International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V71I4P216

Volume 71 Issue 4, 171-189, April 2023
© 2023 Seventh Sense Research Group®

Original Article

An Improved Video Keyframe Detection Technique Leads to Video Authentication

B.S. Kapre[1], A. M. Rajurkar[2], D. S. Guru[3]

[1,2]*Department of Computer Science and Engineering, MGM's College of Engineering, Nanded, India*
[3]*Department of Studies in Computer Science, University of Mysore, Manasgangotri, Mysore, India.*

[1]*Corresponding Author : kapre_bs@mgmcen.ac.in*

Received: 09 December 2022 Revised: 06 February 2023 Accepted: 12 March 2023 Published: 25 April 2023

Abstract - *Codec standard H.264/AVC compression has recently become more popular. Secure sharing of these videos is needed in applications like surveillance, remote sensing, telemedicine, and medical videos. Many times videos get edited, copied, and illegally distributed during transmission over public networks. Therefore, research on video watermarking based on this video compression standard has become very popular. In this research, a blind keyframe-based H.264/AVC video authentication technique is proposed. Embedding a watermark in each video frame increases the size of the video, and as well as it can be easily extracted by the attacker; adding a watermark in all frames is a time-consuming process. To avoid these problems, we propose a novel shot boundary detection algorithm by employing Karl-Pearson Correlation Coefficient (K-PCC) and Absolute Mean Difference (AMD), and SVD is used for the selection of keyframes. The proposed technique uses DWT, DCT, and Arnold Transform to generate a Content-Based Watermark (CBW) for each keyframe, enhancing security. These extracted keyframes are used for embedding and extraction of the watermarks. Further, DCT based embedding technique is used for embedding the watermark. The experimental findings demonstrate that the proposed method outperforms in terms of imperceptibility and robustness and can successfully detect shot boundaries under various camera actions. Also, the summarized video is effectively generated using the presented SVD-based key frame selection technique.*

Keywords - *K-Pearson Correlation Coefficient, Absolute Mean Difference, Video watermarking, Keyframe, DCT.*

1. Introduction

Digital media is a growing field that plays an immense role in almost everything from business to personal life. Social platforms widen the internet world, and the exchange of multimedia data such as images, PDFs, documents, videos, and audio is growing tremendously. Now, the webcast and video-on-request (VOD) administrations have started to rise and spread quickly throughout the world. The advancements in film, web series, YouTube videos, online learning and TV show are blooming. This is leading to an increase in multimedia content usage by unauthorized users. Therefore, data security has become the main concern nowadays. It is required to control unwanted data distribution and protect multimedia data from unauthorized access and misuse. The effective and leading solution to address all these challenges is digital watermarking [1,2,3]. It is used to identify unauthorized access and modifications in digital content. The secret information is embedded into digital content to ensure its security. It has got applications in copyright protection, copy control, fingerprint identification, content authentication, broadcast monitoring, video tracking, etc. Consequently, video watermarking has become an important research area.

The digital watermarking technique has been extensively researched in recent years and has many applications for protecting multimedia content such as images, videos and audio [2,3,4]. These methods involve adding watermark data to the protected multimedia content. The embedded secret information (watermark) in the watermarking system must be undetectable, observable, and resilient to many types of attacks, such as blurring, scaling, cropping, compression, noise, etc. Many researchers have presented different watermarking techniques by considering this requirement [7,8,9]. However, robust and secure video watermarking techniques are still in great demand. All videos are transmitted in compressed form for distributing and storing digital content over the internet. Therefore compressed domain-based video watermarking techniques are more applicable. As a result of this, compressed domain video watermarking has become a hot area of research [10,11]. Compressed videos are available in encoded versions like H.264, H.265, XVID etc. The most recent video coding standard created by the ISO/IEC Moving Pictures Experts Group and the ITU-T Video Coding Experts Group is H.264. The H.264/AVC video compression has gained popularity due to its improved network compatibility and higher compression efficiency. As a result, H.264/AVC-based video watermarking research is getting increasingly active.

Initially, image watermarking techniques were applied to videos directly. However, time complexity and redundancy were increased in these systems, and they failed to solve the new challenges caused by the temporal dimension of video sequences. It has been observed that the video quality degrades if image watermarking techniques are directly applied to it [12]. Hence, the two major challenges that must be considered while designing the video watermarking scheme are: i) keyframe selection and ii) video quality. In the keyframe-based watermarking approach, the watermark coefficients are inserted into the selected keyframes instated of embedding them into all video frames. This efficiently reduces the time complexity and helps improve the watermarked image's visual quality. The work proposed in this paper is inspired by the requirement of a practical video watermarking scheme that provides authentication to H.264/AVC-based compressed video with high imperceptibility and robustness that protects copyright property efficiently. The proposed video watermarking scheme provides several advantages over existing techniques.

The remainder of this paper is organized as follows. Related work is included in Section 2. Section 3 describes the workflow of the proposed methodology's workflow, and Section 4 shows the results of the experiments. The conclusions are summarized in the section.

2. Related Work

With the rapid development of internet technology, digital videos can be easily transmitted, accessed, modified and tampered with by unauthorized users. All videos are transmitted via the internet in compressed format because of their significant bandwidth and storage requirements. So, video watermarking techniques for compressed videos are becoming more applicable. Different types of video watermarking techniques exist for video compression, such as MPEG-4, MPEG-7, and H.264/AVC. Among these, the most widely used and popular compression coding standard at the present time is H.264/AVC. It provides better quality and higher compression efficiency, which is helpful in digital web applications for uploading, downloading, sharing, and recording video information. Three general categories can be used to classify current video watermarking methods: i) frame-by-frame [12,13,14], ii) Region based[15,16,17] and iii) keyframe based[18,19,20-36]. In the frame-by-frame technique, watermark information is inserted into every input video frame. These types of methodologies are very effective and robust against different frame attacks, such as frame dropping, inserting, and swapping. But these techniques suffer from drawbacks such as embedding a watermark in each frame of the video, increasing the size of the video. The algorithms are time-consuming, impractical and cannot resist frame averaging attacks. Many researchers have used a region-based approach for embedding and extracting watermarks to overcome these disadvantages. In these types of approaches, proper regions or moving blocks are selected within a host frame for embedding a watermark. It is revealed that these algorithms provide security, high imperceptibility and robustness against common attacks. The main constraint is that the accuracy of the watermark throughout the extraction process depends on the recovered locations of moving parts or regions. To improve the correctness and to reduce time complexity, the third approach was introduced wherein the representative frames are selected from each shot or scene of the video sequence for embedding and extraction of the watermark. This methodology reduces the huge amount of time required for the watermarking process. Moreover, it avoids frame redundancy and makes the watermarking process more stable and resilient.

In [21], Chen Li et al. presented a semi-fragile video watermarking scheme for compressed domain videos wherein the numerical relationship among DCT non-zero coefficients was considered an authentication code. Initially, the frame number was converted into an 18-bits watermark sequence in this scheme. Then the generated code was embedded into a 4*4 sub-block containing at least three DCT non-zero coefficients. This watermarking technique shows good transparency and tamper detection. In [22] ShahadAlmuzairai et al. introduced a video watermarking system in which a watermark embedding algorithm was applied to audio and visual video streams. In the visual stream, moving blocks were detected, and a watermark was embedded in each moving block using DCT transform, whereas, in the audio stream silence detection algorithm was presented to create space by removing noise. The watermark was inserted in the DWT domain. This technique outperforms in terms of imperceptibility and robustness against geometric distortion and compression. However, it was sensitive to rotation and bilinear-curve attacks. In [23], an object-based video watermarking method was developed, in which a watermark was added to a homogeneous moving object within a video shot. The presented watermarking system is robust against geometric distortion. However, its main requirements are that the selected object must be small and highly textured to improve the watermarked image's visual quality.

In many video applications, shot boundary detection and keyframe extraction play an important role. The watermark was often embedded in the first frame of each scene in key-frame-based video watermarking [24]. Each scene's first and ending frames were chosen as significant frames in [25, 26-27]. Then, in [28], a frame selection strategy based on the object and background information was given. The frame with the highest object-to-background ratio was picked as the keyframe for the present scenario. The selected keyframe has the most information about the scene.

A shot segmentation and block classification-based video watermarking wherein a host frame was detected from each shot for inserting a watermark was introduced in [20]. Initially, the original watermark was divided into small parts based on the number of shots in the input video. Watermark parts were respectively embedded into shots. Further, a host coefficients selection scheme was introduced by adopting block classification in the discrete cosine transformation (DCT) domain. The watermark was inserted using Quantization Index Modulation (QIM). This technique improved the visual quality of watermarked video and provided robustness against Gaussian noise, frame swapping and MPEG compression. In [30] YassineHimeur et al. developed a chaotic encryption-based video watermarking technique, wherein the keyframes were extracted using the Gradient Magnitude Similarity Deviation (GMSD) technique for embedding and extraction of the watermark. A blind and secure watermark embedding and extraction technique was adopted using DWT and SVD. Then, before embedding, a chaotic encryption technique was used to encrypt the watermark. In terms of imperceptibility and resilience, this technique fulfills the requirements of watermarking techniques. However, this system was that if the watermarked key frame was lost from the video sequence, it failed to recover the watermark and provided weak resistance to geometric distortion.

In [31], a semi-blind Speed-Up Robust Features (SURF) features and visual cryptography-based video watermarking scheme. In which a histogram difference of consecutive frames is used for shot detection technique was introduced. Then, an optimal and robust keyframe was selected from each shot. Further, keyframe blocks are classified into two categories, edge blocks and smooth blocks, using a canny edge detection algorithm. According to the type of block, owners' shares were generated. In this technique, watermark embedding was not done; only the owner's shares were considered authentication information during extraction. This method is highly robust against different signal processing and geometric distortion; however, it requires storing shares securely. Soumik Das et al.[32] developed an improved keyframe extraction-based video watermarking scheme. In this method, an efficient keyframe extraction method was introduced using boundary luminosity analysis. The keyframes were extracted from each scene by calculating, comparing and analyzing the mean boundary luminosity of two consecutive frames. The DCT-based watermarking technique was applied on each keyframe to embed a scrambled binary watermark with a secret key. Then the watermark was embedded into seven low-frequency coefficients of DCT. In this scheme watermark extraction process was blind. This method provides high imperceptibility and robustness against compression and filter-based attacks.

Sethuraman et al.[33] have introduced a keyframe-based watermarking technique wherein the structural similarity index metric – absolute difference metric (SSIM-AMD) techniques were adopted to identify non-redundant frames. Then, the entropy–AMD method was used to select the keyframe. Further, DWT is applied to decompose the keyframe into sub-bands. To avoid false-positive attacks, the principal component of the watermark image block was computed and embedded into the middle band of DWT. The watermark image blocks were shuffled using the chaotic map technique to improve the watermarking algorithm's security level. The watermark strength was decided by calculating the scaling factor using the ant colony optimization (ACO) technique. It was observed that this scheme was robust against video processing and false-positive attacks. It provides high performance in terms of imperceptibility and robustness. Roopsingh et al.[34] presented a lossless video watermarking scheme in which the histogram difference technique was employed to detect color motion and motionless frames from the input video. Further, the keyframe identification was made by calculating and comparing the entropy of each motion frame with the average entropy value of motion frames. A scrambled watermark was embedded into each keyframe using linear wavelet transform and Hessenberg transform. Then intelligent gravitational search algorithm was used to extract keyframes efficiently from the input video. This presented video watermarking system was robust against different video processing attacks but provided low visual quality of the watermarked video.

