Proposed Copyright Protection Systems for 3D Video Based on Key Frames

Abeer D. Salman[1], Hala B. Abdulwahab[2]

[1]Department of Computer Engineering Techniques, Al_Maarif University College, Ramadi, Iraq

[2]Department of Computer Science, University of Technology, Baghdad, Iraq

abeer.alnuaimi@yahoo.com, hala_bahjat@yahoo.com

Abstract- *Recently, three Dimension (3D) multimedia technology has grown over a wide range, where the 3D used in many applications such as medical, television, cinema, game, education, etc. The developments also include the computer network, these evaluations make the store and distribute the media very easy and at the same time very critical, where any person can access these data and manipulate it or redistribute it with illegal copyrights. Therefore, it is necessary to find a watermarking technique to keep the security of these media. 3D multimedia, especially 3D Video contains a huge amount of data, trying achievement copyright protection for all these data consumes the computation, and slow down securing process, thus 3D video key frame extraction process is necessary. In this paper, the histogram-based key frame extraction technique is applied to extract essential and important data of the 3D video. A modified invisible, and blind watermark system is suggested in this research for 3D Video copyrights and authentication protection using Dual-Tree Complex Wavelet Transform (DTCWT). The experiment results show that 3D video key extraction system has given high compression ratio (from 4:1 to 5:1) and the protection system could put an invisible watermark while keeping the quality of the 3D video, as well as the proposed watermark system, is robust where the watermark extracted successfully after applying a different types of attack.*

Index Terms*- 3D video, Multiview video, video summarization, Dual tree complex wavelet transform, 3D video dynamic watermark.*

I. INTRODUCTION

Transmission, distribution, and storage of the 3D multimedia data, whether images or videos, have become so easy alongwith keeping the quality of 3D data.These facilities belong to quick evaluations of the computer network and 3D multimedia technology itself [1], such easiness enable users to access, tamper and distribute the multimedia data [2]. When any user can access to multimedia data, he/she can take an unauthorized copy of the data and redistribution it illegally [2]. One of the most important threats regarding the protection of intellectual property of 3D Video is "Video piracy ". Where an unauthorized user without the approval of the copyright owner taking copies of the copyrighted video and then selling or distributing it [3]. A large source of video piracy is camcorder theft where a thief by using a camcorder captures digital movies from the large screen of the cinema, and via the Internet, a thief distributes a copied video over the world without any copyright protection. Filmmakers are facing this type of piracy [4]. Therefore, it's necessary to find techniques for preserving the copyrights of 3D videos [5] [6] [7] [8]. There are more methods for doing this, but the digital watermark is a traditional one [9]. The 3D video has many formats like classical stereo video, video plus depth, multi-view video, and multi-view plus depth video [10]. All these formats result in a huge amount of data. However, traditional watermark methods not suitable for huge video data that increases the need for methods that can abstracting the relevant information and thus making all operations on 3D video efficient such as

indexing, searching, browsing and securing. Unfortunately, researches regarded 3D video keyframe extraction is very little [11].

Multiview Video (MVV) is the format of the 3D video that is used in this paper. The concept of key frame extraction techniques will be surveyed briefly in this paper to show the necessity of using it in MVV. In this research, Dual tree complex wavelet transform (DTCWT) is explained in terms of how it works, its distinctive properties over other transforms, this transform can be used in a proposed watermarking system. During this research, two proposed systems are presented, one for extracting the key frames of 3D video and others for 3D video protection. Where four techniques are proposed for extracting important MVV frames. As well as this work introduces a modified watermark system called three Dimension video copyright and authentication protection (3DVCAP) to protect the copyright of the extracted frames and achieve 3D video authentication.

This paper is organized as follows: Section II presents 3D video keyframe extraction concepts. DTCWT is introduced in section III with some comparisons with other common transforms. Section IV gives the details of the proposed systems for keyframe extraction and protection. Finally, section V shows the experimental results of our systems under the most common quality measurements. The conclusion of the work is given in section VI.

II. KEY FRAME EXTRACTION

The classical video is composed of 20 to 30 frames per second. Video with one-hour length contains about 20x60x60 to 30x60x60 frames. MVV contains more than two classical videos, hence N (20x60x60) or N (30x60x60). N is the number of videos. It is normal that there is much redundant information contained in that number of frames. Therefore, it is essential to extract only the keyframes [12].

The keyframe extraction process can be defined as a powerful tool that selects a set of keyframes to represent video sequences [9]. The benefits of the keyframe extraction process are removing the redundant frames, decrease the computational complexity, makes the recognition efficient, and improve 3D video indexing, browsing & retrieval [12] [13, 14]. It is important to explain some terminologies used in the digital video field such as scenes, shots, and frames. Where any video segment divided into scenes, each video scene consists of one or more video shots that are set of interconnected frames. The hierarchical structure of digital video can be seen in Fig. 1. [11] [15] [16] [17]. Recently, everything going to shrink including digital data, especially video, thus if one or more frames can reflect the important information of a shot, these frames are the keyframes and must be extracted [15]. The video is compressed or shrank if keyframes are extracted from the video. The resulted video is small in size, contain all the vital information, as well as to require less time and resources for processing [17]
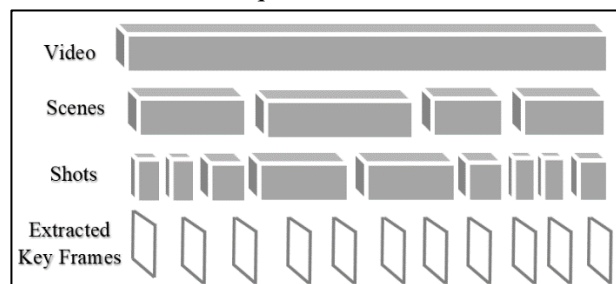

FIG. 1: STRUCTURAL HIERARCHY OF A VIDEO

The same processing chain for extracting important 2D video frames is used in the 3D video with tacking account to the depth and geometric information that is necessary for 3D moving scenes [11] [15]. In video compression, transmission, retrieval, and browsing applications, video analysis is an important and fundamental step [12]. It consists of two processes shot boundary detection SBD and

keyframe extraction process. The first step of video analysis is video segmentation, where depending on spatial-temporal criteria, the input video is segmented into video shots. From each shot, one or more keyframes are taken based on parameters specified previously by the user [11] [15].

