Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.056438

ARTICLE

Adaptive Video Dual Domain Watermarking Scheme Based on PHT Moment and Optimized Spread Transform Dither Modulation

Yucheng Liang[1,2,*], Ke Niu[1,2,*], Yingnan Zhang[1,2], Yifei Meng[1,2] and Fangmeng Hu[1,2]

[1]College of Cryptographic Engineering, Engineering University of the Chinese People's Armed Police Force, Xi'an, 710086, China

[2]Key Laboratory of Information Security of the Chinese People's Armed Police Force, Engineering University of the Chinese People's Armed Police Force, Xi'an, 710086, China

*Corresponding Authors: Yucheng Liang. Email: liangyucheng2000@163.com; Ke Niu. Email: niuke@163.com

Received: 23 July 2024 Accepted: 20 September 2024 Published: 18 November 2024

ABSTRACT

To address the challenges of video copyright protection and ensure the perfect recovery of original video, we propose a dual-domain watermarking scheme for digital video, inspired by Robust Reversible Watermarking (RRW) technology used in digital images. Our approach introduces a parameter optimization strategy that incrementally adjusts scheme parameters through attack simulation fitting, allowing for adaptive tuning of experimental parameters. In this scheme, the low-frequency Polar Harmonic Transform (PHT) moment is utilized as the embedding domain for robust watermarking, enhancing stability against simulation attacks while implementing the parameter optimization strategy. Through extensive attack simulations across various digital videos, we identify the optimal low-frequency PHT moment using adaptive normalization. Subsequently, the embedding parameters for robust watermarking are adaptively adjusted to maximize robustness. To address computational efficiency and practical requirements, the unnormalized high-frequency PHT moment is selected as the embedding domain for reversible watermarking. We optimize the traditional single-stage extended transform dithering modulation (STDM) to facilitate multi-stage embedding in the dual-domain watermarking process. In practice, the video embedded with a robust watermark serves as the candidate video. This candidate video undergoes simulation according to the parameter optimization strategy to balance robustness and embedding capacity, with adaptive determination of embedding strength. The reversible watermarking is formed by combining errors and other information, utilizing recursive coding technology to ensure reversibility without attacks. Comprehensive analyses of multiple performance indicators demonstrate that our scheme exhibits strong robustness against Common Signal Processing (CSP) and Geometric Deformation (GD) attacks, outperforming other advanced video watermarking algorithms under similar conditions of invisibility, reversibility, and embedding capacity. This underscores the effectiveness and feasibility of our attack simulation fitting strategy.

KEYWORDS

Dual-domain; H.264; group of pictures; polar harmonic transform; spread transform dither modulation

1 Introduction

Digital video watermarking is a crucial technology in information security, specifically for protecting digital video copyrights [1]. Recognized as an essential securing technique, it incorporates fundamental principles and key assumptions that guide copyright protection efforts. Advancements in this technology have led to new performance requirements, including imperceptible watermark visibility, enhanced robustness of algorithms, improved anti-counterfeiting capabilities, and reliable watermark extraction. These demands drive the ongoing evolution of digital video watermarking, establishing a strong theoretical foundation for its application in copyright protection. Thus, digital video watermarking evolves beyond a mere technical tool to a fundamental principle in information security, fostering continuous exploration and implementation of copyright protection measures.

Digital video watermarking involves embedding watermark information into videos for purposes such as copyright protection [2], covert communication [3], and integrity verification [4]. Watermarking technologies can be categorized into two main types based on their attack resilience and the reversibility of the embedding process: robust watermarking and reversible watermarking. Robust watermarking [5,6] is resilient to various attacks but permanently alters the original video, making it suitable for contexts like commercial photography and online videos. In contrast, reversible watermarking [7,8] allows for the restoration of the original video from the watermarked version when no attacks occur, ideal for applications requiring original video recovery, such as in medical or remote sensing fields.

The concept of robust reversible watermarking (RRW) has emerged in digital image copyright protection, combining the strengths of both robust and reversible watermarking approaches. RRW allows for the recovery of the original carrier and watermark extraction when the watermarked carrier is intact, but in the event of an attack, it focuses solely on watermark extraction. This dual functionality makes RRW particularly suitable for contexts that demand both copyright protection and the preservation of original content, such as digital art.

RRW can be further classified into two categories based on attack resilience: RRW resistant to common signal processing (CSP) attacks (RRW-CSP) [9–11] and RRW that withstands both CSP [12,13] and geometric deformation (GD) [14,15] attacks (RRW-CG) [16–18]. While RRW-CSP is robust against various CSP attacks like noise, compression, and filtering, it is less effective against GD attacks such as rotation and scaling. Conversely, RRW-CG effectively mitigates both CSP and GD attacks, albeit with a trade-off of reduced embedding capacity or increased distortion levels.

In the early phases of robust reversible watermarking against common signal processing attacks (RRW-CSP), De Vleeschouwer et al. [19] introduced a histogram rotation technique that partitions pixels into embedded blocks, mapping them to circular histograms and embedding watermarks via histogram rotation. While effective against Joint Photographic Experts Group (JPEG) compression, this method suffers from potential overflow/underflow problems. To mitigate this, Ni et al. [20] proposed a revised RRW-CSP approach that categorizes embedded blocks based on their grayscale distribution and applies tailored embedding strategies, ensuring resilience against both JPEG and JPEG2000 formats. Over time, myriad RRW-CSP methods have emerged across spatial [21], transform [22], and other domains [23,24]. Recently, Kumar et al. [10] developed a two-layer RRW-CSP method, differentiating between high-significance and low-significance image planes for watermark embedding and auxiliary information insertion. While this approach remains robust to JPEG attacks, its susceptibility to noise poses limitations. Liang et al. [11] utilized the homomorphic multiplication property of the Paillier cryptosystem for watermark embedding in the encrypted domain, achieving RRW-CSP effectiveness for encrypted images, albeit with reduced visual quality of the watermark.

In comparison, RRW that withstands both common signal processing (CSP) and geometric deformation (GD) attacks (RRW-CG) offers improved adaptability. The pioneering RRW-CG technique by Chrysochos et al. [12] combined paired histogram bins for watermark embedding by interchanging pixel values, demonstrating resistance to GD issues such as rotation but lacking resilience to filtering attacks with a limited embedding capacity. Chang et al. [13] subsequently developed an RRW-CG technique that manipulates differences in DCT coefficients of subsampled images for watermark embedding, providing resistance against noise, compression, and geometric transformations, albeit with room for enhanced overall robustness. Hu et al. [14,15] introduced modern RRW-CG methods using Quantization Index Modulation (QIM) to embed watermarks into Zernike and Polar Harmonic Transform (PHT) moment amplitudes, showing promising resistance against CSP and GD attacks. Further, Tang et al. [16] proposed using pseudo-Zernike moments for watermarking with adaptive normalization for enhanced stability and later devised an attack simulation-based method to optimize robustness and invisibility. Tang et al. [25] also introduced a two-stage RRW-CG scheme with adaptive normalization and embedding, leveraging PHT moments as carriers to strengthen stability against attacks. Lastly, Guo et al. [26] integrated the Integer Wavelet Transform with Zernike moments to embed watermarks in low-frequency wavelet coefficients while preserving reversibility.