To reduce time complexity, a Fibonacci sequence-based video watermarking scheme was developed [35]. The histogram difference was used to detect scene change in the video, and keyframes were selected using the Fibonacci sequence, which was generated using the seed key. In the embedding process, the watermark was first scrambled using Fibonacci-Lucas transform and converted into an encrypted form. The encrypted watermark was embedded using DWT and SVD into the LH sub-band. The presented technique was blind as the original information was not required during the extraction process. This technique enhances video authentication and provides better performance against various attacks. A secure video watermarking system was presented in [36] to solve the ownership problem. In which Graph-Based Transform (GBT) along with SVD and Hyper-chaotic encryption hybrid technique was proposed for watermark embedding. In this technique, grouping identical frames was done by finding histogram differences between two adjacent frames. Keyframe was selected from each scene for embedding and extraction of the encrypted watermark. The proposed technique was found to be robust, secure and imperceptible. However, this technique provides a low average PSNR of watermarked video. Since the watermark was embedded using hybrid combinations of Graph-based transform (GBT) and SVD, the visual quality of this technique was improved. This watermarking technique archives high PSNR and provides robustness against Gaussian noise, sharpening, rotation, blurring, and JPEG compression.

The literature review on video watermarking systems revealed that i) Embedding watermark in all video frames consumes more time. ii) Inserting identical watermark in the keyframes can be easily identified by frame dropping attack iii) Embedding the whole watermark in each frame suffers from the frame-dropping attack. Iv) The block of the watermark embedded in the selected frame can withstand frame-dropping attacks. However, these methods cannot achieve a good trade-off between robustness and imperceptibility.

The main contributions of this work are:

- To improve the efficiency and correctness of shot-to-boundary detection, we have used Karl-Pearson Correlation Coefficient (K-PCC) and Absolute Mean Difference (AMD) to detect shot boundaries.
- SVD-based keyframe extraction algorithm to detect the suitable keyframe for embedding watermark information. This embedding technique provides robustness against frame attacks and improves watermark authentication.
- Computational time analysis for the proposed technique with existing techniques.
- Content-based watermark is generated by selecting proper invariant features and Arnold transform. This shows the semi-fragile property and allows fulfilling the task of the difference between malicious and non-malicious attacks.
- To improve imperceptibility and robustness, DCT-based embedding and extraction techniques are used.
- The tampered area is identified by comparing generated and extracted watermark information.

3. Proposed Work

This section describes the proposed framework of the H.264 video watermarking system in the DCT domain. Initially, preprocessing is done to detect the number of shots using K-PCC and AMD. SVD transform is then used to extract the keyframe from each video shot. To improve the security level of the proposed method, CBW is generated by comparing DC coefficients of DCT by employing DWT and Arnold Transform. Further, the generated CBW is blindly embedded into the DCT domain's respective keyframe.

3.1. Preliminaries

This section provides a detailed review of YCbCr space color, K-PCC, AMD, SVD, DWT, DCT, and Arnold transform to better understand the details of the proposed watermarking scheme.

3.1.1. Color Space Conversion YCbCr color space [44].

The RGB color space consists of Red (R), Green (G) and Blue (B) components, and the YCbCr color spaces consist of luminance Y, Cb and Cr components. Y component indicates the majority of information and overall strength of the frame, while Cb and Cr represent the color information of the frame. YCbCr components are more suitable for image and video watermarking schemes because RGB components are more highly correlated than YCbCr components. The mathematical color transformation of RGB to YCbCr is performed using Eq. (1) andYCbCr to RGB using Eq. (2).

$$\begin{cases} Y = 0.299 \times R + 0.587 \times G + 0.114 \times B \\ C_b = -0.147 \times R - 0.289 \times G + 0.436 \times B \\ C_r = 0.615 \times R = 0.515 \times G + 0.100 \times B \end{cases} \quad (1)$$

$$\begin{cases} R = Y + 1.140 \times C_r \\ G = Y - 0.395 \times C_b - 0.581 \times C_r \\ B = Y + 2.032 \times B \end{cases} \quad (2)$$

3.1.2. Karl-Pearson correlation coefficient (K-PCC) [39]

K-PCC is a popular and widely used statistical measure to calculate the linear relationship between two data sets, P and Q, which is used in many applications such as statistical analysis, image registration, pattern reorganization, etc. The value of K-PCC ranges from -1 to +1. If the two images are absolutely identical, then their K-PCC value is close to +1. If the two images are anti-correlated, in that case, the K-PCC value is equal to -1, and if the two images are completely different, then the correlation between them is 0. The K-PCC calculates the exact degree of correlation between two images in determining whether there is any correlation between them.

K-PCC between two images, P and Q, is calculated in terms of covariance given as in Eq. (3).

$$\rho(P,Q) = \frac{1}{N-1} \times \frac{1}{M-1} \sum_{i=1}^{N} \sum_{j=1}^{M} \left(\frac{P(i,j)-\mu_P}{\sigma_P}\right)\left(\frac{Q(i,j)-\mu_Q}{\sigma_Q}\right) \quad (3)$$

Where,μ_P and σ_P are the mean and standard deviation of P, respectively, and μ_Q and $\sigma_Q$are the mean and standard deviation of Q, respectively.

3.1.3. Absolute Mean Difference [35]

The Absolute mean difference (imabsdiff) is used to measure the absolute difference matrix between two frames for identifying the redundant frames in the shots. It is performed by calculating the mean value of the absolute difference matrix (imabsdiff) between the seed frame$f_{seed}$and successive framesf_{suc_i}. The formula for the average AMD is calculated using Eq.(4)

$$AMD\left(f_{suc_i}, f_{seed}\right) = mean\left(imabsdiff\left(f_{seed} - f_{suc_i}\right)\right) \quad (4)$$

3.1.4. Singular Value Decomposition

Singular Value Decomposition is applied to image A of size M×N and decomposes it into three matrices such as U, S, and V [44]. Among these three matrices, U and V are orthogonal matrices of sizes M×M and N×N, respectively, and S is the diagonal matrix of size M×N. All non-zero elements of the S matrix are in a diagonal position and are called singular values. The singular values of the S matrix are arranged in descending order and provide robustness against attacks, higher invisibility and stability compared to U and V matrices. Therefore, the S matrix of SVD fulfills all requirements of watermarking. The attractive properties of SVD, like stability, robustness and maximum energy packing, are becoming more popular in watermarking research.

$$A = U \times S \times V' \quad (5)$$

3.1.5. Discrete Wavelet Transforms [35]

DWT is a procedure for decomposing an image and video frames into four frequency sub-band, such as LL is a low-frequency sub-band, HH is a high-frequency sub-band, and LH and HL are mid-frequency sub-bands. DWT is more famous because of its resilience against noise and compression attacks. The human visual system aspects are better modulated by DWT transformation than other transformations. As a result, this transformation was used for many applications like image restoration, zooming, compression and transformation [33, 35, 44]. Most important, it is also used in watermarking schemes because of its attractive properties like spatial localization, frequency spread and multi-resolution modeling. Fig. 1 (a) and (b) show the sub-bands obtained after first-level decomposition.

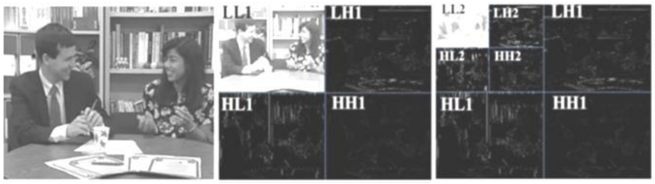


Fig. 1 (a) Original image (b) 1st level DWT decomposition c 2 level DWT decomposition

3.1.6. Discrete Cosign Transform [45]

Discrete cosign transform is employed on the whole image or by partitioning the image into blocks of a specific size. This transform is used to convert an image into low, mid and high-frequency bands. Fig. 2 presents the location of Low, mid, and high frequency with DC coefficient after employing 2D-DCT transform on an 8×8 block. The more common and popular transformation in digital signal processing is DCT. DCT has various applications, like image processing, compression, watermarking, and more [46]. Eq.(6) and Eq.(7) are used to represent 2D-DCT (7). Watermark embedding in low-frequency sub-band cause a visual distortion as human eyes are more sensitive to this frequency. If the high-frequency band is selected for embedding, it will likely be removed during compression. At the same time, the mid-frequency coefficients are more continent than low and high-frequency coefficients. Selecting embedding regions is one of the major concerns of the DCT technique.

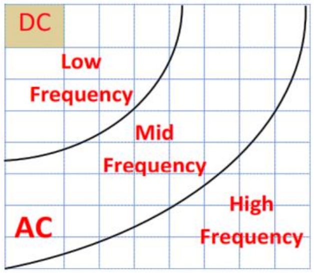


Fig. 2 Elements of the 2D-DCT process

$$F(u,v) = \alpha(u)\alpha(v)\sum_{i=0}^{M-1}\sum_{j=0}^{M-1} f(i,j)\cos\left[\frac{(2i+1)\mu\pi}{2M}\right]\cos\left[\frac{(2j+1)\mu\pi}{2M}\right] \quad (6)$$

Where u,v=0,1,2,3…….M-1, M is size of sequence f(i,j) is an image and F(u,v) is in frequency domain

$$\alpha(u) = \begin{cases} \frac{1}{\sqrt{M}}, u = 0 \\ \frac{2}{\sqrt{M}}, u \neq 0 \end{cases} \quad (7)$$

3.1.7. Arnold Transforms [44]

Arnold transform is used to scramble a pixel location of the original image. This method is iterative and reversible.

The number of iterations during the Arnold period depends on the size of the original image. The Arnold transforms aims to alter the semantics of the original image and transform it into an unreadable or scrambled form. The following Eq. (8) describes the Arnold Transform of a m x m image:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & a \\ b & ab+1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} mod(N) \quad (8)$$

Arnold transform is more useful in the watermarking field because of its scrambling technique. The secret information is encrypted using Arnold Transform to increase security and protect the confidentiality of the watermarking process.

The original secret information cannot be recovered without precise knowledge of the specific Arnold period K.

3.2. Shot Boundary Detection using K-PCC and AMD

In this section, the process of proposed shot boundary detection using K-PCC and AMD is described. In fact, the K-PCC and AMD can recognize shot boundaries in a video stream because they are resistant to changes in illumination, the movement of objects, and camera operations.

Initially, the video is decomposed into frames; the first frame of the video is considered as a seed frame for the first shot. Each video frame is further divided into Red, Green and Blue channels. At first, K-PCC is measured between the seed frame and successive frames for red, green and blue channels and stored in respective matrices named R_PCC, G_PCC and B_PCC, respectively. Next, AMD is used to measure the difference between the seed frame and successive frames for each channel and the mean of the difference matrix is measured for each R, G and B channel and stored in the matrices named RD, GD and BD. Suppose the Karl-Pearson Correlation Coefficient of two frames are greater than the threshold and the mean of difference value is less than the threshold. In that case, the corresponding successive frame is added to the current shot of the seed frame else. It is considered as a seed frame of the next shot. We must choose the appropriate threshold values to detect shots with high accuracy.

For each channel (red, green and blue), the mean and variance of R_PCC, G_PCC and B_PCC are taken as thresholds and computed using Eq 9,10 and 11.

$$ThR_{PC} = \mu_R + \alpha \times \sigma_R^2 \tag{9}$$

$$ThG_{PC} = \mu_G + \alpha \times \sigma_G^2 \tag{10}$$

$$ThB_{PC} = \mu_B + \alpha \times \sigma_B^2 \tag{11}$$

The process of threshold detection for AMD is calculated for each channel using the mean of RD, GD and BD as computed using eq 12,13 and 14:

$$Th_{RD} = \beta \times \mu_{RD} \tag{12}$$

$$Th_{GD} = \beta \times \mu_{GD} \tag{13}$$

$$Th_{BD} = \beta \times \mu_{BD} \tag{14}$$

For the thresholds, the parameters μ define the mean value,σ^2 indicates variance, and scaling factors like α and β are used to make the threshold work under all scenarios and range from 0 to 1. The values of α and β considered in this research are 1, which we found best in our experimentations.