The Shot boundary detection process includes two steps: first computing the visual features of the video frames and forming a feature vector. Computing relevant features are done by one of the following measures: Pairwise Pixel Differences, Statistical Differences (region-based technique), Feature-based (edge changes) [18], and Histograms [12, 13] [17] [19]. While the second step in shot boundary detection is comparing the visual features of sequential frames utilizing particular similarity measures, and to identify shot boundaries some decision criteria are used. Decision criteria can be static, or adaptive thresholds. Adaptive thresholds are the thresholds that depend on the statistics of the visual features used [11, 13, 15]. The keyframes are extracted after detecting the shot boundaries, some techniques consider the first and last frames of each shot as the keyframes [15]. There are other techniques for key-frame extraction like clustering, minimum correlation, minimum reconstruction error (MRE), curve simplification, and etc. [9] [11] [14] [17].

III. 3D VIDEO WATERMARK

As everything goes into development, also Digital watermarking has evolved very fast and used in many applications, mainly in copyrights protection and data authentication also utilized in digital copy tracking, broadcast monitoring, and Content Filtering [2, 3]. Any watermarking system faces the following challenges: imperceptibility, robustness, blind detection, payload, and security of the Watermark [4]. Imperceptibility means that the added watermark does not affect the quality of the hosted media, while the robustness means the watermark successfully can be extracted after an attack. If the watermark extracted without the need for the original media, the system called blind detection. Payload indicates the number of watermark bits. [3]

Digital watermarking can be decomposed into two types one-bit and multi-bit watermarking. In the first type, the detector examines the presence or absence of the watermark, thus the watermark is considered as a verification code, whilst multi-bit watermarking, the function of the detector is to extract watermark bits, bit by bit. The watermark is significant data like an ID or a logo. [2]. The watermarking systems can be classified according to the embedding domain into the spatial domain and frequency domain [1]. Embedding into the spatial domain is simple, and easy to implement, but, this type of embedding is not robust. To gain a watermark system with higher imperceptibility and robustness, the transform domain is used for embedding [2]. Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Singular Value Decomposition (SVD), and Dual tree complex wavelet transform (DT-CWT) are Commonly used transforms [1, 2, 20]. Fourier Transform (FT) very rarely used in image processing systems, spatially in image de-noising applications because it has some limitations such as it can retrieve only the global frequency content of a signal while the information of time is lost, and it's an inability to perform a multi-resolution analysis. To avoid the drawbacks of FT, Wavelet Transform (WT) is used. WT able to return the frequency and time content of a signal, WT also allows for a multiresolution analysis this belongs to the filter banks that WT is used. WT has three inherited types of Continuous Wavelet Transform (CoWT), Discrete Wavelet Transform (DWT), and Complex Wavelet Transform (CWT) [21].

In many image processing and computer vision applications, particularly in the field of digital watermarking, DWT is broadly used due to their attractive features like it can take advantage of the Human Visual System (HVS) features in a preferable way, DWT has the ability to embed watermark in an images with the better visual quality [2] [20] [22]. Additional to previous features DWT has the following benefits: Perfect energy, with substantially support filters, DWT has good reconstruction, no redundancy, and Low computation complexity [1] [23]. DWT's values suffer from Shift Variance, Lack

of Directionality, Absence of Phase Information, and Aliasing problems [4] [21] [22] [23]. This was the motivation to develop a new type of DWT called DT-CWT. This transforms solves the limitations associated with DWT such as [1] [21] [24] [25].

1. Approximate shift-invariance: when geometric distortions like scaling or rotation get applied on the video, the set of coefficients will be different if DWT, DCT, or DFT is used, but with DTCWT, the set of coefficients will remain the same after distortion. This property is necessary when designing a robust video watermark.
2. Good directional selectivity: DWT provides only three directions 0°, 45°, and 90°. While DT-CWT provides
3. six sub-images (three sub-image for each tree), these sub-bands containing complex values resulted from applying positive and negative frequencies of the filters so six sub-bands provide six directions $\pm15^o$, $\pm45^o$, $\pm75^o$ as shown in Fig 2.
4. Limited redundancy: independent number of scales DT-CWT that provides redundancy 2:1 in 1D and 4:1 in 2D.
5. Low computational complexity comparing to other transformations.
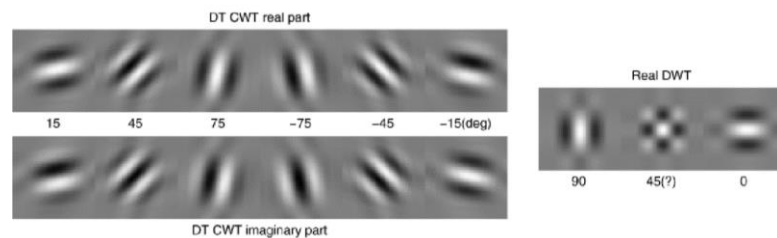6. Better reconstruction

FIG. 2: DTCWT VS. DWT DIRECTIONS

Due to the above attractive features, the DTCWT has become a better embedding domain for implementing effective watermarking systems [1] [2] [21] [24] [25] [26]. To implement DTCWT, two real DWTs are used. The first tree generates the real coefficients of the complex wavelet, while the imaginary part of complex coefficients is generated by the second tree. Different sets of filters areused at each level of decomposition. The phase information of signals is represented by the imaginary coefficients and thus offered general structural information of the image [22] [27]. To recover the original signal, the inverse of two real DWTs is applied to invert the real and the imaginary parts of DTCWT to get two real signals that are finally averaged to build again the signal [2, 21, 23]. Fig 3 and 4 give the steps of implementation of DTCWT analysis and synthesis respectively. [21].
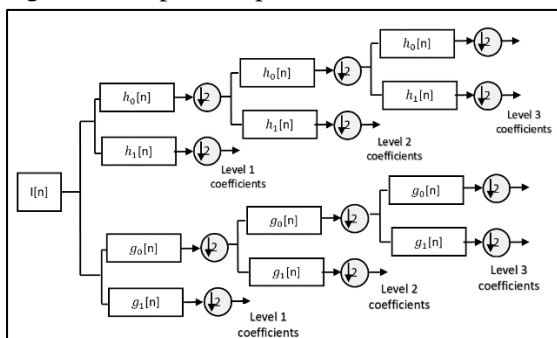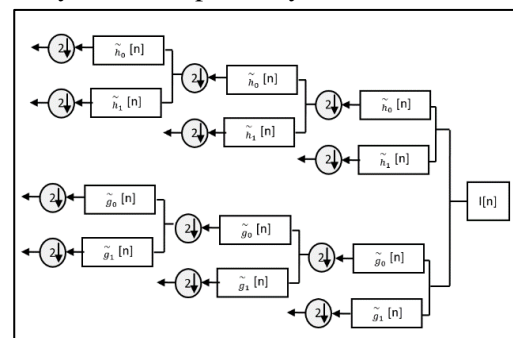

FIG. 3: ANALYSIS FILTER BANK FOR DTCWT

FIG.4: SYNTHESIS FILTER BANKS FOR DTCWT

Usually downsampling process causes aliasing effects, to minimize this effect the designer must use a suitable DT-CWT filter bank [24] as shown in the figure above, the first tree uses $h_0[n]$, $h_1[n]$ as the low and high pass filters while the second tree uses $g_0[n]$, $g_1[n]$ as the low pass and high pass filters.