Though RRW methods are well-developed in digital image copyright protection, research on their application to digital video watermarking is limited. Most video watermarking algorithms have focused on robust and reversible techniques separately, neglecting to leverage their combined benefits. Notable advancements in robust video watermarking include He et al. [5], who enhanced robustness against geometric distortions and synchronization through low-order recursive Zernike moments, and Chen et al. [27], who proposed a rapid watermarking method based on Zernike moments, grouping video frames for watermark insertion using singular value decomposition while maintaining imperceptibility. Asikuzzaman et al. [3] introduced a blind video watermarking technique employing dual-tree complex wavelet transform (DT-CWT) and singular value decomposition (SVD) to combat camera attacks effectively. Shapiro et al. [28] focused on increasing watermark capacity, robustness, and visual imperceptibility through chaotic placement of watermark fragments, achieving successful extraction even under lossy compression. Sharma et al. [29] presented a secured watermarking technique for copyright protection, utilizing Secured Graph Based Transform and Hyperchaotic Encryption. The proposed frame selection algorithm ensures size efficiency and quality preservation, demonstrating robustness against various attacks.

While current digital video watermarking approaches excel in robustness and attack resistance, they often neglect the restoration aspect of copyright protection. The integration of robustness and reversibility is essential for maintaining video integrity during authentication processes.

As short video becomes an increasingly popular form of digital entertainment, there is an urgent need for comprehensive copyright protection and content integrity solutions to safeguard the ownership rights of legitimate users. Existing video robust watermarking algorithms often struggle to balance robustness and reversibility, which poses challenges for effective copyright protection. Our objective is to address these challenges while also providing a mechanism for video information integrity authentication. Inspired by Tang et al. [25], we propose an adaptive dual-domain watermarking scheme for digital video that utilizes the Polar Harmonic Transform (PHT) moment combined with optimized Spread Transform Dither Modulation (STDM). This innovative approach aims to guarantee high robustness while maintaining reversibility, thus offering an effective solution for the ownership protection of short videos. To optimize invisibility during watermark embedding, we designed a key frame selection algorithm based on scene smoothness, allowing us to choose the most effective frames for watermark insertion. Additionally, utilizing spread spectrum watermarking

technology, we distribute each watermark bit across multiple carriers to enhance robustness. Through optimization of the traditional STDM, robust watermark bits are embedded with minimal alterations to the carrier, ensuring resilient watermarking.

In our embedding process, we employ prior knowledge of common target attacks to guide adaptive normalization and embedding processes for robust watermarking. This strategy assesses PHT moment stability through simulated target attacks, enabling dynamic adjustment of moment weights. Our adaptive embedding technique considers watermarked video post-embedding as a labeled candidate, undergoing attack simulations to determine optimal strength for each watermark bit, thereby enhancing robustness and capacity.

Following initial robust watermark insertion via STDM, we integrate reversible watermarking using recursive codes [30] for attack detection and authentication. This combined methodology improves robustness, increases embedding capacity, and maintains invisibility and reversibility. Extensive experimental results reveal our scheme's resilience against CSP attacks such as Additive White Gaussian Noise (AWGN), JPEG, and JPEG2000, along with GD attacks like rotation and scaling. Compared to existing video robust watermarking algorithms, our approach excels in robustness, invisibility, and embedding capacity, facilitating effective video attack detection, authentication, and original video restoration.

The proposed dual-domain watermarking-CG scheme represents a significant advancement in digital video copyright protection. By optimizing the existing image-oriented RRW framework, this approach achieves superior robustness and embedding capacity compared to current video watermarking algorithms, while ensuring invisibility and reversibility. It also introduces recovery capabilities and the ability to detect attacks on watermarked videos.

A key contribution is the development of a frame selection algorithm based on scene smoothness, which minimizes distortion during watermark embedding, thus enhancing robustness and invisibility. Additionally, an adaptive normalization strategy is introduced, which evaluates PHT moment stability against multiple known attacks, allowing for dynamic adjustment of normalized weights to improve robustness.

Furthermore, the adaptive watermark embedding technique refines the traditional single-stage quantizer, resulting in enhanced robustness with reduced reversible watermarks, achieving a balance between robustness and invisibility.

The structure of this paper is as follows: Section 2 reviews the existing two-stage RRW framework and introduces PHT moment and STDM technology. Section 3 details the proposed video-oriented dual-domain watermarking-CG scheme based on PHT moments and optimized STDM. Experimental results and analysis are presented in Sections 4 and 5, followed by conclusions in Section 6.

2 Related Work

This section briefly describes the origin and thinking of the two-stage RRW framework for digital images, and introduces GOP, PHT moment and STDM technology as the basic knowledge.

2.1 RRW Framework for Digital Images

Coltuc et al. [9,31] introduced the initial two-stage RRW approach, where robust watermark bits are embedded into the DCT coefficients of a grayscale image to create an intermediate image marked with the robust watermark. The residual information between the original image and the marked intermediate image is then calculated and reversibly embedded back into the original image

pixels. This process aims to mitigate the impact of robust watermark embedding on the grayscale image. However, a challenge arises as the direct embedding of the reversible watermark can introduce distortion to the pre-existing robust watermark in the image, potentially compromising its robustness due to the interdependent embedding domains. To address this issue, Wang et al. [32] proposed a two-stage RRW method employing independent embedding domains. By utilizing Haar wavelet transform, the grayscale image is partitioned into high-frequency and low-frequency independent domains for embedding the robust and reversible watermarks, respectively. This approach severs the connection between embedding domains, thereby eliminating mutual influence between the two sets of embedded information. The two-stage RRW embedding framework by Coltuc and Wang is illustrated in Fig. 1.

