9, Jul. 2020.

[4] A. Zotin, M. Favorskaya, A. Proskurin, and A. Pakhirka, "Study of digital textual watermarking distortions under internet attacks in high resolution videos," *Proc. Comput. Sci.*, vol. 176, pp. 1633–1642, Jan. 2020.

[5] S. Jie, L. Qi, and S. Chun-Lin, "Robust video watermarking using a hybrid DCT-DWT approach," *J. Electron. Sci. Technol.*, vol. 18, no. 2, pp. 1–10, Jun. 2020.

[6] D. Hai-Yang, Z. Ya-Jian, Y. Yi-Xian, and Z. Ru, "Robust blind video watermark algorithm in transform domain combining with 3D video correlation," *J. Multimedia*, vol. 8, no. 2, pp. 161–167, Apr. 2013.

[7] O. S. Faragallah, "Efficient video watermarking based on singular value decomposition in the discrete wavelet transform domain," *AEU-Int. J. Electron. Commun.*, vol. 67, no. 3, pp. 189–196, Mar. 2013.

[8] W. Yu, D. Hu, N. Tian, and Z. Zhou, "A novel search method based on artificial bee colony algorithm for block motion estimation," *EURASIP J. Image Video Process.*, vol. 2017, no. 1, pp. 1–14, Dec. 2017.

[9] S. A. K. Mostafa and A. Ali, "Multiresolution video watermarking algorithm exploiting the block-based motion estimation," *J. Inf. Secur.*, vol. 7, no. 4, pp. 260–268, 2016.

[10] T. Bernatin and G. Sundari, "Comparative analysis of different diamond search algorithms for block matching in motion estimation," *ARPN J. Eng. Appl. Sci.*, vol. 12, no. 11, pp. 3550–3553, 2017.

[11] M. Jakubowski and G. Pastuszak, "Block-based motion estimation algorithms—A survey," *Opto-Electron. Rev.*, vol. 21, no. 1, pp. 86–102, Jan. 2013.

[12] S. P. Metkar and S. N. Talbar, "Fast motion estimation using modified orthogonal search algorithm for video compression," *Signal, Image Video Process.*, vol. 4, no. 1, pp. 123–128, Mar. 2010.

[13] N. Ishihara and K. Abe, "A semi-fragile watermarking scheme using weighted vote with sieve and emphasis for image authentication," *IEICE Trans. Fundamentals Electron., Commun. Comput. Sci.*, vol. 90, no. 5, pp. 1045–1054, May 2007.

[14] N. Leelavathy, E. V. Prasad, and S. S. Kumar, "Video watermarking techniques: A review," *Int. J. Comput. Appl.*, vol. 104, no. 7, pp. 24–30, Oct. 2014.

[15] T. R. Singh, K. M. Singh, and S. Roy, "Video watermarking scheme based on visual cryptography and scene change detection," *AEU-Int. J. Electron. Commun.*, vol. 67, no. 8, pp. 645–651, Aug. 2013.

[16] P. Rasti, S. Samiei, M. Agoyi, S. Escalera, and G. Anbarjafari, "Robust non-blind color video watermarking using QR decomposition and entropy analysis," *J. Vis. Commun. Image Represent.*, vol. 38, pp. 838–847, Jul. 2016.

[17] S. M. Youssef, A. A. ElFarag, and N. M. Ghatwary, "Adaptive video watermarking integrating a fuzzy wavelet-based human visual system perceptual model," *Multimedia Tools Appl.*, vol. 73, no. 3, pp. 1545–1573, Dec. 2014.

[18] M. U. Celik, G. Sharma, A. M. Tekalp, and E. Saber, "Lossless generalized-LSB data embedding," *IEEE Trans. Image Process.*, vol. 14, no. 2, pp. 253–266, Feb. 2005.

[19] C.-W. Tang and H.-M. Hang, "A feature-based robust digital image watermarking scheme," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 950–959, Apr. 2003.

[20] A. Joseph and K. Anusudha, "Robust watermarking based on DWT SVD," *Int. J. Signal Image Secur.*, vol. 1, no. 1, pp. 1–5, 2013.

[21] N. M. Makbol and B. E. Khoo, "Robust blind image watermarking scheme based on redundant discrete wavelet transform and singular value decomposition," *Int. J. Electron. Commun.*, vol. 67, no. 2, pp. 102–112, Feb. 2013.

[22] Z.-M. Lu, H.-Y. Zheng, and J.-W. Huang, "A digital watermarking scheme based on DCT and SVD," in *Proc. 3rd Int. Conf. Intell. Inf. Hiding Multimedia Signal Process. (IIH-MSP)*, Kaohsiung, Taiwan, Nov. 2007, pp. 241–244.