DTCWT at each level of decomposition generates two complex low frequency sub-bands (LL_a and LL_b) and six high frequency complex sub-bands (HL_a, LH_a, HH_a, HL_b, LH_b and HH_b) that are reflecting the six directional filters at angles of ±15◦, ±45◦ and ±75◦ [2] [23] [27]. Some related works are presented in our previous paper [28].

IV. PROPOSED SYSTEMS

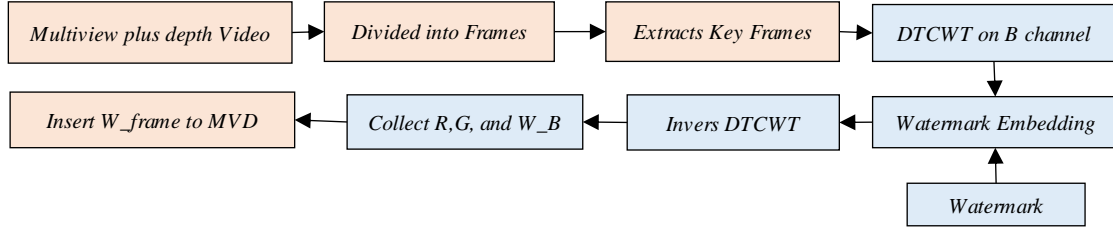The General Steps of the proposed system explained in Fig 5.

FIG 5: BLOCK DIAGRAM OF PROPOSED SYSTEM

A. Keyframe Extraction of 3D Video (3DVKE)

As known that the color of the image is one of the essential visual features. In the case of small camera motions, Color histograms based techniques are the best choice for keyframe extraction where it is easy to compute and robust. The Color histogram approaches depend on frame histogram differences where it's based on the fact that the two consecutive frames with unchangeable objects and the background will have little variation in their histograms. For a classical video, histogram-based keyframe extraction computes the difference between the histogram of the consecutive frames by applying Eq.1 where $F_{(I)}$ and $F_{(I+1)}$ are the first and second frames respectively. After that, the summation of the computed difference is computed according to Eq.2,

$$D\left(F_{(I)}, F_{(I+1)}\right) = \frac{\left|(hist_{(I)} - his_{(I+1)})^2\right|}{his_{(I+1)}} \qquad (1)$$

$$S(I) = \sum_{1}^{256} D\left(F_{(I)}, F_{(I+1)}\right) \qquad (2)$$

The computed difference consists of 256 values. After computing the summation values of all frames in the video, the Eq 3 normalized the summation of the video. A threshold value of the resulted normalized summations is computed by Eq.4. Any frame has normalized summations greater than the threshold value consider keyframe. [13] [17] [19]

$$Nor\_S(I) = \frac{S(I) - \min(S)}{\max(S) - \min(S)} \qquad (3)$$

$$T = mean((Nor\_S) \qquad (4)$$

In this stage of the proposed system, keyframes of MVV are extracted to take the important scenes from 3D Video. In the proposed 3DVKE system, the above histogram technique was applied in MVV to find the differences between the views, where there is no research before applying this technique for 3D video. At first, we are working on the three datasets, as explained in Table I below.

TABLE I: MULTIVIEW VIDEO DATASET

No	Dataset Name	No. camera	No. views per camera	Frame size
1	Breakdancers	8	100	768x1024
2	Ballet	8	100	768x1024
3	Soccer-Game-Goal	20	300	1080x1920

These datasets are downloaded from the Interactive Visual Media group at Microsoft Research and Fujii Laboratory, Department of Electrical Engineering and Computer Science, Nagoya University.

Each dataset used N cameras, each camera contains M views as shown in Fig 6. Four procedures are proposed to deal with the views of the cameras alone or mix between them in order to summarize the number of views. The first proposed algorithm of **3DVKE** tries to extract keyframes of MVV depending on Views, where k keyframes are extracted from M frames see Fig 7 and algorithm 1. While the second algorithm extracts the frames according to the camera as explained in algorithm 2, at the end of algorithm k keyframes are acquired of N Cameras as shown in Fig 8. The third algorithm takes all frames that are extracted from the algorithm 1 and 2.
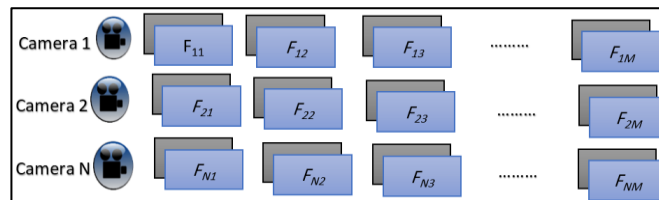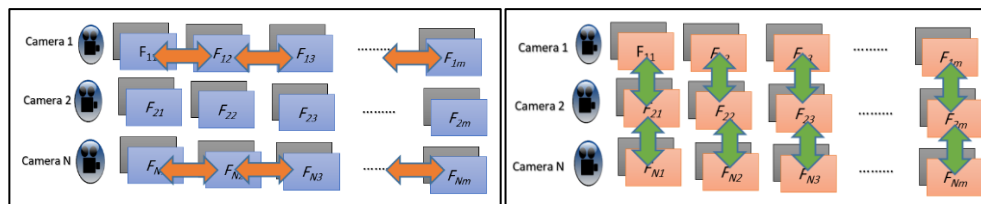

FIG. 6: MULTIVIEW VIDEO

FIG. 7: 3DVKE DEPEND ON VIEWS   FIG. 8: 3DVKE DEPEND ON CAMERA

Algorithm 1:

Algorithm Name: 3DVKE depend on Views
Input: Multiview Video
Output: Set of key frames

For each camera I from 1 to N do
  For each frame J from 1 to M-1 do
    Step 1: Convert RGB color frame (I, J) into grayscale color
    Step 2: Calculate the histogram of each pixel in the frame (I, J).
    Step 3: Calculate the differences $D(F_{(I, J)}, F_{(I, J+1)})$ between the pixels of each two adjust frames(I, J) and (I, J+1) according to the Eq. 1.
    Step 4: Find the summations S(I, J) of the pixel differences according to Eq. 2
  End For
End For

For I=1 to N do
  For J=1 to M-1 do
    Step 5: Normalization of summations Nor_S(I, J) according to Eq. 3.
    Step 6: Compute Adaptive threshold value of the camera frames T(I, :) depends on the normalized summations Nor_S(I, J) as follow Eq. 4:
    Step 7: frames J is key_frame if their Nor_S(I, J) greater than the threshold value T(I). "*k frame of M frames*"
  End for
End for

The final proposed algorithm works by extracting the frames that have high differences in their histogram comparing with adjusting frames belong to the same camera and next camera. The details of the final procedure explained in algorithm 3. After measuring the Compression ratio of

the proposed algorithms, algorithm 3 gives a high compression ratio compared with other proposed algorithms.