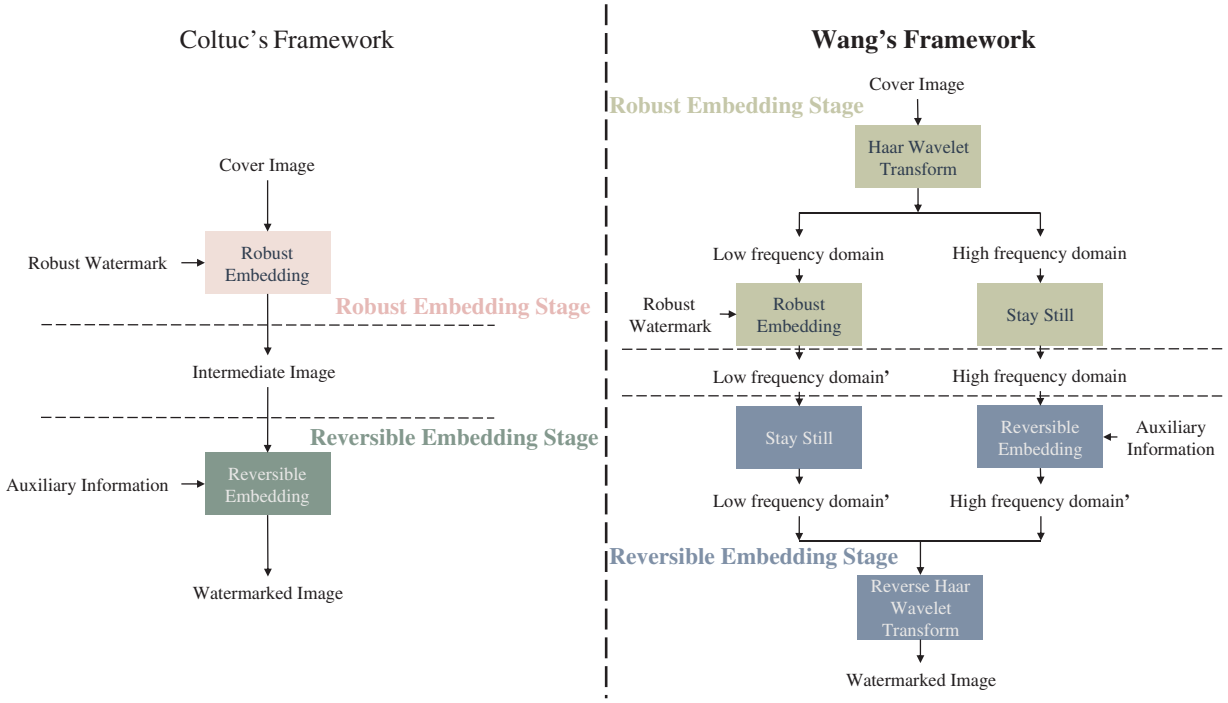


Figure 1: Coltuc's two-stage RRW framework *vs*. Wang's two-stage RRW framework

Building upon Wang et al.'s notion of independent embedding domains, we have devised a novel dual-domain watermark embedding framework tailored for digital videos. This framework entails the selection of video frames from short videos, application of PHT on these frames, normalization of resulting PHT moments, and embedding of robust watermarks using optimized STDM. By incorporating these steps, our approach not only facilitates attack detection authentication in short videos but also allows for video restoration, rendering our proposed solution highly resilient against CSP and GD attacks.

2.2 Group of Pictures

The concept of Group of Pictures (GOP) [33] is fundamental in digital video coding, delineating a continuous scene within a video sequence. Each GOP comprises a sequence of reference frames and non-reference frames that depict the video's ongoing motion and alterations. Typically, a GOP consists of I frames, P frames, and B frames, whose interrelationships and ordering are pivotal in the video compression and decoding processes.

Intra Frame (I frame): An I frame serves as a standalone reference frame containing complete image information, independent of other frames for decoding. I frames are often utilized as starting or transitional points in a video sequence, preserving the foundational structure and content information.

Predictive Frame (P frame): A P frame acts as a predictive frame, leveraging previously decoded I or P frames to forecast picture content in the current frame. P frames store solely residual information pertinent to the prediction, effectively reducing the video file size.

Bi-directional Frame (B frame): The B frame, a bidirectional predictive frame, predicts the current frame's image content utilizing preceding and subsequent I or P frames. Integration of B frames optimize inter-frame correlation, enhancing video compression rates and quality.

The MP4 (H.264) video format [33] is widely adopted in digital video contexts. By tactically configuring the GOP structure and interval, video encoding efficiency can be enhanced, facilitating superior compression and transmission. Optimal selection of GOP size and frame type sequence ensures a harmonized blend of video encoding compression rates, image quality, and real-time performance, thereby elevating video encoding efficacy and user experience.

In our approach, we identify a single frame exhibiting the highest picture smoothness from each GOP as the dual-domain watermark embedding domain corresponding to that GOP. This strategy, while upholding a balance between invisibility and robustness, lessens distortion resulting from attacks on specific frame types, substantially fortifying the watermark's security.

2.3 Polar Harmonic Transform

Polar Harmonic Transform (PHT) refers to a set of three orthogonal moments grounded in harmonic functions, as introduced by Yap et al. [34]. These orthogonal moments encompass the polar complex exponential transform (PCET), polar cosine transform (PCT), and polar sine transform (PST). Notably, PHT is characterized by its exclusion of complex factorials, gamma terms, and extensive summations, contributing to favorable computational complexity and data stability. PHT offers several advantages over common frequency domain transformations like the Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT). It possesses enhanced robustness to image and video manipulations, such as compression and resizing, ensuring embedded watermarks remain detectable. Its superior frequency localization enables precise watermark embedding and minimizes the risk of unauthorized removal.

Additionally, PHT supports multiscale analysis, allowing efficient representation at various resolutions, which is particularly beneficial for adaptable video applications. Its low computational complexity, especially compared to DWT, is significant for real-time processing without sacrificing quality. Furthermore, PHT achieves effective energy compaction, optimizing frequency component usage for discreet watermarking with minimal visual impact. This section primarily elucidates the mathematical formulation of PHT. Given a video frame of size $H \times H$ within a brief video, its corresponding frame function is denoted as $F(x, y)(x, y \in [1, H])$. The projection of this frame function onto the unit circle yields $F(x_s, y_t)({x_s}^2 + {y_t}^2 \leq 1)$. $F(x_s, y_t)$ can be represented as a linear combination of the PHT moment $M_{n,k}$ and the PHT basis function $V_{nk}(x_s, y_t)$, as depicted in Eq. (1).

$$F(x_s, y_t) = \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} M_{n,k} V_{n,k}(x_s, y_t) \tag{1}$$

where $M_{n,k}$ is the nth-order PHT moment with k repetitions, which is composed of the inner product of $F(x_s, y_t)$ and $V_{n,k}(x_s, y_t)$, as shown in Eq. (2).

$$M_{n,k} = \sum_{s=0}^{H-1} \sum_{t=0}^{H-1} F(x_s, y_t) [V_{n,k}(x_s, y_t)]^* \Delta x_s \Delta y_t \tag{2}$$

where $()^*$ represents the complex conjugate operation, $(x_s, y_t) = \left(\frac{2s+1-H}{H}, \frac{2t+1-H}{H}\right)$, $\Delta x_s = \Delta y_t = \frac{2}{H}$. Since the polar function is defined on the unit circle, its function $V_{n,k}(x_s, y_t)$ can be decomposed into radial basis function $R_n(r)$ and angular basis function $A_k(\theta) = e^{k\theta \cdot i}$, as shown in Eq. (3).