[23] S. D. Roy, X. Li, Y. Shoshan, A. Fish, and O. Yadid-Pecht, "Hardware implementation of a digital watermarking system for video authentication," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 2, pp. 289–301, Feb. 2013.

[24] H. Flórez and M. Argáez, "A model-order reduction method based on wavelets and POD to solve nonlinear transient and steady-state continuation problems," *Appl. Math. Model.*, vol. 53, pp. 12–31, Jan. 2018.

[25] M. Asikuzzaman and M. R. Pickering, "An overview of digital video watermarking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 9, pp. 2131–2153, Sep. 2018.

[26] Y. Himeur and A. Boukabou, "A robust and secure key-frames based video watermarking system using chaotic encryption," *Multimedia Tools Appl.*, vol. 77, no. 7, pp. 8603–8627, Apr. 2018.

[27] M. Fallahpour, S. Shirmohammadi, and M. Ghanbari, "A high capacity data hiding algorithm for H.264/AVC video," *Secur. Commun. Netw.*, vol. 8, no. 16, pp. 2947–2955, 2015.

[28] R. Chamlawi and A. Khan, "Digital image authentication and recovery: Employing integer transform based information embedding and extraction," *Inf. Sci.*, vol. 180, no. 24, pp. 4909–4928, Dec. 2010.

[29] Q. Liu, S. Yang, J. Liu, P. Xiong, and M. Zhou, "A discrete wavelet transform and singular value decomposition-based digital video watermark method," *Appl. Math. Model.*, vol. 85, pp. 273–293, Sep. 2020.

[30] S. Ponni alias Sathya and S. Ramakrishnan, "Fibonacci based key frame selection and scrambling for video watermarking in DWT–SVD domain," *Wireless Pers. Commun.*, vol. 102, no. 2, pp. 2011–2031, Sep. 2018.

[31] K. Meenakshi, K. Swaraja, K. Padmavathi, and G. Kurana, "A robust blind oblivious video watermarking scheme using undecimated discrete wavelet transform," in *Intelligent System Design*, vol. 1171, 2021, pp. 169–177.

[32] M. Favorskaya and A. Zotin, "Robust textual watermarking for high resolution videos based on code-128 barcoding and DWT," *Proc. Comput. Sci.*, vol. 176, pp. 1261–1270, Jan. 2020.

[33] I. Nouioua, N. Amardjia, and S. Belilita, "A novel blind and robust video watermarking technique in fast motion frames based on SVD and MR-SVD," *Secur. Commun. Netw.*, vol. 2018, pp. 1–17, Nov. 2018.

[34] J. Yang and S. Li, "An efficient information hiding method based on motion vector space encoding for HEVC," *Multimedia Tools Appl.*, vol. 77, no. 10, pp. 11979–12001, May 2018.

[35] D. R. Galiano, A. A. D. Barrio, G. Botella, and D. Cuesta, "Efficient embedding and retrieval of information for high-resolution videos coded with HEVC," *Comput. Electr. Eng.*, vol. 81, pp. 1–15, Jan. 2020.

[36] M. Anurag and S. Monika, "Uncompressed video watermarking using motion vectors and back propagation network," *Int. J. Eng. Res. Technol.*, vol. 3, no. 8, pp. 1023–1030, 2014.

[37] Y. Xiaoyan, W. Chengyou, and Z. Xiao, "A survey on robust video watermarking algorithms for copyright protection," *Appl. Sci.*, vol. 8, no. 10, pp. 1–26, 2018.

[38] [Online]. Available: https://media.xiph.org/video/derf/

[39] C.-T. Li and H. Si, "Wavelet-based fragile watermarking scheme for image authentication," *J. Electron. Imag.*, vol. 16, no. 1, pp. 1–9, 2007.

[40] C.-T. Li and Y. Yuan, "Digital watermarking scheme exploiting nondeterministic dependence for image authentication," *Opt. Eng.*, vol. 45, no. 12, pp. 127001-1–127001-6, 2006.

[41] O. Hosam, "Attacking image watermarking and steganography—A survey," *Int. J. Inf. Technol. Comput. Sci.*, vol. 11, no. 3, pp. 23–37, Mar. 2019.

[42] W. Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets," *Appl. Comput. Harmon. Anal.*, vol. 3, no. 2, pp. 186–200, Apr. 1996.