At the endof this stage, and after extracting a set of the keyframes from MVV, the frames address file must be created which consists of (cam$_{number}$,frame$_{number}$). This file sends to the end-user to help him in the verification process.

Algorithm 2:

Algorithm Name: 3DVKE depend on Camera
Input: Multiview Video
Output: Set of key frames

For each View J from 1 to M do
  For each camera I from 1 to N-1 do
    Step 1: Convert RGB color frame (I, J) into grayscale color
    Step 2: Calculate the histogram of each pixel in the frame (I, J).
    Step 3: Calculate the differences $D(F_{(I, J)}, F_{(I+1, J)})$ between the pixels of two frames of adjusting cameras according to the Eq. 1.
    Step 4: Find the summations S(I, J) of the pixel differences according to Eq. 2
  End for
  End for
For J from 1 to M do
  For I from 1 to N-1 do
    Step 5: Normalization of summations Nor_S(I, J) according to Eq. 3.
    Step 6: Compute the Adaptive threshold value T(:, J) of the frame depends on the normalized summations as follow Eq. 4
    Step 7: frame J is keyframe if their Nor_S(I, J) greater than T(J). "*k frame of N Cameras*"
  End for
  End for

Algorithm 3:

Algorithm Name: 3DVKE
Input: Multiview Video
Output: Set of key frames

  Step 1: Calculate the differences D1 between the pixels of each two adjust frames of each camera as Algorithm 1.
For I from 1 to N-1 do
 For J from 1 to M-1 do
   Step 2: Calculate the differences $D2(D1_{(I, J)}, D1_{(I+1, J)})$ between the differences that are the output from step1 by applying Eq.1
   Step 3: Find the summations S2(I, J) of the D2 according to Eq. 2
  End for
 End For
For I from 1 to N-1 do
 For J from 1 to M-1 do
   Step 4: Normalization of summations Nor_S2(I, J) according to Eq. 3:
   Step 5: Compute the Adaptive threshold value T(:, J) depends on the Nor_S2 as Eq. 4:
   Step 6: The frame that has Nor_S2 greater than the T(J) is a keyframe.
  End for
 End for

B. Three Dimension Video Copyright and Authentication Protection (3DVCAP)

The general watermark system was explained in Fig 9. The key frames from the previous stage are watermarked by the proposed system that consists of the following parts:

- Input: The input of the system is a set of key frames.

- Watermark: Two watermarks are tested in 3DVCAP. Watermark 1 (W1): is a binary logo with size 32 x 74=2368 bits as shown in Fig 10-a. Watermark 2 (W2): is also a binary logo with size 32 x 48=1536 bits as shown in Fig 10-b.
- Embedding process: the Proposed Embedding process that uses Dual-Tree Complex Wavelet Transform in their work.
- Extraction process: Extraction process that aims to extract the embedded watermark without the need to the original frames.
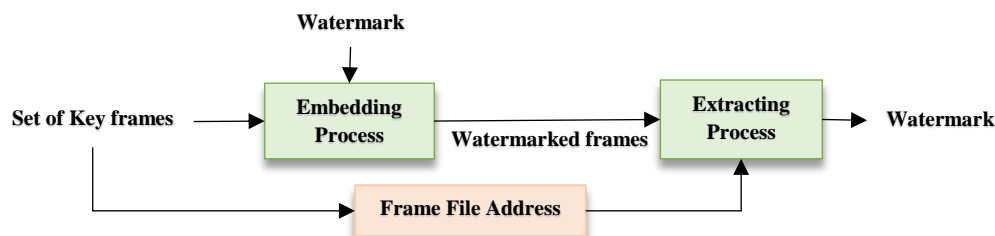

FIG. 9: WATERMARK SYSTEM IN GENERAL

(a)   (b)

FIG. 10: EMBEDDED WATERMARKS

The *embedding process* works as follows: The watermark is embedded in the B channel of RGB (Red, Green, and Blue) frame because HVS is less sensitive to the blue channel. According to watermark bits, Level 2 $\pm45°$ real and/or, imaginary high-frequency sub-bands of DTCWT are used for embedding. Each sub-band is divided into n (8x8) blocks. The total number of blocks must be equal to or greater than watermark bits. Also, the watermark bit is divided into subparts if needed depending on sub-band blocks. Each watermark bit is embedded into an isolated block. Watermark not embedded in the blocks sequentially, instead, it embedded dynamically. Therefore, in the proposed 3DVCAP, the Zigzag scan method is used for rearranging the blocks.

After rearranging the blocks, the *W* bit is replaced with one of the pixels of (8x8) block, this pixel called "victim pixel". The victim pixel varying from one block to another. The process of determining the pixel in one block done as follow proposed equations: Initially, V_{row} and $V_{column} =1$

$$V_{row} = [(V_{row} * \text{PB} + c) \, mod \, 8] + 1 \qquad (5)$$

$$V_{column} = [(V_{column} * \text{NB} + c) \, mod \, 8] + 1 \qquad (6)$$

V_{row} and V_{column} are the row and column number respectively, PB is the number of the previous block, NB is number of the next block, *c* is a counter begins with 1 and ends to the number of the keyframes. Modular 8 is used to restrict the position of victim pixel in the 8x8 block. As shown the victim pixel is changed from one block to another. This means that positions of the W bits are not static and are changed dynamically. This process was used firstly in this paper. The IDTCWT is applied for watermarked sub-bands and the R, G, watermarked B channels are contacted together again explained in algorithms 4.