$$V_{n,k}(x_s, y_t) = R_n(r) A_k(\theta) \tag{3}$$

where $r = \sqrt{x_s^2 + y_t^2}$, $\theta = tan^{-1}\left(\frac{y_t}{x_s}\right)$

The radial basis functions for PCET, PCT, and PST are defined [25] as shown in Eqs. (4) to (6).

$$R_n^{PCET}(r) = \frac{1}{\sqrt{\pi}} e^{2\pi n r^2 \cdot i} \tag{4}$$

$$R_n^{PCT}(r) = \begin{cases} \frac{1}{\sqrt{\pi}}, n = 0 \\ \sqrt{\frac{2}{\pi}} \cos(\pi n r^2), n > 0 \end{cases} \tag{5}$$

$$R_n^{PST}(r) = \sqrt{\frac{2}{\pi}} \sin(\pi n r^2) \tag{6}$$

We integrate the attributes of MP4 (H.264), a prevalent format for short videos [33], and apply PHT to the video frames to extract the PHT moments within the unit circle as the embedding domain for the robust watermark. This methodology enables the capture of enhanced features and structures of the frame in polar coordinates, reinforcing its defense against a wider range of gradient descent attacks, thereby aligning with the requisites of our proposed solution.

2.4 Spread Transform Dither Modulation

Chen et al. [35] introduced two robust watermarking techniques: Quantization Index Modulation (QIM) and Spread Transform Dither Modulation (STDM). Both methods embed watermarks through the quantization of the jitter signal. STDM, an advancement of QIM, disperses a resilient watermark bit across multiple carriers to enhance robustness against random noise attacks. In STDM, the vector x earmarked for embedding in the polar coordinates of the video frame is randomly projected onto a specific direction vector u. Following this, the robust watermark bit is embedded by perturbing the vector x, transformed into a scalar through dithering, as illustrated in Eq. (7).

$$y = x + \left(Q^w\left(x^T u, \Delta\right) - x^T u\right) \cdot u, w \in \{0, 1\} \tag{7}$$

where $()^T$ represents the transposition operation and Δ is the quantization step size. $Q^w()$ is a single-stage quantizer, which is defined as shown in Eq. (8).

$$Q^w\left(x^T u, \Delta\right) = \left[\frac{x^T u + \beta(w)}{\Delta}\right] \Delta - \beta(w) \tag{8}$$

where [] represents the rounding operation, $\beta(w)$ is the dither value corresponding to the watermark bit w, and $\beta(1) = \beta(0) + \frac{\Delta}{2}$. When extracting the robust watermark bit, STDM extracts the robust watermark from the robustly watermarked video by using minimum distance decoding, as defined in Eq. (9).

$$w' = \arg\min_{w\in\{0,1\}} \left|(y')^T u - Q^w\left((y')^T u, \Delta\right)\right| \tag{9}$$

where w' is the extracted robust watermark bit, and y' is the robust watermark vector extracted from the robustly watermarked video.

3 A Dual-Domain Video Watermarking Scheme Based on Adaptive Normalized PHT Moment and Optimized STDM

This scheme presents a dual-domain watermarking framework for digital videos, integrating adaptive normalization and enhanced Spread Transform Dither Modulation (STDM) techniques. The framework consists of three primary components: dual-domain watermark embedding, attack detection and authentication, and watermark extraction with video reconstruction.

In the initial embedding phase, both robust and reversible watermarks are incorporated into the original video through distinct embedding processes. The subsequent attack detection and authentication phase assesses the integrity and distortion of the marked video, determining whether it has been subjected to external attacks. Finally, in the watermark extraction and video reconstruction stage, if the video remains undamaged, the dual-domain watermark is retrieved. Conversely, in cases of attack or content loss, only the robust watermark can be extracted. This section outlines the embedding algorithm framework, describes the key frame selection algorithm, and provides an overview of the components involved.

3.1 Video-Oriented Dual-Domain Watermark Embedding Framework

Leveraging the RRW framework designed for digital images, we have developed a dual-domain watermarking-CG framework specifically for digital videos. By utilizing Polar Harmonic Transform (PHT), key frames are mapped onto the unit circle in polar coordinates. The interior of the circle represents the low-order section, characterized by smooth frequency transitions and containing the key frame's essential structural elements, which serves as the embedding domain for the robust watermark, emphasizing stability and imperceptibility. Conversely, the exterior region encompasses high-order details and textures, providing a suitable space for the reversible watermark due to its ample coefficient availability. The embedding process consists of two stages: the robust embedding stage and the reversible embedding stage, as illustrated in Fig. 2.

Initially, PHT is applied to all frames within the key frame set X to derive the PHT moment $M_{n,k}$ for dual-domain watermark embedding. In the robust embedding phase, $M_{n,k}$ is segmented into a low-order region, $M_{nL,kL}$ (representing the texture component of the inscribed circle), and a high-order region. The suitable PCET moment, PCT moment, and PST moment are extracted from $M_{nL,kL}$ and normalized adaptively. Subsequently, the robust watermark W is embedded into the inscribed circle of the PHT domain using optimized STDM, generating a quantization error d_q. The inverse PHT process is then applied in conjunction with the high-order region to produce the robust watermark key frame set X_c'.