In this work, for the experimentation, three different Datasets are used with different types of videos such as news, cartoon, educational, ephemeral, and historical and lectures etc. It is observed that the videos with a greater number of shots will have a greater number of keyframes. The proposed shot boundary detection algorithm is described in Algorithm 1

Algorithm 1: Shot boundary detection

```
Initially, the host video is preprocessed into a number of frames ($F_1, F_2, F_3, \ldots F_n$ )
k=1, j=1
$SF_i = F_i$   // At first, the first frame is selected as the seed frame
$Vshot_j(k) = F_i$   // Add seed frame as the first frame in the Ist shot.
For i ← 2: NumberofFrames
Decompose each frame $F_i$ into RGB channels, such as $F_i\_R,\ F_i\_G\ and\ F_i\_B$
// Calculate K-PCC between the j^th shot of  seedframe with successive frames
        $R\_PCC(i)_R$ ← PCC($SF_j$_R, $F_i\_R$)
        $G\_PCC(i)_G$ ← PCC($SF_j$_G, $F_i\_G$)
        $B\_ PC(i)_B$ ← PCC($SF_j$_R, $F_i\_B$)
        // Calculate the Absolute Mean Difference between the j^th shot of the seedframe with successive frames
        $RD(i)_R$ ← mean(AMD($SF_j$_R, $F_i\_R$))
        $GD(i)_G$ ←mean(AMD($SF_j$_G, $F_i\_G$))
        $BD(i)_B$ ←mean(AMD($SF_j$_R, $F_i\_B$))
IF    ($R_{PCC(i)_R} > ThR_{PC}$ && $G_{PCC(i)_R} > ThG_{PC}$ && $B_{PCC(i)_R} > ThB_{PC}$ && $RD(i)_R < Th_{RD}$ && $GD(i)_R > Th_{GD}$ &&  $BD(i)_R < Th_{BD}$
        // Add a successive frame in the j^th shot
            $Vshot_j(k) = F_i$
            k=k+1
  Else
            k=1, j=j+1
        // Consider successive frame as seedframe
            $SF_j = F_i$
        // Add seedframe as the first frame in the j^th shot.
            $Vshot_j(k) = SF_j$
            k=k+1
        Decompose SF into RGB channels, such as $SF_j$_R, $SF_j$_G and $SF_j$_B channels
    End
  End
```

3.3. Key–Frame Extraction using SVD

A robust keyframe extraction technique from each shot is proposed in this work. Watermark embedding and extraction are done in those keyframes only and not in all the frames. To find such a keyframe from each shot, we have proposed SVD based keyframe extraction technique. Initially, SVD is applied on each keyframe of the shot to obtain U, S and V matrices. In which S is the diagonal matrix and U and V are orthogonal matrices. Singular values of the S matrix are more stable, and a small number of singular values represent large signal energy information. In addition to this, they are also robust to rotation, scaling, compression, transposition and noise.

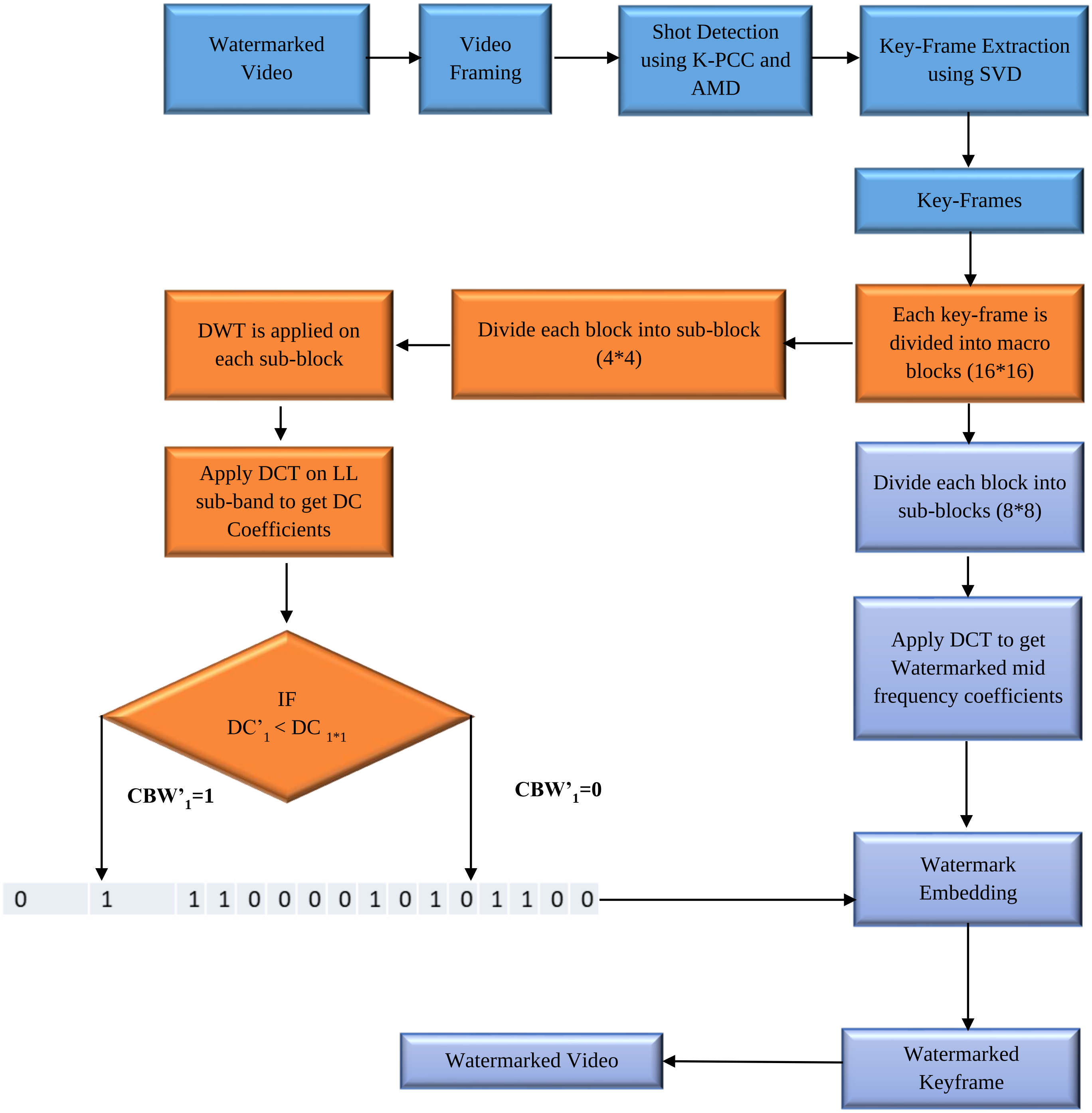

Watermarked Video
Video Framing
Shot Detection using K-PCC and AMD
Key-Frame Extraction using SVD
Key-Frames
Each key-frame is divided into macro blocks (16*16)
Divide each block into sub-block (4*4)
DWT is applied on each sub-block
Apply DCT on LL sub-band to get DC Coefficients
IF DC'$_1$ < DC $_{1*1}$
CBW'$_1$=1
CBW'$_1$=0
0 1 1 1 0 0 0 0 1 0 1 0 1 1 0 0
Divide each block into sub-blocks (8*8)
Apply DCT to get Watermarked mid frequency coefficients
Watermark Embedding
Watermarked Keyframe
Watermarked Video

Fig. 3 Watermarking embedding process

By considering all these features, we have used SVD for the extraction of the keyframe. Initially, each shot frame is separated into red, green and blue channels. SVD is then employed on each channel to obtain k-singular values. Each singular value is normalized, and then the entropy of k-singular values is measured for each channel of the frame and stored in the matrices R_Ent, G_Ent and B_Ent, respectively. Maximum of R_Ent, G_Ent and B_Ent is selected to get a single entropy value denoted by Max_RGB for a frame of the shot. The same process is repeated for each frame of the shot, and the calculated maximum values are stored into $Max_{RGB_{F_i}}$ array. After that, the maximum valued frame is detected as a keyframe for the current shot. The same procedure is applied on each shot to detect the keyframe. The proposed algorithm for keyframe extraction is presented in Algorithm 2.

3.4. Watermarking Process

The watermarking process is the process of embedding the watermark into the source video. The flow of the watermark embedding process is shown in Fig. 3. Initially video is preprocessed into a number of frames and grouped into various shots by applying KPCC and AMD techniques. Then, the keyframes are detected by employing SVD; furthermore, DWT-DCT-based Content Based Watermark (CBW) is generated for each keyframe by applying Arnold Transform and generated CBW is embedded into DCT mid-frequency coefficients. The watermarking process is explained briefly in the following section. The watermarking process is divided into two steps: i) Watermark embedding and ii) Watermark detection process

3.4.1. Watermark Embedding Process

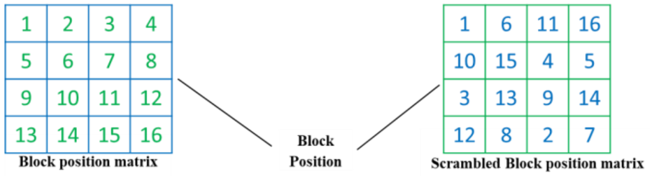
Initially, the keyframe is transformed from RGB color space to YCbCr color space because RGB color space is highly correlated and unsuitable for frequency domain watermarking such as DCT [19]. Y channel is called the luminance channel, whereas; the Cb and Cr channels are called blue chrominance and red chrominance, respectively. Although, the luminance component is more sensitive to the human Visual System than the chrominance components (Cb and Cr). We have used the keyframe's luminance (Y) channel for embedding the watermark. The reason behind this is that the JPEG compression discards a lot of chrominance information during chroma subsampling, and the watermark will not sustain against compression techniques if it is embedded in the chrominance part. The transformation from RGB color space to YCbCr color space is done using the following Eq. (1)

Content-based Watermark Generation

In the proposed authentication code generation, each keyframe's luminance (Y) channel is initially divided into non-overlapping blocks of size 4×4 and is scrambled using Arnold transform as shown in fig. 4. Then, the blocks of the original keyframe are rearranged according to the scrambled position matrix. The original keyframe and its scrambled block are shown in fig. 5. Each frame consists of macro-blocks (16×16) which may be divided into 4×4 sub-block partitions for motion prediction. Each 16×16 macro-block contains 16 sub-blocks of size 4×4.