Algorithm 4:

Algorithm Name: 3DVCAP Embedding **Input:** Set of key frames and watermark W **Output:** Watermarked keyframes

Start
For m =1 to number of keyframes
 Step 1: Apply level 2 DTCWT on B channel
 Step 2: Select $\pm45°$ sub-bands for embedding
 Step 3: Determine Max and Min Values in each sub-band
 Step 4: Divide sub-band into N (8x8) Blocks
 Step 5: Read N-Blocks in Zigzag Fashion
 Step 6: For i=1 to length (W)
  • Determine the index of Victim pixel V in each block for replacement depending on Eq.5 and Eq.6.
  • Read watermark bits
   o If W(i)=1
    ▪ Replace pixel $V_{(row,\ column)}$ with Max
   o Else
    ▪ Replace pixel $V_{(row,\ column)}$ with Min
 Step 7: End for i
 Step 8: Apply inverse DTCWT
 Step 9: Collect R, G, and watermarked B
End for
End Algorithm

On the other away side for the *verification process*. B channel of the watermarked keyframes has taken according to the frame address file that determines the watermarked frame number with its associated camera number from authenticated MVV. DTCWT is applied, the sub-bands are also divided into blocks, then the blocks are read by Zigzag method, finally, the victim pixels are determined for an extraction watermark bit like the same way that is indicated at the encoder as explained in algorithms 5. 3DVCAP is a blind system, the watermark can be extracted without need to the original video.

Algorithm 5:

Algorithm Name: 3DVCAP Extraction **Input:** Authenticated MVV and frame address file **Output:** watermark bits

Start
Determine keyframes depending on frame address file
For m=1 to number of keyframes
Step 1: Apply level 2 DTCWT on watermarked B channel of the keyframe(m)
Step 3: Select $\pm45°$ sub-bands for embedding
Step 4: Compute Average Value in each sub-band
Step 5: Divide sub-band into N (8x8) Blocks
Step 6: Read N-Blocks in Zigzag Fashion
Step 7: For i=1 to end of Blocks
 • Determine the index of Victim pixel V in each block for extracting
  o If value $V_{(row,column)}$>avg
   ▪ **W**(i)=1
  o Else
   ▪ **W**(i)=0
Step 8: End For i
 Step 9: Return **W** as a watermark bits from the keyframe(m)
End for
End Algorithm

V. Experiment Results

As mentioned previously, three datasets are used for testing the **3DVKE** and **3DVCAP** systems. For 3DVKE, compressions ratios are the measure used for comparing all proposed algorithms as

explained in Table II. You can notice that the fourth method explained in algorithm 3 gives a high compression ratio for all datasets.

TABLE II: COMPRESSION RATIO OF PROPOSED SUMMARIZATION METHODS.

Method No.	Data Set No.	Total framers No.	Summarized frames No.	Compression Ratio
Method 1	Dataset1	800	269	2.70
	Dataset2	700	272	2.57
	Dataset3	6000	3459	1.73
Method 2	Dataset1	800	360	2.22
	Dataset2	700	289	2.42
	Dataset3	6000	3021	1.98
Method 3	Dataset1	800	515	1.55
	Dataset2	700	449	1.55
	Dataset3	6000	4728	1.26
Method 4	Dataset1	800	195	4.10
	Dataset2	700	188	3.72
	Dataset3	6000	1067	5.62

For example, the breakdancer dataset contains 8 (0 to 7) cameras with 100 frames per camera. If algorithm 3 is used for key frame extraction, the result will be as follows: the first frame of camera 0 and camera 5 is keyframes, the second frame of camera 3 is keyframe also. Fig 11 shows examples of some selected frames

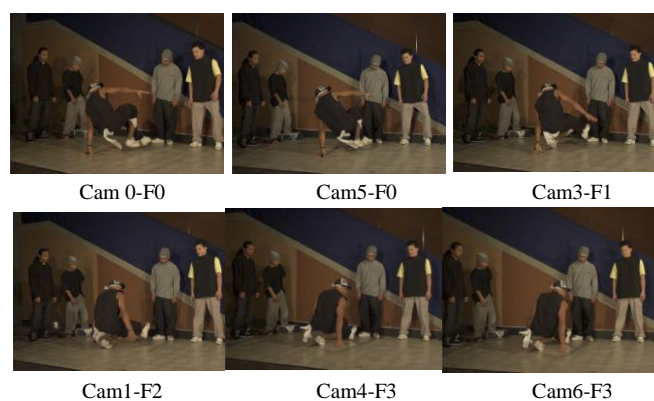| Cam 0-F0 | Cam5-F0 | Cam3-F1 |
| Cam1-F2 | Cam4-F3 | Cam6-F3 |

FIG 11: EXAMPLES OF EXTRACTED FRAMES

To test the proposed **3DVCAP** system Breakdancers and Ballet datasets are used, the frame size is 768x1024 pixels. In the case of using W1 for copyrights and authentication protection, W1 size is 2368 bits, this means, 2368 blocks are needed. While Level 2 sub-bands size is 192 x 256=49152/64=768 blocks. Therefore, we need to use $\pm45°$ real and imaginary sub-bands to get 3072 blocks enough to W1 bits. W1 bits are divided into four parts (w_1, w_2, w_3, and w_4), each w embedded into sub-bands. In another Case 2 when W2 is embedded in the frame, the W2 size is 32 x 48=1536 bit, this means we need 1536 blocks. The number of 8 x8 blocks of level 2 sub-band is 768 *2=1536 blocks exactly equal the size of W2, therefore W2 is embedded in $\pm45°$ real only sub-bands. W2 bits are divided into two parts (w_1, w_2), each w embedded into sub-bands. Fig 12 illustrates embedding and extracting W1 and W2. Many quality measurements that used to measure the quality 3DVCAP system such as Peak Signal Noise Ratio (PSNR in dB), Mean square Error (MSE), and structural similarity (SSIM), Normalize Correlation (NC), Structural Content (SC), and Normalized Cross-Correlation (NCC).

FIG. 12: EMBEDDING AND EXTRACTING W1 AND W2

The important features of any watermark system are achieved by the proposed **3DVCAP** system like *Imperceptibility, Blind Detection, payload and Robustness to Attacks*. Invisible watermark inserted to the extracted frames with keeping the quality of each frame where the value of PSNR influences between (61.5-63.5) for the different number of frames as well as the MSE values are [0.03-0.05], and the SIMM values are [0.9994-0.9997] as shown in Fig 13. Watermark bits are extracted without need the original frames, therefore the **3DVCAP** system is a blind system. Different watermark with the different number of bits is embedded successfully in summarized frame with no effect on the quality of the frames. Table IV shows the Quality measures of Frame without any type of attack. W1 and W2 extracted without error with NCC, SSIM, and SC equal to 1.