[43] J. Feng, A. Huang, and Y. Chen, "A novel scene change detection algorithm for H.264/AVC bitstreams," *Comput. Intell. Ind. Appl.*, vol. 1, pp. 712–716, 2008.

[44] W. Zeng and W. Gao, "Shot change detection on H.264/AVC compressed video," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Kobe, Japan, May 2005, pp. 3459–3462.

[45] Z. Yu and Z. Lin, "Scene change detection using motion vectors and DC components of prediction residual in H.264 compressed videos," in *Proc. 7th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, Singapore, Jul. 2012, pp. 990–995.

[46] R. O. Preda, "Semi-fragile watermarking for image authentication with sensitive tamper localization in the wavelet domain," *Measurement*, vol. 46, no. 1, pp. 367–373, Jan. 2013.

[47] H. Liu and M. Steinebach, "Semi-fragile watermarking for image authentication with high tampering localization capability," in *Proc. 2nd Int. Conf. Automated Prod. Cross Media Content Multi-Channel Distrib. (AXMEDIS)*, Leeds, U.K., Dec. 2006, pp. 143–152.

[48] K. Atul, *Cryptography and Network Security*. New York, NY, USA: McGraw-Hill, 2003.

RAFI ULLAH received the M.S. degree in computer system engineering from GIKI, Pakistan, in 2006, and the Ph.D. degree from PIEAS, Pakistan, in 2010. He is currently working as an Associate Professor with the Department of Computer Science, National University of Technology, Pakistan. His research interests include image and video processing, digital watermarking, and EEG signals.

SULTAN DAUD KHAN received the B.Sc. degree (Hons.) in computer engineering from the University of Engineering and Technology, in 2005, the M.Sc. degree (Hons.) in electronics and communication engineering from Hanyang University, South Korea, in 2010, and the Ph.D. degree in computer science from the University of Milano-Bicocca, in 2016. He is currently working as an Associate Professor with the Department of Computer Science, National University of Technology, Pakistan. He has published several papers in reputed journals and conferences. His research interests include crowd analysis, action recognition and localization, object detection, visual tracking, multi-camera, and airborne surveillance using deep learning techniques. He is an active reviewer of prestigious journals.

MOHIB ULLAH (Member, IEEE) received the bachelor's degree in electronic and computer engineering from the Politecnico di Torino, Italy, in 2012, the master's degree in telecommunication engineering from the University of Trento, Italy, in 2015, and the Ph.D. degree in computer science from the Norwegian University of Science and Technology (NTNU), Norway, in 2019. He is currently a Postdoctoral Research Fellow with NTNU, where he is involved in several industrial projects related to video surveillance. He teaches courses on machine learning and computer vision. He has published more than 30 peer-reviewed journals, conferences, and workshop articles. His research interests include medical imaging, crowd analysis, object segmentation, behavior classification, and tracking. He has served as a Program Committee Member for the International Workshop on Computer Vision in Sports (CVsports). He has also served as the Chair for the Technical Program at the European Workshop on Visual Information Processing. He is a Reviewer of well-reputed conferences and journals, including *Neurocomputing* (Elsevier), *Neural Computing and Applications* (Elsevier), *Multimedia Tools and Applications* (Spring), IEEE ACCESS, the *Journal of Imaging*, IEEE CVPRW, IEEE ICIP, and IEEE AVSS.

FADI AL-MACHOT received the German Diploma degree in computer science from the University of Potsdam, in 2010, the Ph.D. degree in computer science from Klagenfurt University, Austria, in 2013, and the Habilitation degree in applied computer science from the University of Lübeck, Germany, in 2020. He is currently working as an Associate Professor in machine learning (ML) with the Norwegian University of Life Sciences (NMBU), Norway. His work has been patented and published in peer-reviewed international conferences and journals. His research interests include deep learning, neural-symbolic learning, video understanding, cognitive modeling, and zero/few-shot learning. He is an active reviewer of well-known journals.

HABIB ULLAH received the M.S. degree in electronics and computer engineering from Hanyang University, Seoul, South Korea, in 2009, and the Ph.D. degree in information and communication technology from the University of Trento, Trento, Italy, in 2015. From 2015 to 2016, he was an Assistant Professor of electrical engineering with COMSATS University Islamabad, Wah Campus, Pakistan. From 2016 to 2020, he was an Assistant Professor with the College of Computer Science and Engineering, University of Ha'il, Ha'il, Saudi Arabia. In 2020, he was a Postdoctoral Researcher with The Arctic University of Norway, Tromsø, Norway. He is currently an Associate Professor with the Norwegian University of Life Sciences, Ås, Norway. His research interests include computer vision and machine learning.