Further, DWT is applied on each 4×4 sub-block to get four sub-bands LL, LH, HL and HH and DCT is employed on the LL band to get DC-coefficient. All 16-DC coefficients of each macro-block (16×16) are compared, and a 15-bit authentication code (*CBW*ᵢ) is generated using the following Eq (15).

$$CBW_i = \begin{cases} 1 & if\ DC_i < DC_{i+1} \\ 0 & else \end{cases} \tag{15}$$



1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Block position matrix

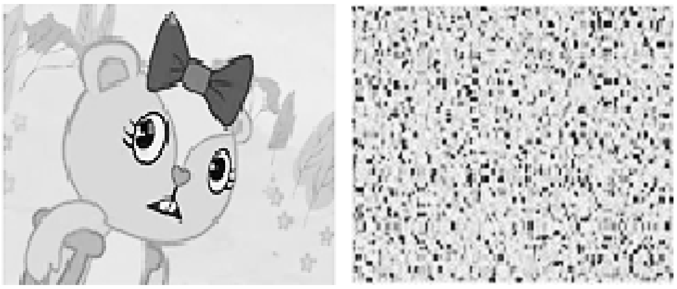
1	6	11	16
10	15	4	5
3	13	9	14
12	8	2	7

Scrambled Block position matrix

Fig. 4 Scrambled block positions

Algorithm2 :- Key-Frame Extraction

1. *Initially, the video is divided into a number of shots by using Algorithm 1*
2. *Each of the shots is separated into Red, Green and Blue channels.*
3. *SVD is employed on each channel to get three matrices as following Equation:*
   $A = USV^T$
   Where, U and V^T *are orthogonal matrices, and S is a diagonal matrix. The diagonal elements of the S matrix have the following property.*
   $$s_1 > s_2 > s_3 > s_4 \dots \dots \dots > s_N$$
4. *All the singular values* s_1, s_2, s_3, s_4 *,----,* s_i *, ------ ,*s_N *are normalized using following formula.*
   $$s_i = \frac{s_i}{\sum_{k=1}^{N} s_k}$$
5. *Entropy of the singular value is measured using the following formula.*
   $Ent_{F_i} = \sum_{k=1}^{N} s_k (\log\left(\frac{1}{s_k}\right))$
6. *All steps, step No2 to step No5, are repeated for each channel of the frame and store the calculated entropy value in R_Ent, G_Ent and B_Ent .arrays.*
7. *Maximum of R_Ent, G_Ent and B_Ent to get a single entropy value for a frame of a shot.* $\mathrm{Max}_{\mathrm{RGB}_{\mathrm{F_i}}} = max(R\_Ent, G\_Ent, B\_Ent)$
8. *Calculate entropy values for each frame of the shot by repeating steps No. 1 to step No. 5.*
9. *Finally, the maximum entropy-valued frame is detected as a keyframe of a current shot using the following eq.*
   $\mathrm{KF}_{\mathrm{S_I}} = \max\left(\mathrm{Max}_{\mathrm{RGB}_{\mathrm{F_i}}}\right)$
10. *Above all, steps are repeated for each shot of the video frame.*



Original Y-channel of key-frame **Scrambled block after applying Arnold Transform**

Fig. 5 Scrambled block keyframe

Watermark Embedding Process

In the proposed H.264/AVC video watermark embedding scheme, the generated CBW for each macro-block (16×16) is embedded into the same macro-block. Before embedding, each macro-block is divided into 8×8 sub-block. Then, DCT is applied on each sub-block (8×8) to embed CBW into the mid-frequency coefficients of DCT. The watermark embedding process is explained using the following steps;

Step 1: Watermarked video is decomposed into RGB channels.
Step 2: Shots are detected using K-PCC and AMD.
Step 3: SVD is applied to detect the keyframe.
Step 4: a 15-bit content-based watermark (CBW'$_i$) is generated for each macro-block (16×16) of the keyframe.
Step 5: Each (16×16) block of the keyframe is divided into (8×8) blocks.
Step 6: DCT is applied on each (8×8) sub-block to get 15-mid frequency coefficients.
Step 7: The generated $\boldsymbol{CBW_j}$ for the macro-block (16×16) is embedded into mid-frequency coefficients of each (8×8) sub-block using Eq- (16)

$$IM'_{mid_{x,y}} = \begin{cases} IM_{mid_{x,y}} + C & if\ IM_{mid_{x,y}} \geq 1\ and\ CBW_j = 1 \\ IM_{mid_{x,y}} - C & if\ IM_{mid_{x,y}} < 1\ and\ CBW_j = 0 \end{cases} \quad (16)$$

Where $IM_{mid_{x,y}}$ is the mid-frequency coefficient, IM'$_{mid_{x,y}}$ is the watermarked mid-frequency coefficients and CBW$_j$ is the jth content-based watermark**.**

Step 8: Inverse DCT is applied to get the watermarked frame.
Step 9: Step- 4 to Step-8 are applied on each keyframe to get watermarked video.

3.4.2. Watermark Detection Process

In this section, we present a watermark detection scheme which is a very simple process. For watermark detection, similar steps are mentioned in watermark embedding. The flow of the watermark detection process is shown in Fig. 6. Initially, KPCC and AMD are applied to the watermarked video to detect shots. From each shot, a watermarked keyframe is extracted using SVD. Then, watermarked mid-frequency coefficients are extracted from each sub-block (8×8) by applying DCT. The similarity between extracted and detected watermarks is checked using Normal correlation coefficients. The Watermark detection process is explained using the following steps.

Step 1: Watermarked video is decomposed into RGB channels.
Step 2: Shots are detected using KPCC and AMD.
Step 3: SVD is applied to get watermarked keyframe.
Step 4: Watermarked mid-frequency coefficients are extracted from each sub-block (8×8) by applying DCT and CBW''$_j$ is detected using Eq.17.
Step 5: Then, a Content-based watermark CBW''$_j$ is regenerated for the frame by applying Arnold transform, DCT and DWT using Eq. (15)
Step 6: The regenerated CBW''$_j$ and extracted CBW'$_j$ watermarks are compared using the Normal correlation Coefficient (NCC) as in Eq.18.

$$W'_j = \begin{cases} 0 & if\ IM'_{mid_{x,y}} < 0 \\ 1 & if\ IM'_{mid_{x,y}} \geq 1 \end{cases} \quad (17)$$

$$C_j = NCC(CBW'_j\ ,\text{CBW''}_j) \quad (18)$$

3.4.3. Tampered Detection Process

In the tampered detection process, the content-based watermark CBW is extracted from the watermarked keyframe and compared with the content-based watermark CBW" generated from the frame with the second maximum entropy value of the same shot. The difference between the two watermarks is calculated by using the following Eq. (19).

$$D\_AC_{i,j} = |CBW' - CBW''| \quad (19)$$

$D\_AC_{i,j}$ is denotes the difference between generated and extracted content-based watermarks. Whether the area is tampered with or not is analyzed according to the distribution of 0's and 1's. If the difference between generated and extracted watermark is equal to '0', which indicates that both the watermarks are the same, the video is not tampered with.

However, the distribution of 1 in the different images indicates malicious manipulation. The tampered area is identified and located using dense and sparse points.

The dense point in the difference image is defined as the pixel with at minimum one 1 pixel in each of its 8 surrounding pixels. It is a sparse point in all other aspects. The malicious or content-based manipulations are recognized using the following Eq (20).

$$T = \frac{Number\ of\ Dense\ point(ND)}{Number\ of\ Dense\ Point\ (ND) + Number\ of\ Sparse\ Point\ (NS)} \quad (20)$$

If T>= Threshold, the manipulation is malicious; otherwise, it is content-based. If it is content-based manipulation, the authentication passes; otherwise, it locates the malicious modification regions. If a pixel in the difference frame is the dense point, then define the four neighborhoods as dense points. Then, find the areas of 4×4 sub-blocks that match the dense point.

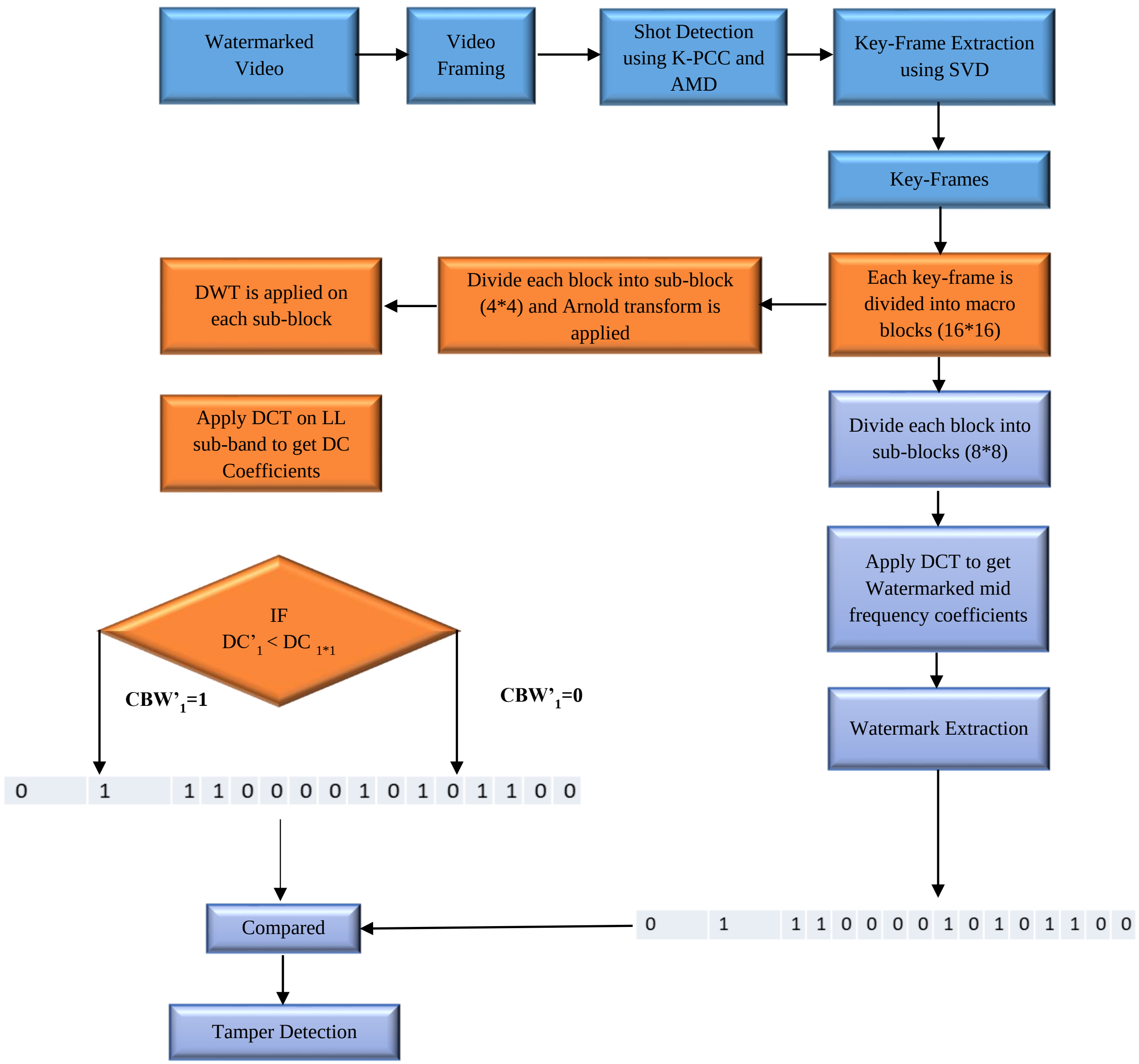

Watermarked Video
Video Framing
Shot Detection using K-PCC and AMD
Key-Frame Extraction using SVD
Key-Frames
Each key-frame is divided into macro blocks (16*16)
Divide each block into sub-block (4*4) and Arnold transform is applied
DWT is applied on each sub-block
Apply DCT on LL sub-band to get DC Coefficients
Divide each block into sub-blocks (8*8)
Apply DCT to get Watermarked mid frequency coefficients
Watermark Extraction
IF DC'₁ < DC 1*1
CBW'₁=1
CBW'₁=0
0 1 1 1 0 0 0 0 1 0 1 0 1 1 0 0
0 1 1 1 0 0 0 0 1 0 1 0 1 1 0 0
Compared
Tamper Detection

Fig. 6 Watermarking detection process

4. Experimental Results

Experiments are conducted to evaluate the performance of the proposed keyframe-based video watermarking scheme. The proposed technique is implemented using MATLAB21a on a computer system. Under this framework, three datasets are used for experimentation. In Dataset-1 total of 300 video sequences in the standard quarter common intermediate format (QCIF) and common intermediate format (CIF) are accessible at https://www.npdi.dcc.ufmg.br/VSUMM. The video sequences include news, carphone, tempete, akiyo, container, mobile, coastguard, flower, highway, silent, etc., consisting of 300 to 500 frames with 30 frames per second frame rate. Dataset-2 consist of 70 RGB videos of 1 to 4 minutes duration with a frame rate of 30 frames per second and frame size of 352 ×240. It includes videos in several categories, including historical, instructional, documentary, and lecture videos. This video dataset is accessible at https://www.sites.google.com/site/vsummsite /download. In Dataset 3 contain 110 RGB videos downloaded from YouTube of several categories, such as cartoons, news, sports, commercial TV shows, and home videos, with a duration of 1 to 10 min. The Dataset-3 is present at https://www.sites.google.com/site/vsummsite/download.