TABLE IV: QUALITY MEASUREMENTS WITHOUT ATTACK

Measures Name	Quality measures of Frame after embed W1	Quality measures of Frame after embed W2
MSE	0.037838	0.34247
PSNR	62.3515	62.7846
NCC	0.99971	0.99972
SC	0.99972	0.99973
SSIM	0.99626	0.99677
Measures Name	Quality measures of extracting W1	Quality measures of extracting W2
NCC_W	1	1
SSIM	1	1
SC	1	1

Some of spatial and geometry attacks are applied on the watermarked frames, Applied Spatial attack includes (salt & pepper 0.03, Poisson, speckle, Gaussian, Gamma Correction, Adjustment, Histogram Equalization, and Gaussian low pass filter). While Geometry attack includes (crop25%", crop50%, JPEG, JPEG 2000 compression). Some combination attacks are also applied to the watermarked frames like (crop25% with Poisson noise, crop50% with Gamma Correction, and JPEG" with Gaussian). Quality measurements are also computed for the frame after implementing these attacks. Table V shows the results of PSNR, MSE, SSIM, NCC, and SC of the watermarked frame after an attack.
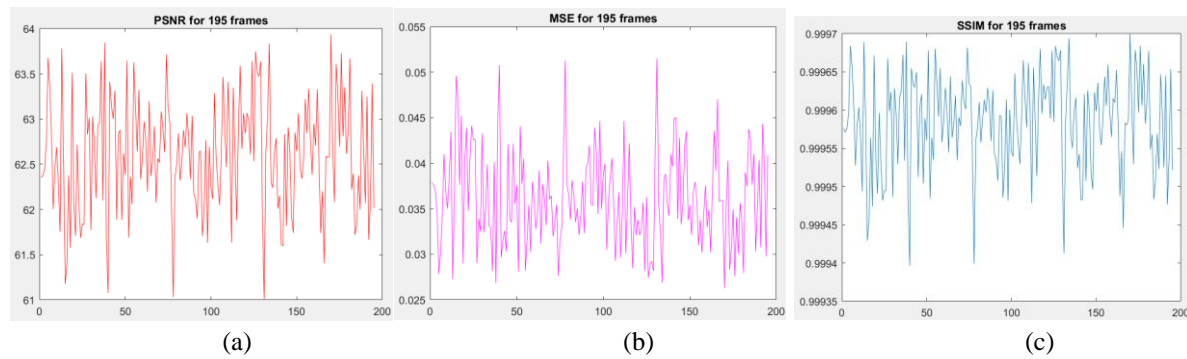
FIG. 13: TESTS FOR 195 SUMMARIZED FRAME A) PSNR   B) MSE   C) SSIM

TABLE V: QUALITY MEASUREMENTS OF WATERMARKED FRAME WITH DIFFERENT TYPE OF ATTACKS

Quality measure	MSE		PSNR		NCC		SC		SSIM	
Attack	**W1**	**W2**	**W1**	**W2**	**W1**	**W2**	**W1**	**W2**	**W1**	**W2**
Salt & pepper	312.8204	305.3113	23.1779	23.2834	0.85663	0.85582	0.89003	0.89278	0.4388	0.44223
Poisson	24.3229	24.2958	34.2706	34.2755	0.94768	0.94691	0.99296	0.99235	0.8594	0.85969
speckle	90.2283	90.2637	28.5774	28.5757	0.89105	0.8901	0.97797	0.97807	0.71941	0.71916
Gaussian	14.5569	14.6109	36.5001	36.484	0.96175	0.96176	0.99499	0.99513	0.88019	0.8801
Adjustment	250.4016	250.4109	24.1444	24.1443	1	1	0.63953	0.63952	0.95685	0.95685
Gamma Correction	730.3192	730.2981	19.4957	19.4958	0.81017	0.8102	2.7388	2.7388	0.68012	0.68017
Histogram Equalization	7913.721	7913.8131	9.147	9.1469	1	1	0.17627	0.17627	0.45751	0.45761
Gaussian low pass filter	13.3985	13.3961	36.8603	36.861	0.97035	0.97041	1.0029	1.0028	0.95622	0.95647
crop25%	386.487	386.4562	22.2595	22.2598	0.93266	0.93268	1.1074	1.1074	0.75057	0.75057
crop50%	2627.028	2627.0211	13.9362	13.9362	0.26494	0.26498	2.9341	2.9341	0.49998	0.49999
JPEG	2.0499	2.0497	45.0136	45.0139	0.98625	0.98614	1.0001	1.0001	0.97956	0.97994
JPEG 2000	0.34986	0.34947	52.6919	52.6967	0.99628	0.99637	0.99694	0.99699	0.99717	0.99716
crop25%+noise	406.8257	406.7795	22.0367	22.0372	0.93105	0.93112	1.1001	1.0994	0.64197	0.64229
crop50%+Gama	2959.516	2959.5371	13.4186	13.4186	0.34288	0.34284	10.2701	10.270	0.31682	0.31687
JPEG+Gaussian	17.8628	17.9165	35.6113	35.5983	0.95857	0.95865	0.99429	0.9944	0.91646	0.91654

The watermarks are extracted after all attacks with NCC values Ranging from 1 to 0.8, but the NCC decreased to 0.7 with geometry attack especially crop and crop with noise. Fig 14 explains the quality measurements NCC, SSIM, and SC for the extracted watermark. The forms of the extracted W2 under attacks are shown in Fig 15.
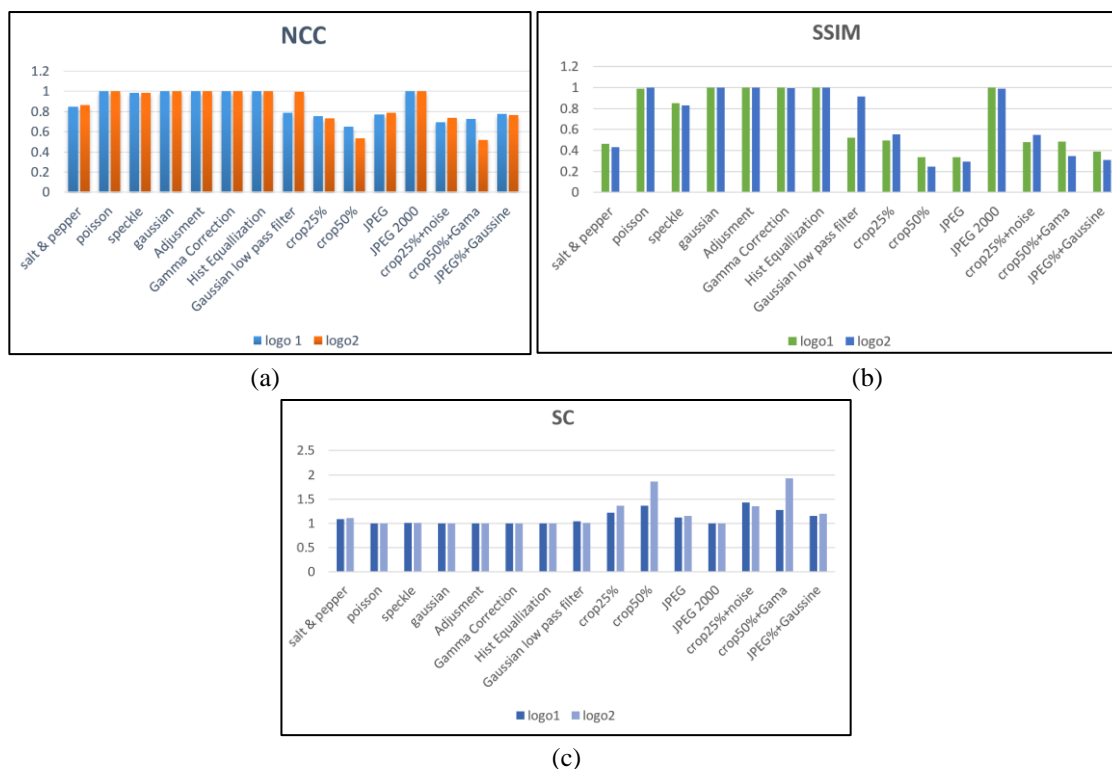
FIG. 14: EXTRACTED WATERMARK QUALITY MEASURES (A) NCC   (B) SSIM   (C) SC