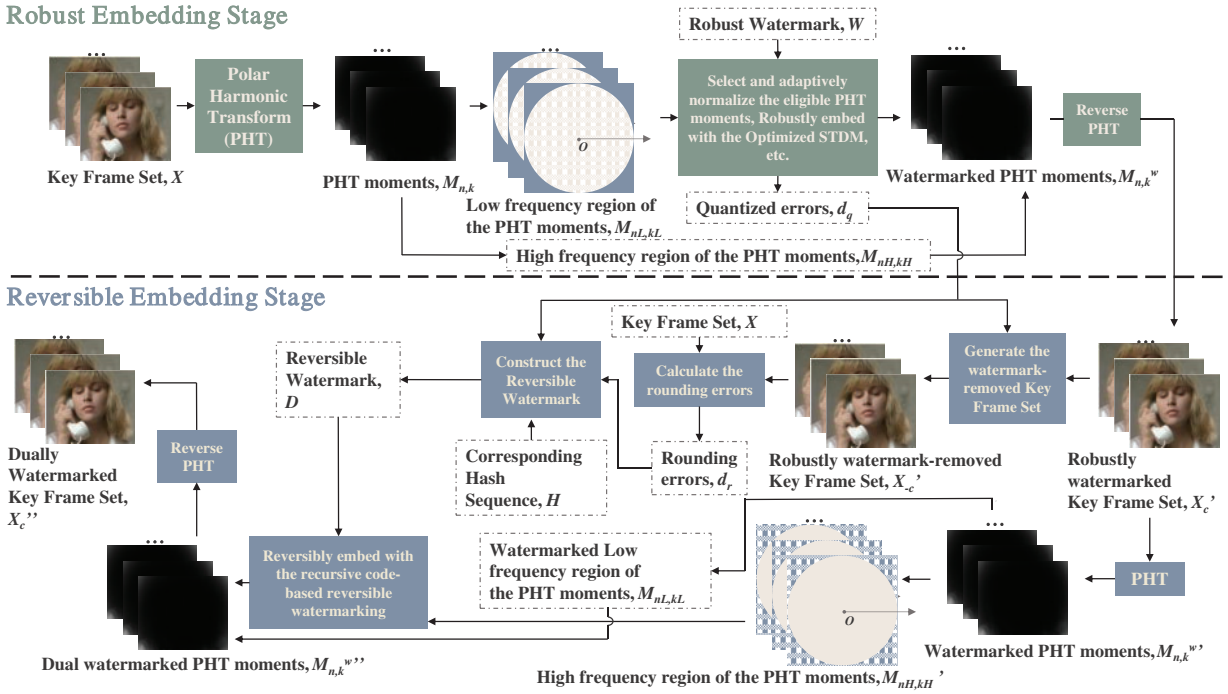


Figure 2: Video-oriented dual-domain watermark embedding framework

In the reversible embedding stage, the key frame X'_{-c} devoid of the robust watermark is created utilizing X'_c and the quantization error d_q from the robust embedding phase, facilitating the computation of the rounding error d_r. Following this, the position length of the LSBs for embedding the hash sequence H is determined, and the reversible watermark D is formed by merging the two errors. Subsequently, D is embedded into the high-order region, $M_{nH,kH}$ (located outside the inscribed circle but within the rectangle) of X'_c, employing recursive code-based reversible watermarking technology. The resulting reversible watermarked region and robust watermarked region undergo inverse PHT collectively to yield the dual-domain watermarked keyframe set X''_c. It is pertinent to highlight that this framework entails procedures such as adjusting the robust embedding strength by simulating multiple attacks, with detailed specifics elaborated in Section 2.3.

3.2 Keyframe Selection Algorithm

3.2.1 Design of Key Frame Selection Algorithm

This study introduces a key frame selection algorithm (KSA) based on the GOP in the context of H.264 video coding standard. By processing digital video data, all GOP groups are extracted, and the key frame characterized by the smoothest transitions is chosen as the optimal location for watermark information embedding, as depicted in Fig. 3.

Utilizing the classic test video Suzie as a case study, the initial step involves segmenting the video to extract individual video frames. These frames are then categorized into I frames, P frames, and B frames based on intra-frame and inter-frame prediction, leading to the formation of multiple Group of Pictures (GOPs). Subsequently, the Y channel components and U channel components of all frames within the same GOP are selected, and the difference between the adjacent bit planes of these two

channels is computed. This difference for each frame is represented as a histogram difference, and the cumulative difference calculation is outlined in Eqs. (10) and (11).

$$D_n^{\ Y} = \sum_{l=1}^{L} \left| HV_n^{\ Y}(l) - HV_n^{\ Y}(l+1) \right| \tag{10}$$

$$D_n^{\ U} = \sum_{l=1}^{L} \left| HV_n^{\ U}(l) - HV_n^{\ U}(l+1) \right| \tag{11}$$

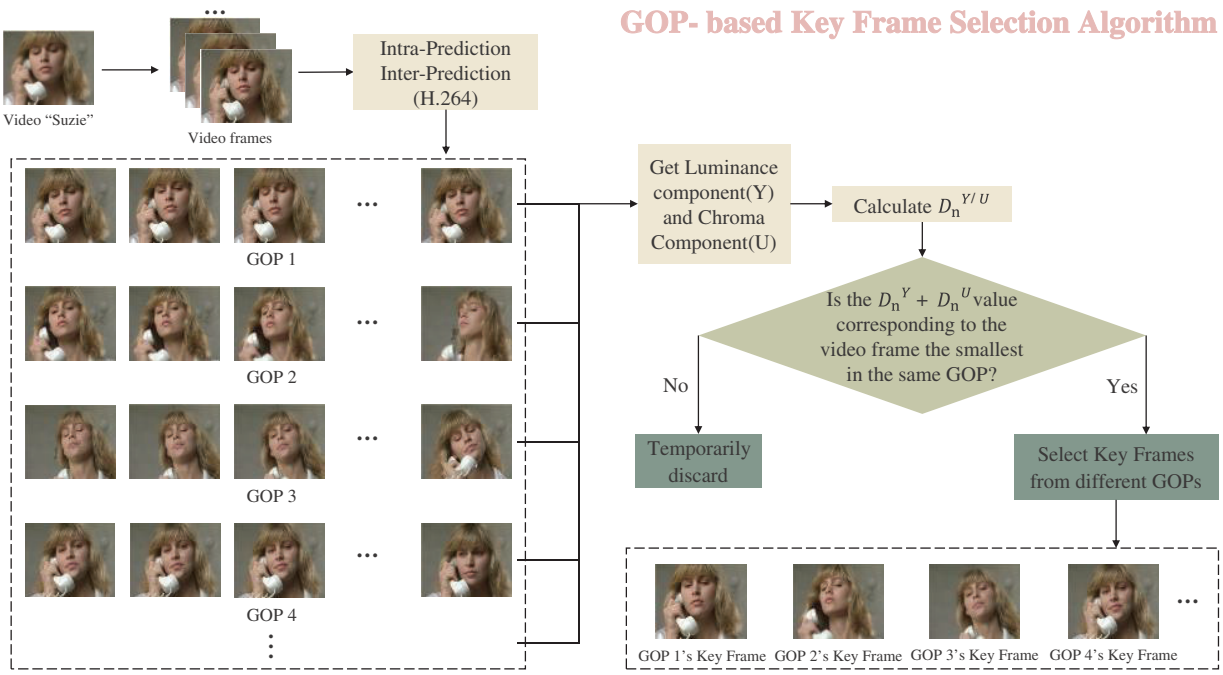


Figure 3: Framework of GOP-based key frame selection algorithm

In this context, $D_n^{\ Y}$ denotes the difference of the Y channel histogram for the nth frame within the same GOP, while $D_n^{\ U}$ signifies the disparity of the U channel histogram for the nth frame in the identical GOP. Additionally, $HV_n^{\ Y}$ corresponds to the Y channel histogram value and $HV_n^{\ U}$ corresponds to the U channel histogram value of the nth frame within the GOP. Here, L is defined as the total number of bit planes, with a fixed value of 8.

The aggregation of histogram differences between adjacent bit planes within a video frame provides valuable insights into texture variations in the corresponding scene [29]. This metric is instrumental in identifying the frame with the smoothest texture within the same Group of Pictures (GOP), which is subsequently selected as the key frame for watermark embedding. The selection rationale is based on the representation of pixel value changes through bit plane disparities; a lower sum of histogram differences indicates minimal variations between neighboring pixels, resulting in smoother frame details and enhanced concealment capacity for the embedded watermark. Thus, the key frame is determined by finding the video frame with the smallest difference between the Y channel and U channel within the same GOP, as shown in Eq. (12).

$$D_n = D_n^{\ Y} + D_n^{\ U} \tag{12}$$

In the context of the grouped GOPs, let the nth frame represent the frame with the minimal difference, denoted as D_n, between the Y channel histogram and U channel histogram. By comparing the cumulative histogram differences of each frame within the grouped frame, the frame exhibiting the least sum of histogram differences is designated as the optimal embedding frame for the dual-domain watermarks W and M.