4.1. Assessment of Robustness and Imperceptibility

Several experiments are carried out to determine the imperceptibility and robustness of the proposed video watermarking technique. The robustness of the proposed technique is evaluated by employing various image processing attacks such as filtering, blurring, compression, noise addition, and so on, as well as frame synchronization attacks such as frame averaging, frame dropping, and frame swapping on watermarked video. The perceptual quality of the watermarked video to the human visual system is determined using Peak Signal to Noise Ratio (PSNR). In these experiments, the PSNR is measured for every keyframe of the video sequence. The high value of PSNR indicates high watermark invisibility, and the low value of PSNR indicates the low perceptual quality of the watermarked video. PSNR is calculated using the following Eq. 21 [30].

$$PSNR = 10log_{10}\frac{255^2}{\frac{1}{N\times M}\sum_{x=0}^{N-1}\sum_{y=0}^{M-1}(Ofr_{x,y}-Wfr_{x,y})^2} \qquad (21)$$

Where, Ofr and denote the original video frame and the corresponding watermarked video frame, respectively; (x, y) is the position of a pixel in and Wfr; and N × M is the size of each frame.

Normalized cross-correlation (NCC), which is the difference between the original watermark (w) and extracted watermark w', is used to measure watermark immunity against various attacks. The NCC is computed using the following Eq. 22 [30].

$$NCC = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M} w(i,j)w'(i,j)}{\sum_{i=1}^{N}\sum_{j=1}^{M} w^2(i,j)w'^2(i,j)} \qquad (22)$$

Where w and w' are the original and extracted watermark, and N and M represent rows and columns of the watermark, respectively.

4.2. Shot Detection Evaluation

The snapshots of sample videos of three Datasets used for experimentation are shown in Fig. 7. The shots in the videos are detected using the proposed K-PCC and AMD-based shot detection scheme. The value of K-PCC between the seed frame and the successive frame is measured and compared with the threshold. If the value of K-PCC is greater than the threshold, then that frame is considered a frame of the current seed frame's shot; otherwise, it is considered a seed frame for the next shot. The absolute mean difference is calculated between the seed frame and successive frame, and the mean difference is measured and compared with a threshold. If the value of AMD is less than the threshold, then that frame is considered a frame of the current seed frame's shot; otherwise, it is considered a seed frame for the next shot. Examples of shots detected by estimating the K-PCC and AMD difference are shown in Fig. 8 and Fig. 9, respectively, for the 'silent' video from Dataset-1.

Fig. 7 Snapshots of sample videos

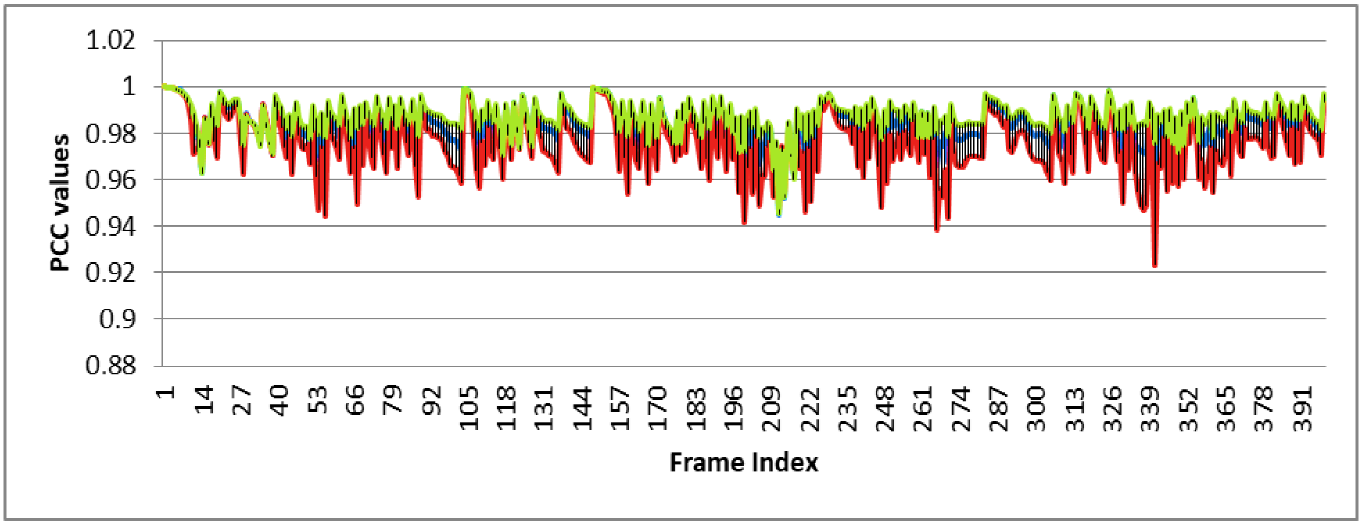


Fig. 8 K-PCC value between seed frame and successive frames

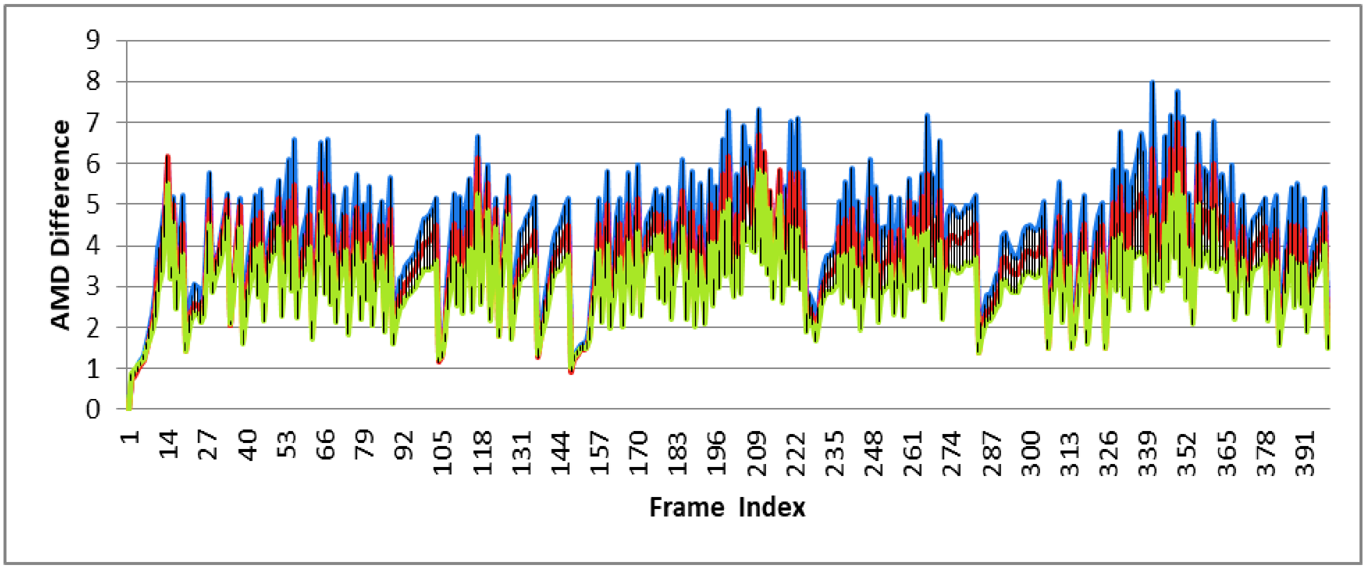


Fig. 9 AMD between seed frame and successive frames

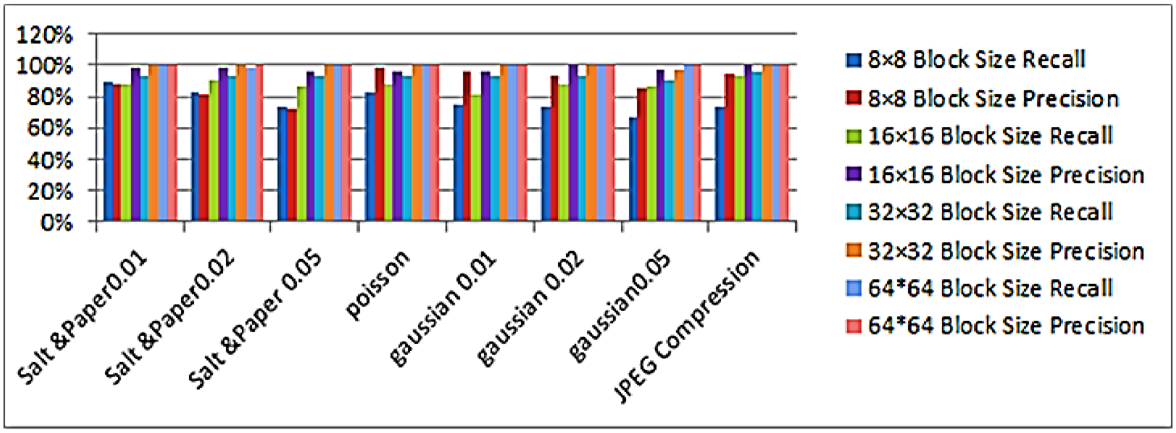


Fig. 10 Precision and recall values of 'Salesman' video

4.3. Evaluation of Keyframes Extraction

This article discusses the findings of a proposed SVD-based keyframe extraction scheme that attempts to reduce the time required to watermark video data while maintaining the overall content of the original video. In this work, the extracted keyframes are matched with the ground truth, which was produced manually after five human observers viewed the videos.

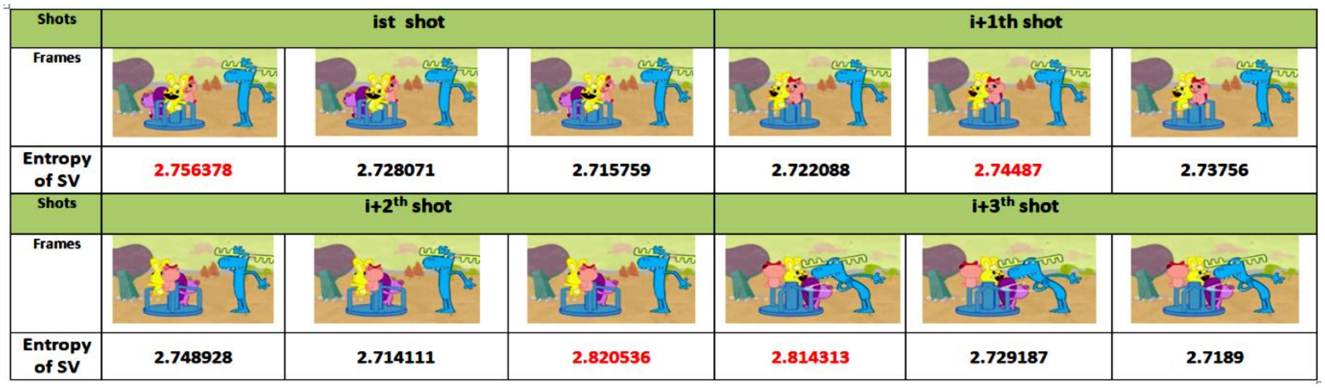
Then, a comparison between keyframes discovered using the suggested method and the actual data is calculated. A real affirmative occurs when the extracted keyframe is recognized as a keyframe by both humans and the suggested approach. False positives occur when the extracted keyframe is identified as a keyframe by the proposed method but not by humans, and false negatives occur when the extracted keyframe is identified as a keyframe by humans but not by the proposed method. The true positive frame (TPF), false positive frame (FPF), and false negative frame are used to measure the precision and recall (FNF).