VI.   CONCLUSION

Multimedia data, particularly the 3D video has been emerged in a wide range of applications such as medicine, education, television, cultures, etc. With these evolutions, the danger of threats is increased, where any user can access 3D video on the network, copying it, manipulating it, or redistributing it illegally. Therefore, it is necessary to propose new techniques to ensure the authentication and copyrights of the 3D video. Watermark techniques are the best choice for achieving these security requirements. The huge amount of 3D video data makes using a classical watermarking technique not suitable, thus the researcher must think of methods for reducing 3D video data. In this research, four algorithms are proposed for extracting the keyframes of MVV, these algorithms summarize the data according to histogram difference. The histogram is the most common method that is easy to implement and requires less computation complexity. 3DVKE algorithm has a high compression ratio where it varies [4:1 to 5:1]. The proposed watermark system 3DVCAP enables the users to embed invisible and robust watermark in the keyframes, where 3DVCAP uses modified embedding algorithm that embeds watermark bits in DTCWT coefficient blocks in dynamic way where the coefficients blocks are read in zigzag fashion, this means the watermark bits are not placed in sequential blocks such as the first bit embedded in the first block and the second bit in the second block etc. Positions of the embedding watermark bits inside the blocks are not static,however they are changing continuously according to the suggested equations, and thus an attacker cannot determine or guess the watermark bit locations. The proposed 3DVCAP system could extract the watermark after applying almost spatial, geometrical, and combination attacks with NCC near to 1. The payload of watermark has not effecte on the quality of the frames where the imperceptibility remains high after embedding different watermark lengths and where the PSNR values are between 62 to 64 db.This belongs to using DTCWT that provides us with six to twelve high-frequency sub-band coefficients. 3DVCAP is a blind system where the original frames aren't required at the extraction process, only the watermarked frames and frame address file are required.
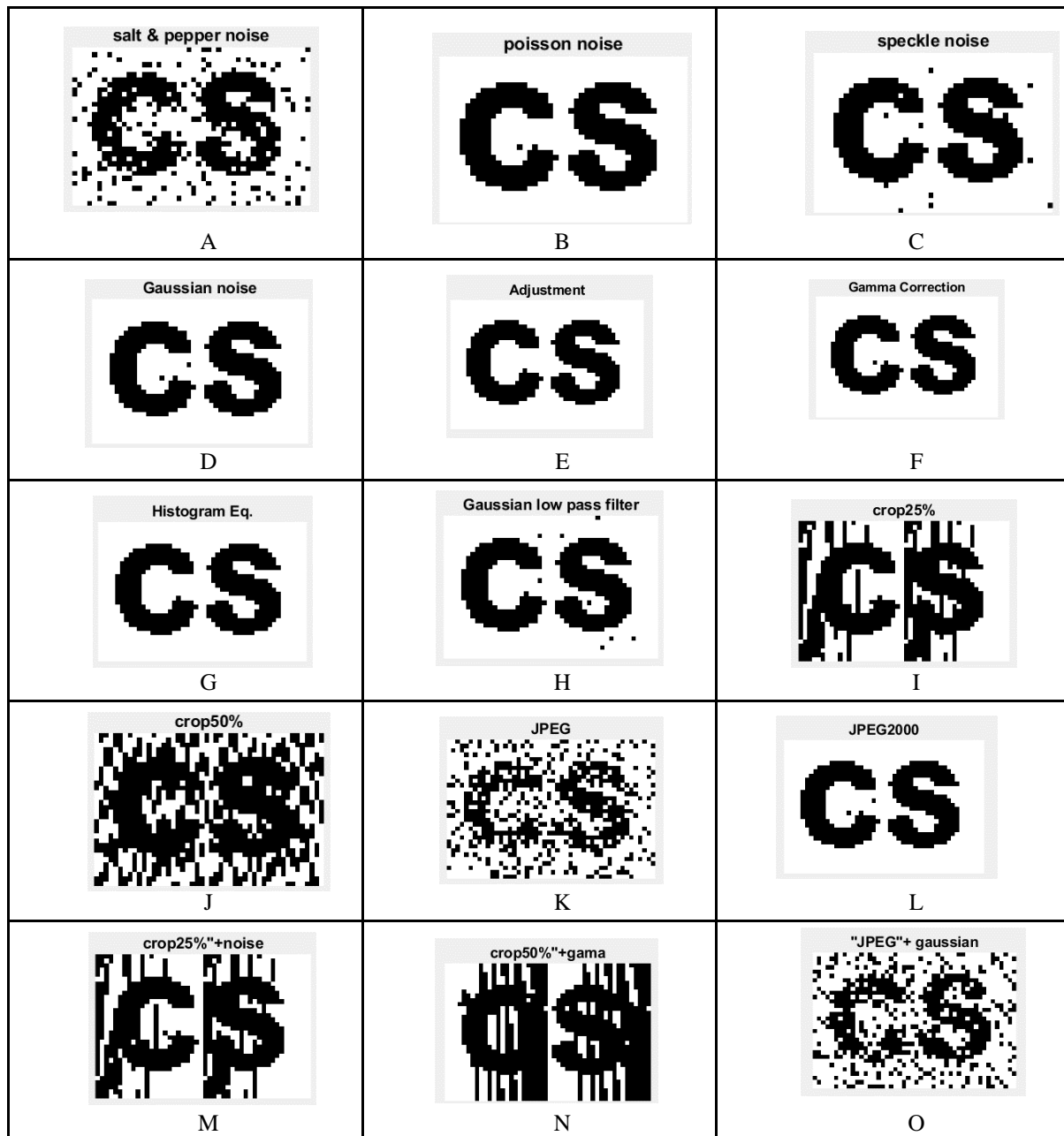
FIG. 15: FORMS OF THE EXTRACTED W2 UNDER ATTACKS

REFERENCE

[1] S. Mabtoul, E. I. Elhaj, and D. Aboutajdine, "Robust semi-blind digital image watermarking technique in dt-cwt domain," *International Journal of Computer Science,* vol. 4, pp. 8-12, 2009.