3.2.2 Contribution of KSA to the Proposed Algorithm

Taking the Y channel as an example, the Y channel of a frame is represented as a matrix, and each element of the matrix represents the Y channel value of a pixel, as shown in Eq. (13).

$$I(i,j) = \prod_{l=1}^{L} b_l(i,j) \tag{13}$$

The binary representation of a specific pixel from the 8th to the 1st bit plane is denoted as b_8 to b_1. To compute the inter-frame difference $D_n{}^Y$, the difference value for the Y channel of each pixel is derived by defining the difference matrix DM, as depicted in Eq. (14).

$$DM(i,j) = \sum_{l=1}^{L} |b_{l+1}(i,j) - b_l(i,j)| \tag{14}$$

where $DM(i,j)$ represents the difference between the pixel in the ith row and jth column of the difference matrix. As per the definition of geometric interval [36], the difference matrix DM is interpreted as the geometric distance and is normalized to DM', which signifies the distance from the sample point to the hyperplane in the feature space. The normalization process is detailed in Eq. (15).

$$DM' = w^T X + \frac{B}{\|w\|} \tag{15}$$

Within the key frame selection algorithm, where w denotes the weight vector, B is the feature vector, and DM' represents the bias term, pixels with $DM' > 0$ are categorized as positive, while those with $DM' < 0$ are classified as negative. For effective classification, proximity between the model's predicted outcomes and the actual results is crucial. It indicates a precise division between positive and negative classes, highlighting the proximity of each sample point to the hyperplane [36]. A smaller sum of differences between adjacent bit planes in the Y and U channels indicates closer proximity between the predicted and actual outcomes, resulting in reduced distances between pixel types and the hyperplane. This underlines the correlation between a concentrated distribution of Y channel values and smoother image quality, a principle applicable to the U channel as well. In essence, a frame within the same GOP featuring minimal differences between adjacent bit planes in the Y and U channels exhibits smoother attributes, designating it as the key frame for the dual-domain watermarking scheme. Opting for a key frame with gradual transitions aids in embedding the watermark effectively. Ultimately, the scheme's efficacy lies in selecting smoothly transitioning key frames, enhancing its overall performance. The contributions of KSA to the proposed scheme are summarized as follows:

1. Enhanced Invisibility: Smoothed video frames maintain visual integrity, minimizing perceptible distortions and ensuring that the embedded watermark information seamlessly blends with keyframes without compromising their quality.

2. Improved Robustness: The gradual transitions in pixel values facilitate easier extraction of the watermark from smoothed video frames. This enhancement boosts the accuracy and efficiency of the

extraction process, fortifying the resilience of the watermark information and mitigating potential losses from transmission errors or image processing.

3.3 Dual-Domain Watermark Embedding for Digital Video

The dual-domain watermark embedding process comprises two key stages: the robust embedding stage and the reversible embedding stage. The robust embedding stage is primarily implemented to safeguard video copyrights, while the reversible embedding stage serves the dual purposes of attack detection authentication and original video recovery. Within the reversible watermark, components include the quantization error resulting from watermark embedding, the rounding error in key frame reconstruction, and the integration of the least significant bit (LSB) for attack detection authentication.

3.3.1 Robust Watermark Embedding Stage

The robust embedding stage encompasses the PHT moment calculation, PHT moment selection, adaptive moment normalization, and adaptive watermark embedding components. Within this process, common target attacks serve as the foundation for both the adaptive normalization of moments and robust watermark embedding stages, with each having distinct methodologies and objectives in their respective attack simulation tests.

In the adaptive normalization of moments, the stability of PHT moments against target attacks is analyzed under various conditions of orders and repetitions through extensive attack simulations on digital videos. The normalization weight is adaptively determined based on the stability observed. On the other hand, the robust watermark embedding stage conducts attack simulations on individual digital videos to be embedded, adjusting the optimal embedding strength of each robust watermark bit based on the ensuing watermark detection performance. Balancing robustness and computational complexity are essential in this stage, and the selection of key attacks representative of common signal processing (CSP) attacks, such as AWGN, JPEG2000, and mean filtering, ensures strong robustness while reducing computational overhead significantly.

By strategically choosing representative CSP attacks to replace a multitude of simulated attacks, the robustness of the embedded watermark is effectively maintained, while computational complexity is minimized, achieving an optimal performance equilibrium. The detailed process of the robust watermark embedding stage is illustrated in Algorithm 1.

Algorithm 1: Robust embedding ()

Input: Robust Watermark W, Original Key Frames X;
Output: Robust Watermarked Key Frames $X_c^{'}$, Quantized Errors d_q;

1. $M_{n,k} = \sum_{s=0}^{H-1} \sum_{t=0}^{H-1} F_X(x_s, y_t)\,[V_{n,k}(x_s, y_t)]^* \Delta x_s \Delta y_t$.// **Calculate PHT moments.**
2. $M = Key \otimes \left\{ {M_{n,k}}^{PCET}, {M_{n,k}}^{PCT}, {M_{n,k}}^{PST} \right\} = \{M_{n1,k1}, M_{n1,k1}, \ldots, M_{nL_P,kL_P}\}$.// **Select the eligible PHT moments.**
3. ${M_{ni,ki}}^{R} = \dfrac{M_{ni,ki}}{M_{0,0}} \times \dfrac{10^3}{WT_{ni,ki}}$.// **Adaptively normalize every eligible PHT moment.**
4. $M_i^{Rw} = M_i^R + \left(Q_z^w \left(M_i^{RT} u, \Delta\right) - M_i^{RT} u\right) u, w \in \{0, 1\}$.// **Adaptively embed W by bit.**
5. $d_q = \sum_{j=1}^{total\ W\ bits} \left(Q_z^w \left(M_j^{RT} u, \Delta\right) - M_j^{RT} u\right)$.// **Calculate the total quantized errors.**
6. ${M_{ni,ki}}^{w} = {M_{ni,ki}}^{Rw} \times M_{0,0} \times \dfrac{WT_{ni,ki}}{10^3}$.// **Inverse normalize the normalized M_i^{Rw}.**

(Continued)

Algorithm 1 (continued)