The recall and precision are computed using the following Eq. 23 and Eq. 24, respectively:

$$Recall = \frac{TPF}{TPF+FNF} * 100\ \% \tag{23}$$

$$Precision = \frac{TPF}{TPF+FPF} * 100\ \% \tag{24}$$

The performance of the proposed keyframe extraction technique is evaluated for various block sizes of 16 × 16, 32 × 32,64 × 64 and the whole frame using precision and recall. It is revealed that the higher the block size better the precision and recall.



Shots	ist shot			i+1th shot		
Frames						
Entropy of SV	2.756378	2.728071	2.715759	2.722088	2.74487	2.73756
Shots	i+2th shot			i+3th shot		
Frames						
Entropy of SV	2.748928	2.714111	2.820536	2.814313	2.729187	2.7189

Fig. 11 Different shots of video 'V11.flv' with entropy values of each frame.

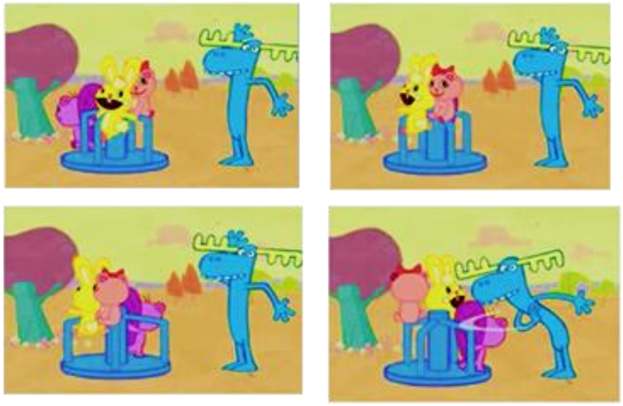


Fig. 12 Keyframes of each shot of the video 'V11.flv'

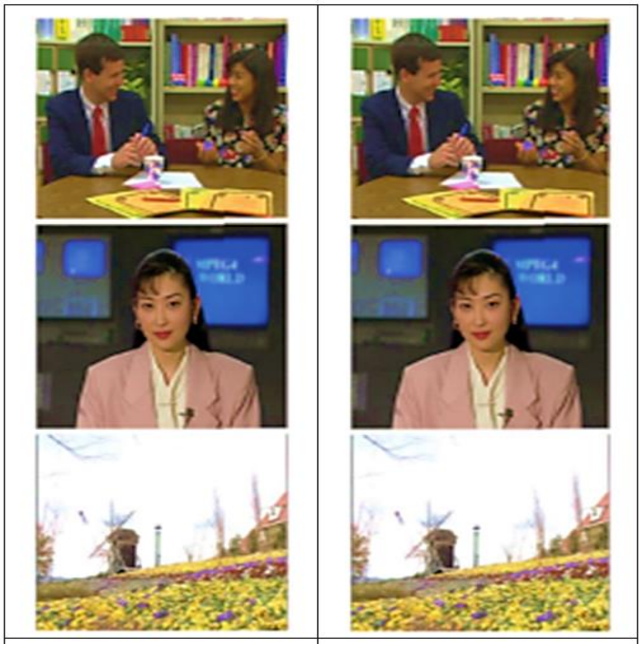
Hence, we have chosen the whole frame to extract the keyframe from the shot. Even if we purposely corrupt the frame with different types of noise, it is observed that the precision and recall achieved are 100%. Fig. 10 shows the precision and recall values of a sample video, 'Salesman' of Dataset-1.

Using the proposed SVD-based keyframe extraction algorithm, SVD is applied on every frame of every shot. The entropy of singular values is measured by extracting singular values from the diagonal matrix of SVD. The proposed keyframes extraction technique detects representative keyframes accurately and semantically from videos of long sequences or with more transitions. The results of keyframe detection by the proposed scheme for the 'V11.flv' video of Dataset-2 are given in Fig. 11, wherein different shots of a 'V11.flv' video and entropy of singular values obtained are shown. The entropy value represented in red color indicates the maximum entropy value frame, which is a keyframe of that shot. Fig. 12 shows accurately and semantically extracted keyframes of each shot of video 'V11.flv' of Dataset-2.

4.4. Imperceptibility Test

The imperceptibility measures perceptual quality between the original and watermarked video. In order to evaluate the perceptual quality of the watermarked video, the PSNR values are measured. A greater PSNR suggests improved performance in terms of imperceptibility. To test the imperceptibility of the proposed watermarking scheme, the original frame is compared with the watermarked frame of the video. Fig. 13 shows the result of the imperceptibility of watermarked video from Dataset-1, Dataset-2 and Dataset-3, in which the first column indicates original frames and the second column represents watermarked frames. It is observed that using the proposed scheme, the perceptual video quality remains unchanged.

To evaluate the invisibility of the proposed video watermarking approach in an objective way, we used PSNR (peak signal-to-noise ratio) as an evaluation parameter by comparing the original video and the watermarked video. A greater PSNR indicates improved performance in terms of imperceptibility.



a) Original Frame **b) Watermarked frame**

Fig. 13 Original (first column) and watermarked (second column) video frames of Dataset-1

Table 1. shows the average PSNR values of the watermarked videos

Dataset	Video	PSNR
Dataset-1	Silent	**63.21**
	Foreman	**63.13**
	News	**63.35**
Dataset-2	v21	**63.34**
	v23	**62.21**
	v24	**63**
Dataset-3	v11	**62.99**
	v12	**63.49**
	v13	**62.9**

Table 1 shows the average PSNR values of three videos of each dataset. To eliminate the effect of randomness, we provide the average of the PSNR values for all the frames in the video. From Table , It is observed that the average PSNR values of watermarked videos of three datasets are 63.56dB, 62.85dB and 63.13dB, respectively. It is found that the average PSNR of all the watermarked videos of all Datasets are above 62.85dB, which proves the good imperceptibility of the watermarked videos.

4.5. Robustness Test

We have used NCC to evaluate the robustness of the proposed video watermarking approach against different attacks. The NCC value close to one indicates less video distortion and better algorithm robustness. The experimental results for testing the robustness of the proposed method against different attacks for five videos of Dataset-1 are shown in Fig. 14 and Fig. 15. It is revealed that the proposed algorithm is robust against various attacks like frame averaging, frame swapping and frame dropping, filtering, the addition of noise, compression, blur, and brightness change. The PSNR measure is used for checking the quality of watermarked images. For the experimentation, 25% of watermarked frames are considered to undergo frame dropping, frame averaging and frame swapping attacks. The PSNR values of attacked videos are found in the range of 61dB to 63dB. The PSNR values of different attacked watermarked videos are presented in Fig 14. The Normalized cross-correlation (NCC) values are measured between the detected watermark and the original watermark after applying various attacks on the watermarked video, which are shown in Fig. 15. It is observed that the NCC value of the detected watermark ranges from 0.9 to 1.

4.6. Evaluation of Tampered Area Location

Fig. 16 shows that the proposed watermark scheme can detect and locate malicious areas. The proposed video watermarking is used to detect and localize malicious manipulation. During tamper detection, the content-based watermark is extracted from temporal attacks; however, the proposed method additionally detects and localizes spatial attacks, i.e. video frame content forgeries. Instead of the binary sequence, the watermark image itself is employed to identify forgery in this case. The following experiment was carried out to demonstrate the concept. An arbitrary video frame fragment was modified by assigning random values to the pixels. The watermark was then extracted from the frame. The “tampered” frame is shown in Fig. 16(a), while Fig. 16(b) shows the watermark extracted from the modified frame.

4.7. Comparative Analysis of Proposed Keyframe Extraction

This section compares the proposed scheme's result with recent popular keyframe extraction approaches. We have compared our results of the SVD-based keyframe with GSMD [30], RPCA-KFSA[38] and FASAM [37] schemes. The keyframes extracted by these methods are compared with the corresponding ground truth keyframes [30,37,38].

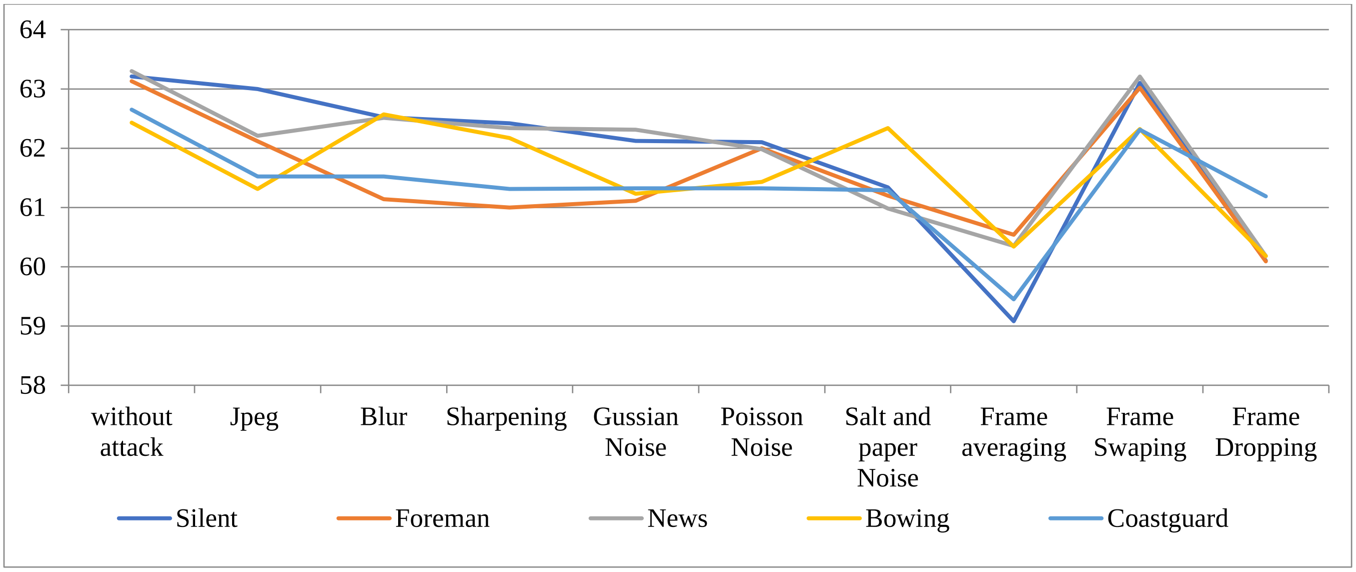


Fig. 14 PSNR values of watermarked videos of Dataset-1 after applying various attacks