[2] K. Zebbiche, F. Khelifi, and K. Loukhaoukha, "Robust additive watermarking in the DTCWT domain based on perceptual masking," *Multimedia Tools and Applications,* vol. 77, pp. 21281-21304, 2018.

[3] M. Asikuzzaman and M. R. Pickering, "An overview of digital video watermarking," *IEEE Transactions on Circuits and Systems for Video Technology,* 2017.

[4] M. Asikuzzaman, M. J. Alam, A. J. Lambert, and M. R. Pickering, "Imperceptible and robust blind video watermarking using chrominance embedding: a set of approaches in the DT CWT domain," *IEEE transactions on Information Forensics and Security,* vol. 9, pp. 1502-1517, 2014.

[5] D. Dhaou, S. B. Jabra, and E. Zagrouba, "An Efficient Group of Pictures Decomposition based Watermarking for Anaglyph 3D Video," 2018.

[6] S. Rana and A. Sur, "3D video watermarking using DT-DWT to resist synthesis view attack," in *Signal Processing Conference (EUSIPCO), 2015 23rd European*, 2015, pp. 46-50.

[7] W. El-Shafai, E.-S. M. El-Rabaie, M. El-Halawany, and F. E. A. El-Samie, "Efficient multi-level security for robust 3D color-plus-depth HEVC," *Multimedia Tools and Applications,* pp. 1-27, 2018.

[8] J. W. Salih, S. H. Abid, and T. M. Hasan, "Imperceptible 3D Video Watermarking Technique Based on Scene Change Detection," *International Journal of Advanced Science and Technology,* vol. 82, pp. 11-22, 2015.

[9] Y. Shi, H. Yang, M. Gong, X. Liu, and Y. Xia, "A fast and robust key frame extraction method for video copyright protection," *Journal of Electrical and Computer Engineering,* vol. 2017, 2017.

[10] A. D. Salman and H. B. Abdulwahab, "Overview: 3D Video from capture to Display," in *2018 1st Annual International Conference on Information and Sciences (AiCIS)*, 2018, pp. 259-268.

[11] L. Ferreira, L. A. da Silva Cruz, and P. Assuncao, "Towards key-frame extraction methods for 3D video: a review," *EURASIP Journal on Image and Video Processing,* vol. 2016, p. 28, 2016.

[12] A. Kumthekar and J. Patil, "Key frame extraction using color histogram method," *International Journal of Scientific Research Engineering & Technology (IJSRET),* pp. 207-214, 2013.

[13] A. Nasreen and G. Shobha, "Key frame extraction from videos-A survey," *International Journal of Computer Science & Communication Networks,* vol. 3, p. 194, 2013.

[14] S. Pandey, P. Dwivedy, S. Meena, and A. Potnis, "A survey on key frame extraction methods of a MPEG video," in *Computing, Communication and Automation (ICCCA), 2017 International Conference on*, 2017, pp. 1192-1196.

[15] J. D. Gibson, "H A N D B O O K O F IMAGE AND VIDEO PROCESSING," *Academic Press Series in Communications, Networking, and Multimedia,* 2000.

[16] A. Ioannidis, V. Chasanis, and A. Likas, "Weighted multi-view key-frame extraction," *Pattern Recognition Letters,* vol. 72, pp. 52-61, 2016.

[17] P. M. S. V. Kavita Sahu, "Key Frame Extraction From Video Sequence : A Survey," *International Research Journal of Engineering and Technology (IRJET),* vol. 04 May-2017 2017.

[18] M. A. Talab, S. N. H. S. Abdullah, and M. H. A. Razalan, "Edge direction matrixes-based local binary patterns descriptor for invariant pattern recognition," in *2013 International Conference on Soft Computing and Pattern Recognition (SoCPaR)*, 2013, pp. 13-18.

[19] J. S. Boreczky and L. A. Rowe, "Comparison of video shot boundary detection techniques," *Journal of Electronic Imaging,* vol. 5, pp. 122-129, 1996.

[20] P. Loo and N. Kingsbury, "Watermarking using complex wavelets with resistance to geometric distortion," in *2000 10th European Signal Processing Conference*, 2000, pp. 1-4.

[21] D. k. N. Y. Padma, P.Rakesh kumar, "Dual Tree Complex Wavelet Transform for Medical Image Fusion," *International Research Journal of Engineering and Technology (IRJET),* vol. 03 Sep 2016.

[22] A. Kushwaha, A. Khare, O. Prakash, J.-I. Song, and M. Jeon, "3D medical image fusion using dual tree complex wavelet transform," in *2015 International Conference on Control, Automation and Information Sciences (ICCAIS)*, 2015, pp. 251-256.

[23] D. S. Reddy, S. Varadarajan, and M. Giriprasad, "2D dual-tree complex wavelet transform based image analysis," *Contemporary Engineering Sciences,* vol. 5, pp. 127-136, 2012.

[24] S. U. Faruq, K. Ramanaiah, and K. Soundararajan, "ANovel ALGORITHM FOR IMAGE DENOISING USING DT-CWT," *Signal & Image Processing : An International Journal (SIPIJ)* vol. 8, No.3 2017.

[25] 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 transactions on information forensics and security,* vol. 3, pp. 466-474, 2008.

[26] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and computational harmonic analysis,* vol. 10, pp. 234-253, 2001.

[27] J. Yadav and K. Sehra, "Large Scale Dual Tree Complex Wavelet Transform based robust features in PCA and SVD subspace for digital image watermarking," *Procedia computer science,* vol. 132, pp. 863-872, 2018.

[28] H. B. A. w. Abeer D. Salman, "Study Analysis to New Trend for 3D Video Watermark," *REVISTA AUS,* 26-2 (2019) 2019.