7. $X_c^{'} = X + X_{c1} = X + \left[\sum_{i=1}^{L_p}\left(\left(M_{ni,ki}{}^{w} - M_{ni,ki}\right)V_{ni,ki} + \left(M_{ni,ki}{}^{w*} - M_{ni,ki}{}^{*}\right)V_{ni,ki}{}^{*}\right)\right]$.// **Yield the candidate watermarked key frames.**
8. AWGN ($\sigma^2 = 0.03$), JPEG2000 ($C_r = 100$), mean filtering ($W_s = 5 \times 5$) $\rightarrow X_c^{'}$.// **Perform three attack simulations.**
9. Calculate watermarked PHT moments, repeat Step 1.
10. Select the watermarked eligible PHT moments, repeat Step 2.
11. Normalize every watermarked eligible PHT moment, repeat Step 3.
12. $\widetilde{w}_i = \arg\min_{\widetilde{w}_i \in \{0,1\}} \left| \widetilde{M_i^{RwT}} u - Q^w\left(\widetilde{M_i^{RwT}} u, \Delta\right)\right|$.// **Detect the Robust Watermark.**
13. If $\widetilde{w}_i \neq w_i$ and $z \leq Z$:
    Record the positions of the incorrect watermark bits and increase the z at these positions, re-embed these bits and repeat Steps 9∼13 till $\widetilde{w}_i = w_i$.
14. If $\widetilde{w}_i = w_i$:
    Take the corresponding candidate watermarked key frames $X_c^{'}$ as the final Robust Watermarked Key Frames, yield the corresponding total Quantized Errors d_q.

In Step 2, the restrictions on the order n and the number of repetitions k for the three combination moments corresponding to PHT differ [34]. To ensure these combination moments fall within a reasonable range, we define their ranges as follows: $\{M_{n,k}{}^{PCET}, 0 \leq n \leq N, k \neq 4l, l \in \mathbb{Z}\}$, $\{M_{n,k}{}^{PCT}, 0 \leq n \leq N, k \geq 0, k \neq 4l, l \in \mathbb{Z}\}$, $\{M_{n,k}{}^{PST}, 1 \leq n \leq N, k \geq 0, k \neq 4l, l \in \mathbb{Z}\}$. Concurrently, a random index Key is utilized to select the embedding moment from combination moment set $\left\{M_{n,k}{}^{PCET}, M_{n,k}{}^{PCT}, M_{n,k}{}^{PST}\right\}$, with the number of indices determined by the number of bits W. In Step 3, $WT_{ni,ki}$ represents the adaptive weight and is calculated as per Eq. (16).

$$WT_{ni,ki} = a_1 + a_2 n_i + a_3 l_i + a_4 n_i^2 + a_5 n_i l_i + a_6 l_i^2 + a_7 n_i^3 + a_8 n_i^2 l_i + a_9 n_i l_i^2 + a_{10} l_i^3 \tag{16}$$

$M_{0,0}$ represents the zero-order zero-repetition moment, with 10^3 denoting a constant. Despite the adaptive normalization of PHT moments, variations in the stability of normalized PHT moments of the same order and repetitions across different keyframes persist when confronted with identical attacks, often stemming from fixed embedding strengths. To address this issue, we introduce an embedding strategy that relies on simulated attack tests to dynamically determine the appropriate embedding strength.

Initially, the watermark bit is embedded with a lower strength to generate a potential robust watermark keyframe. Subsequently, a representative CSP attack is executed on the frame. Successful extraction of the watermark bit from the attacked frame indicates the optimal embedding strength. Should extraction fail, the embedding strength is incrementally increased until successful extraction is achieved for all robust watermark bits. This approach ensures optimal robustness of the video while maintaining invisibility.

In Step 4, building upon the aforementioned strategy, we introduce a multi-level Spread Transform Dither Modulation (STDM) with Z levels of embedding strength to enhance both robustness and invisibility. Higher levels correspond to greater embedding strength. Following the embedding strategy, the 1-bit robust watermark bit is inserted into the P-bit PHT matrix through the refined STDM approach, denoted as $M_i^R = \{\left|M_{n((i-1)P+1),k((i-1)P+1)}{}^{R}\right|, \left|M_{n((i-1)P+2),k((i-1)P+2)}{}^{R}\right|, \ldots \left|M_{n(iP),k(iP)}{}^{R}\right|\}$. Here, $()^T$ signifies the transposition operation, and $Q_z{}^{w}()$ represents the multi-level STDM under the zth embedding strength level, as detailed in Eq. (17).

$$Q_z{}^{w}\left(M_i{}^{RT}u,\Delta\right)=\begin{cases}Q^{w}\left(M_i{}^{RT}u,\Delta\right)-\dfrac{Z-z}{Z}\cdot\dfrac{\Delta}{4}-\left|\left[M_i{}^{RT}u\right]-M_i{}^{RT}u\right|,\left|M_i{}^{RT}u-Q^{w}\left(M_i{}^{RT}u,\Delta\right)\right|>\dfrac{\Delta}{4}\,and\,M_i{}^{RT}u\le Q^{w}\left(M_i{}^{RT}u,\Delta\right)\\ Q^{w}\left(M_i{}^{RT}u,\Delta\right)+\dfrac{Z-z}{Z}\cdot\dfrac{\Delta}{4}+\left|\left[M_i{}^{RT}u\right]-M_i{}^{RT}u\right|,\left|M_i{}^{RT}u-Q^{w}\left(M_i{}^{RT}u,\Delta\right)\right|>\dfrac{\Delta}{4}\,and\,M_i{}^{RT}u>Q^{w}\left(M_i{}^{RT}u,\Delta\right)\\ M_i{}^{RT}u,\qquad\qquad otherwise\end{cases}\tag{17}$$

The optimized Spread Transform Dither Modulation (STDM) differs from traditional STDM as it operates as a multi-level quantizer, aligning with the diverse embedding strength specifications of the embedding strategy. Varied quantization step sizes are associated with different embedding strengths, where larger step sizes indicate higher embedding strengths and vice versa. This adaptive feature enables the optimized STDM to effectively allocate the optimal embedding strength to each watermark bit, enhancing robustness while preserving invisibility. The fundamental principles of traditional STDM and optimized STDM are depicted in Fig. 4.