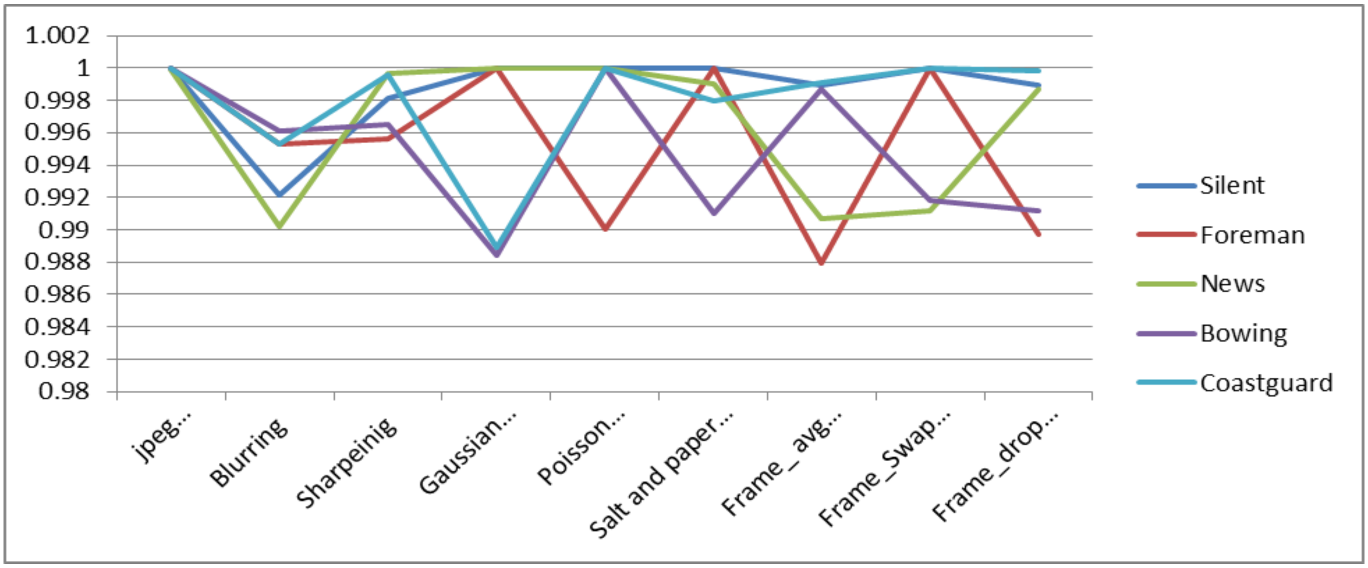


Fig. 15 NCC values of watermark after applying various attacks

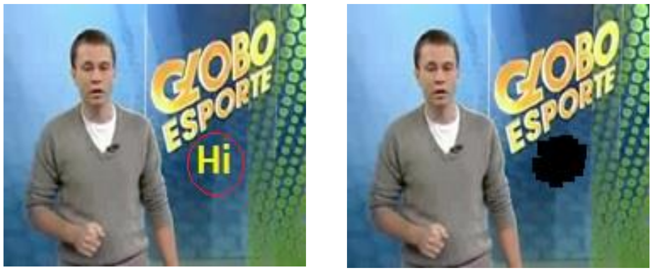


Fig 16 a) Tampered 8th keyframe b) Tampered area localization.

Table 2. Recall and precision achieved by different techniques for 'silent video'

	GSMD [30]	RPCA-KFE [38]	FASAM [37]	proposed
Recall	95	96	86	100
Precision	93	95	82	100

Table 2 shows average recall and precision values for all the above schemes. Recall and precision values obtained by the proposed method are 100%. So it is proved that the proposed SVD–based keyframe extraction scheme outperforms the performances of GSMD[30], RPCA-KFSA[38] and FASAM[37] schemes. From the experimentation, it is revealed that the proposed keyframe extraction scheme is less complicated, segments shots accurately and extracts the keyframes from each shot while fulfilling strong robustness.

4.8. Comparative Analysis of the Proposed Video Watermarking Scheme

Comparison results of the proposed video watermarking scheme with existing recent schemes [22,30,31,33,34,35] are presented in this section. A detailed description of these schemes is already given in related work. Table 3 illustrates the results of comparing the proposed video watermarking scheme with existing schemes under different attacks.

Table 3. NCC variations of the watermark in various attacks compared with existing methodology

Attacks	**[10]**	**[22]**	**[31]**	**[35]**	**[33]**	**[34]**	**Proposed**
Compression	0.997	0.997	0.9883				0.9991
Gaussion Noise	1	0.97	0.9844	0.94	0.92	0.9881	1
Salt and Paper	1	0.98	0.9844		0.87	0.9878	1
Blurring	0.978				0.95		0.9911
Median Filtering	1	0.988	0.9965	0.95	0.93	0.999	0.9989
Frame averaging	0.99	0.967	0.99	0.93	0.9		0.9993
Brighten		0.987			0.94		1
Frame Dropping				0.97	0.95		1
Frame Swapping				0.99	0.92		1

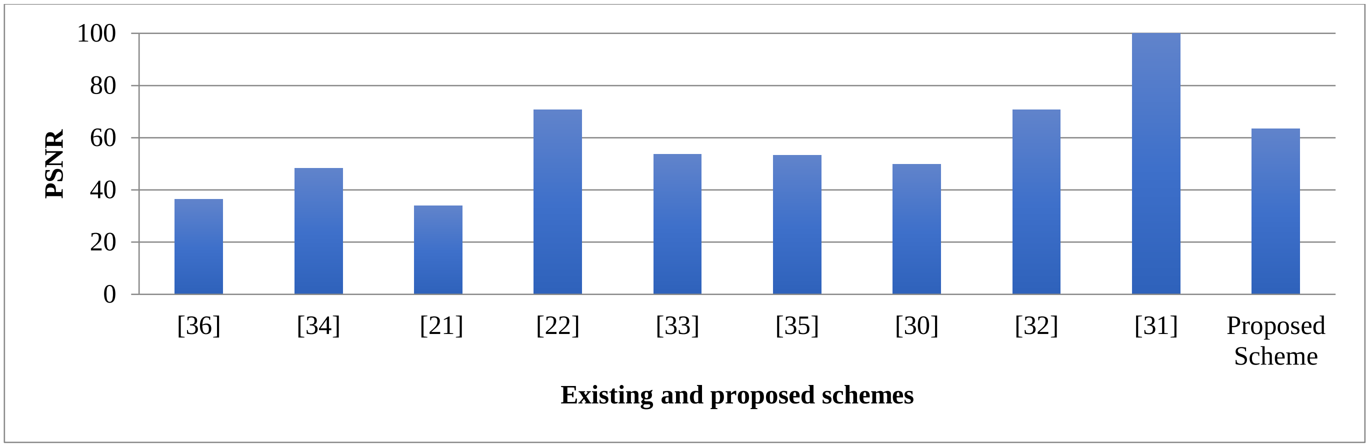


Fig. 17 PSNR values of watermark extracted from the video after various attacks

To evaluate the robustness of the proposed scheme, we compare the proposed scheme with six recent video watermarking schemes, which we have described in the related work. Table 3 shows the results of comparing the proposed scheme with other recent schemes under different kinds of attacks like a blur, brighten, Gaussian noise attack, Salt and Pepper, median filtering, frame dropping and averaging. It is observed that the proposed scheme outperforms all the above-mentioned existing schemes. As can be seen from Table 3, the average NCC value of the proposed scheme is higher than 99% for almost all the attacks. When the NCC is higher than 0.8, it means that this algorithm has decent robustness. The average NCC values of the proposed scheme are comparatively greater than the mentioned schemes for various kinds of attacks like a blur, brighten, Gaussian noise attack, Salt and Pepper, median filtering, frame dropping and averaging. For example, for the Gaussian noise attack, Salt and Pepper, brighten, frame dropping, frame averaging and frame swapping attacks of the proposed scheme achieve an NCC of 100%. Whereas NCC obtained by recent schemes are considerably lower than the proposed scheme. However, the NCC value obtained using [35] and [33] is lower than other schemes.

Table 4 illustrates the comparison of the proposed video watermarking scheme with the existing schemes in terms of payload capacity, PSNR and time complexity comparison. Table 4 shows that the proposed scheme can achieve a high PSNR of 63.13 dB, a high payload capacity of up to 396 bits/frame and archives min time of up to 34.34 sec compared to other schemes. Compared to all existing schemes, our proposed method performs well in terms of imperceptibility, capacity and time. As a result, the proposed video watermarking system is a successful applicant for real-time applications.

The PSNR values obtained from the proposed approach are compared to those obtained from existing systems, as illustrated in Fig.16 PSNR values of 40 dB and above are thought to be better in human visual perception [43].

Table 4. Payload Capacity, PSNR and Time Complexity comparison for 'Foreman' video sequence

Evaluation	[41]	[45]	[30]	Proposed Method
Capacity (bits/frame)	2	64	396	**396**
PSNR	46.86	49.11	45.31	**63.13**
TIME	76.31	136.39	41.16	**34.35**

The average PSNR values of proposed and current schemes are shown in Fig. 16 [21,22,30-36]. Fig.17 shows that the suggested system has a PSNR value greater than 62.8dB, which is better than the existing schemes [21,22,30-36] except [22,31]. As in [31], authors have considered only boundary blocks of selected keyframes for embedding and extracting watermarks. Whereas in [22], authors have only generated shares for copyright protection instated of embedding them into the host videos; therefore, it shows high PSNR, i.e. infinity. Therefore, the present watermarking scheme provides good imperceptibility.

5. Conclusion

This paper presents a blind, efficient and secure video watermarking system for H.264/AVC videos. Initially, we developed a novel and efficient shot detection technique using K-PCC and AMD. The reason behind the use of K-PCC and AMD is that it is robust against illumination changes, object motion and camera operation. As embedding watermarks in every frame is time consuming, we have proposed a robust and efficient SVD-based keyframe extraction algorithm. Singular values are more robust to Gaussian noise, Salt and pepper noise, compression, transpose, and scaling. Further, to improve the video authenticity and robustness of the content-based watermark (CBW), three robust mathematical transforms DCT, DWT, and Arnold transform, are employed to generate CBW. For embedding generated watermarks, DCT is used to improve the robustness and imperceptibility of the host video.

Moreover, the extraction function allows a blind extraction of the embedded authentication code. Promising results have been achieved in terms of imperceptibility and robustness.

The proposed video watermarking scheme is tested on three Datasets containing different types of videos like animations, news, cartoon, flowers, personal interviews, etc. The experimental results demonstrate that the proposed video watermarking scheme is robust to potential attacks such as Gaussian noise, Salt and paper noise, blurring, filtering, compression, frame dropping, frame averaging, frame swapping etc. Encouraging results have been achieved in terms of imperceptibility, robustness, and with respect to the capacity of insertion and complexity.

References

[1] Sandra Eliza Fontes Avila et al., “VSUMM: A Mechanism Designed to Produce Static Video Summaries and a Novel Evaluation Method,” *Pattern Recognition Letters*, vol. 32, no. 1, pp. 56–68, 2011. [CrossRef] [Google Scholar] [Publisher link]

[2] Yassine Himeur, and Bachir Boudraa, “Secure and Robust Audio Watermarking System for Copyright Protection,” *2012 24th International Conference on Microelectronics (ICM)*, 2012. [CrossRef] [Google Scholar] [Publisher link]

[3] Nazeer Muhammad, and Nargis Bibi, “Digital Image Watermarking using Partial Pivoting Lower and Upper Triangular Decomposition into the Wavelet Domain,” *IET Image Processing*, vol. 9, no. 9, pp. 795–803, 2015. [CrossRef] [Google Scholar] [Publisher link]

[4] Dawen Xu, Rangding Wang, and Yun Q. Shi, “Data Hiding in Encrypted h.264/AVC Video Streams by Codeword Substitution,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 4, pp. 596–606, 2014. [CrossRef] [Google Scholar] [Publisher link]

[5] Nilesh Dubey, and Hardik Modi, “A State of Art Comparison of Robust Digital Watermarking Approaches for Multimedia Content (Image and Video) Against Multimedia Device Attacks,” *International Journal of Engineering Trends and Technology*, vol. 70, no. 8, pp. 132-139, 2022. [CrossRef] [Google Scholar] [Publisher link]

[6] K.Duraisamy, “Video Source Tracking and Copyright Protection in Video Watermarking,” *SSRG International Journal of Communication and Media Science*, vol. 1, no. 1, pp. 6-11, 2014. [CrossRef] [Publisher link]

[7] L. Agilandeeswari, and K. Ganesan, “A Robust Color Video Watermarking Scheme Based on Hybrid Embedding Techniques,” *Multimedia Tools and Applications*, vol. 75, pp. 8745–8780, 2016. [CrossRef] [Google Scholar] [Publisher link]

[8] Yassine Himeur, and Karima Ait-Sadi, “Joint Color and Texture Descriptor using Ring Decomposition for Robust Video Copy Detection in Large Databases,” *2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2015. [CrossRef] [Google Scholar] [Publisher link]

[9] Yassine Himeur, Karima Ait-Sadi, and Abdelmalik Ouamane, “A Fast and Robust Key-frames Based Video Copy Detection Using BSIF-RMI,” *2014 International Conference on Signal Proceedings and Multimedia Application (SIGMAP)*, 2014. [Google Scholar] [Publisher link]

[10] Dawen Xu, Rangding Wang, and Jicheng Wang, “A Novel Watermarking Scheme for H.264/AVC Video Authentication,” *Signal processing: Image Communication*, vol. 26, no. 6, pp. 267-279, 2011. [CrossRef] [Google Scholar] [Publisher link]

[11] T. Wiegand et al., “Overview of the H.264/AVC Video Coding Standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol.13, no. 7, pp. 560–576, 2003. [CrossRef] [Google Scholar] [Publisher link]

[12] Hanieh Khalilian, and Ivan V. Bajic, “Video Watermarking with Empirical PCA-based Decoding,” *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4825–4840, 2013. [CrossRef] [Google Scholar] [Publisher link]

[13] Tanfeng Sun et al., “A Novel Differential Energy Video Watermarking Based on Watson Visual Model,” *2009 Second International Symposium on Electronic Commerce and Security*, 2009. [CrossRef] [Google Scholar] [Publisher link]

[14] C.H. Wu et al., “A Flexible H.264/AVC Compressed Video Watermarking Scheme using Particle Swarm Optimization Based Dither Modulation,” *AEU- International Journal of Electronics and Communications*, vol. 65, no. 1, pp. 27-36, 2011. [CrossRef] [Google Scholar] [Publisher link]

[15] Jiang Xuemei, Liu Quan, and Wu Qiaoyan, “A New Video Watermarking Algorithm Based on Shot Segmentation and Block Classification,” *Multimedia Tools and Applications*, vol. 62, pp. 545–560, 2013. [CrossRef] [Google Scholar] [Publisher link]

[16] Ta Minh Thanh et al., “Robust Semi-blind Video Watermarking Based on Framepatch Matching,” *AEU-International Journal of Electronics and Communications*, vol. 68, no. 10, pp. 1007-1015, 2014. [CrossRef] [Google Scholar] [Publisher link]

[17] Nisreen I. Yassin, Nancy M. Salem, and Mohamed I. Eladawy, “Block Based Video Watermarking Scheme using Wavelet Transform and Principle Component Analysis,” *International Journal of Computer Science Issues*, vol. 9, no. 1, pp. 296–301, 2012. [Google Scholar] [Publisher Link]

[18] Ching-Sheng Hsu, and Young-Chang Hou, “A Visual Cryptography and Statistics Based Method for Ownership Identification of Digital Images,” *World Academy of Science, Engineering and Technology*, vol. 2, pp. 172–175, 2005.

[19] Majid Masoumi, and Shervin Amiri, “A Blind Scene-based Watermarking for Video Copyright Protection,” *AEU-International Journal of Electronics and Communications*, vol. 67, no. 6, pp. 528–535, 2013. [CrossRef] [Google Scholar] [Publisher link]

[20] Jiang Xuemei, Liun Quan, and Wu Qiaoyan, “A New Video Watermarking Algorithm Based on Shot Segmentation and Block Classification,” *Multimedia Tools and Applications*, vol. 62, pp. 545–560, 2013. [CrossRef] [Google Scholar] [Publisher link]

[21] Chen Li et al., “A Semi-Fragile Video Watermarking Algorithm Based on H.264/AVC,” *Wireless Communications and Mobile Computing*, vol. 2020, 2020. [CrossRef] [Google Scholar] [Publisher link]

[22] Shahad Almuzairai, and Nisreen Innab, “Video Watermarking System for Copyright Protection based on Moving Parts and Silence Deletion,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 2, 2019. [CrossRef] [Google Scholar] [Publisher link]

[23] Sibaji Gaj, Ashish Singh Patel, and Arijit Sur, “Object Based Watermarking for H.264/AVC Video Resistant to Rst Attacks,” *Multimedia Tools and Applications*, vol. 75, pp. 3053-3080, 2016. [CrossRef] [Google Scholar] [Publisher link]

[24] Y. Tonomura et al., “VideoMAP and VideoSpaceIcon: Tools for Automatizing Video Content,” *In: Proceedings ACM INTERCHI’93 Conference*, 1993.

[25] Hirotada Ueda, Takafumi Miyatake, and Satoshi Yoshizawa, “IMPACT: An Interactive Natural-motion-picture Dedicated Multimedia Authoring System,” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 343-350, 1991. [CrossRef] [Google Scholar] [Publisher link]

[26] Yong Rui, T.S. Huang, and S. Mehrotra, “Exploring video Structure Beyond the Shots,” *Proceedings IEEE International Conference on Multimedia Computing and Systems (ICMCS)*, 1998. [CrossRef] [Google Scholar] [Publisher link]

[27] Alex Pentland et al., “Video and Image Semantics: Advanced Tools for Telecommunications,” *IEEE MultiMedia*, vol. 1, pp. 73–75, 1994. [CrossRef] [Google Scholar] [Publisher link]

[28] Zhonghua Sun, and Ping Fu, “Combination of Color and Object Outline Based Method in Video Segmentation,” *Proceedings of SPIE-The International Society for Optical Engineering*, pp. 61-69, 2004. [CrossRef] [Google Scholar] [Publisher Link]

[29] Shaik Hedayath Basha et al., “Video Watermarking using DWT and Elgamal for Authentication and Security,” *SSRG International Journal of Electronics and Communication Engineering*, vol. 5, no. 1, pp. 4-9, 2018. [CrossRef] [Publisher link]

[30] Yassine Himeur, and Abdelkrim Boukabou, “A Robust and Secure Key-frames Based Video Watermarking System using Chaotic Encryption,” *Multimedia Tools and Applications*, vol. 77, pp. 8603-8627, 2018. [CrossRef] [Google Scholar] [Publisher link]

[31] Zhila Bahrami, and Fardin Akhlaghian Tab, “A New Robust Video Watermarking Algorithm Based on SURF Features and Block Classification,” *Multimedia Tools and Applications*, vol. 77, pp. 327-345, 2018. [CrossRef] [Google Scholar] [Publisher link]

[32] Soumik Das, Monalisa Banerjee, and Atal Chaudhuri, “An Improved Video Key-frame Extraction Algorithm Leads to Video Watermarking,” *International Journal of Informatioan Tecnology*, vol. 10, pp. 21-34, 2018. [CrossRef] [Google Scholar] [Publisher link]

[33] Sethuraman Ponni alias Sathya, and Srinivasan Ramakrishnan, “Non-redundant Frame Identification and Keyframe Selection in DWT-PCA Domain for Authentication of Video,” *IET Image Processing*, vol. 14, no. 2, pp. 366-375, 2020. [CrossRef] [Google Scholar] [Publisher link]

[34] Roop Singh, Himanshu Mittal, and Raju Pal, “Optimal Keyframe Selection-based Lossless Video-watermarking Technique Using IGSA in LWT Domain for Copyright Protection,” *Complex & Intelligent Systems*, vol. 8, pp. 1047-1070, 2022. [CrossRef] [Google Scholar] [Publisher link]

[35] S. Ponni alias Sathya, and S. Ramakrishnan, “Fibonacci Based Key Frame Selection and Scrambling for Video Watermarking in DWT–SVD Domain,” *Wireless Personal Communications*, vol. 102, pp. 2011-2031, 2018. [CrossRef] [Google Scholar] [Publisher link]

[36] Chirag Sharma et al., “A Secured Frame Selection Based Video Watermarking Technique to Address Quality Loss of Data: Combining Graph Based Transform, Singular Valued Decomposition, and Hyperchaotic Encryption,” *Security and Communication Networks*, 2021. [CrossRef] [Google Scholar] [Publisher link]

[37] Naveed Ejaz, Irfan Mehmood, and Sung Wook Baik, “Feature Aggregation Based Visual Attention Model for Video Summarization,” *Computers and Electrical Engineering*, vol. 40, no. 3, pp. 993–1005, 2014. [CrossRef] [Google Scholar] [Publisher link]

[38] Chinh Dang, and Hayder Radha, “RPCA-KFE: Key Frame Extraction for Video Using Robust Principal Component Analysis,” *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3742–3753, 2015. [CrossRef] [Google Scholar] [Publisher link]

[39] M.G. Kendall, *The Advanced Theory of Statistics, 4th Edition*, Macmillan, 1979.

[40] P.S. Dinesh et al., “Digital Fragile Watermarking Video Based on Discrete Wavelet Transformation,” *SSRG International Journal of Computer Science and Engineering*, vol. 6, no. 11, pp. 5-9, 2019. [CrossRef] [Publisher link]

[41] Zhi Li, Xiao-Wei Chen, and Jianhua Ma, “Adaptively Imperceptible Video Watermarking Based on the Local Motion Entropy,” *Multimedia Tools and Applications*, vol. 74, pp. 2781–2802, 2015. [CrossRef] [Google Scholar] [Publisher link]

[42] Wang Xiang-yang et al., “A New Robust Digital Watermarking using Local Polar Harmonic Transform,” *Computers and Electrical Engineering*, vol. 46, pp. 403–418, 2015. [CrossRef] [Google Scholar] [Publisher link]

[43] David R. Bull, *Communicating Pictures: A Course in Image and Video Coding*, Cambridge: Academic Press, 2014. [Google Scholar] [Publisher link]

[44] Amal Hammami, Amal Ben Hamida, and Chokri Ben Amar, “Blind Semi-fragile Watermarking Scheme for Video Authentication in Video Surveillance Context,” *Multimedia Tools and Applications*, vol. 80, pp. 7479-7513, 2021. [CrossRef] [Google Scholar] [Publisher link]

[45] Preeti Garg, and R. Rama Kishore, “Performance Comparison of Various Watermarking Techniques,” *Multimedia Tools and Applications*, vol. 79, pp. 25921-25967, 2020. [CrossRef] [Google Scholar] [Publisher link]

[46] Mohammad Moosazadeh, and Gholamhossein Ekbatanifard, “An Improved Robust Image Watermarking Method using DCT and YCoCg-R Color Space,” *Optik*, vol. 140, pp. 975-988, 2017. [CrossRef] [Google Scholar] [Publisher link]