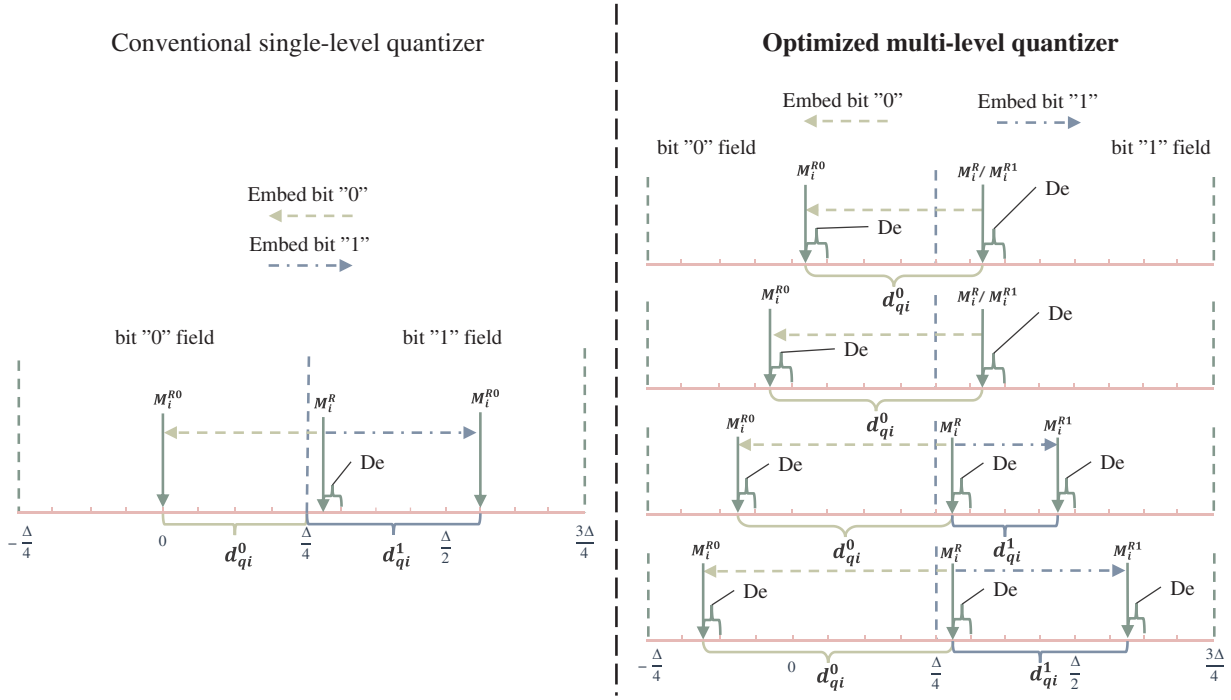


Figure 4: Optimized STDM-based watermarking method *vs*. conventional STDM-based watermarking method. The conventional single-level quantizer with $\beta_i(0)=0$ and $\beta_i(1)=\dfrac{\Delta}{2}$, where $\Delta=16$. The optimized multi-level quantizer that makes the $d_{qi}{}^{j}$ to be an integer. The *De* in the figure denotes the decimal part of the normalized PHT moment, which corresponds to the $\left[M_i{}^{RT}u\right]-M_i{}^{RT}u$ terms in Eq. (17). $M_i{}^{Rw}=\{M_i{}^{R0},M_i{}^{R1}\}$. $d_{qi}{}^{j}=\{d_{qi}{}^{0},d_{qi}{}^{1}\}$

It is essential to highlight that the key frame of the candidate robust watermark denotes the frame under evaluation. This frame undergoes scrutiny to determine if its current robust embedding strength suffices and if it qualifies as the final robust watermark key frame post the simulated attack test, serving as input for subsequent stages. In Step 5, the computed quantization error d_q is integrated outside the unit circle as part of the reversible watermark. Step 6 involves the reverse process of Step 2, where the normalized robust watermark mark moment is reciprocally normalized to produce the potential

robust watermark mark key frame. Here, $M_{ni,ki}{}^{w}$ represents the PHT moment of the potential robust watermark mark key frame.

In Step 7, $M_{ni,ki}{}^{*}$ is the conjugate of $M_{ni,ki}$, with PCET component moment denoted as $M_{-ni,-ki}$, while both the PCT and PST component moments are represented by $M_{ni,-ki}$. $M_{ni,ki}{}^{w*}$ is the conjugate of $M_{ni,ki}{}^{w}$, the PCET component moment is $M_{-ni,-ki}{}^{w}$, and both the PCT and PST component moments are labeled as $M_{ni,-ki}{}^{w}$. The generation of the compensated key frame X_{c1} leads to the creation of the candidate robust watermark key frame, resolving the issue of frame visual quality deterioration that may arise from directly utilizing the PHT moment for key frame reconstruction [25].

Steps 8 to 14 outline the process of the attack simulation test in this phase. Various attacks, including AWGN with a variance of $\sigma^2 = 0.03$, JPEG2000 compression with a ratio of $C_r = 100$, and mean filtering with a window size of $W_s = 5 \times 5$, are applied to assess the resistance of the video post robust watermark embedding on a bit-by-bit basis. If the extracted robust watermark bits remain unchanged compared to the original watermark bits under the attack scenario, the key frame marked with the candidate robust watermark at that moment becomes the final robust watermark key frame. However, if differences persist and the embedding strength has not peaked at Z, the embedding strength value is escalated by one level z, and the robust watermark bits are re-embedded into the original key frame. This process iterates until the condition of parity between both watermark bits is achieved.

3.3.2 Reversible Watermark Embedding Stage

The reversible embedding stage mainly includes four parts: de-watermarking video generation, reversible watermark construction, reversible watermark embedding, and attack detection authentication sequence embedding, which are introduced in detail through Algorithm 2 below:

Algorithm 2: Reversible embedding ()

Input: Robust Watermarked Key Frames $X_c^{'}$, Quantized Errors d_q;
Output: Dual Watermarked Key Frames $X_c^{''}$;

1. $M_i{}^{wr} = \sum_{s=0}^{H-1} \sum_{t=0}^{H-1} F_X{}^{w}(x_s, y_t) [V_{n,k}{}^{w}(x_s, y_t)]^{*} \Delta x_s \Delta y_t$.
2. $M_i{}^{Rwr} = \dfrac{M_i{}^{w}}{M_{0,0}} \times \dfrac{10^3}{WT_i}$.// **Calculate the normalized PHT moments of $X_c^{'}$.**
3. $M_i{}^{RrT}u = M_i{}^{RwrT}u - d_{qi}$.// **Calculate the normalized PHT moments of robust-watermarked-removed key frames.**
4. $M_i{}^{RrT}$: $M_{ni,ki}{}^{rT} = M_{ni,ki}{}^{RrT} \times M_{0,0} \times \dfrac{WT_{ni,ki}}{10^3}$.// **Inverse normalize the $M_i{}^{RrT}$.**
5. $X_{-c}^{'} = X_c^{'} + X_{c2} = X_c^{'} + \left[\sum_{i=1}^{L_p} \left(\left(M_{ni,ki}{}^{r} - M_{ni,ki}{}^{wr}\right) V_{ni,ki}{}^{r} + \left(M_{ni,ki}{}^{r*} - M_{ni,ki}{}^{wr*}\right) V_{ni,ki}{}^{r*}\right)\right]$.// **Generate the robust-watermark-removed key frames.**
6. $d_r = X - X_{-c}^{'}$.// **Calculate the rounding errors.**
7. $D = \{d_q, d_r, b_l\}$.// **Construct Reversible Watermark, where b_l is the LSB of the corresponding predefined pixel in the key frames used to store H.**
8. $X_c^{'} + D \xrightarrow{\text{recursive code-based reversible embedding}} X_{cd}^{'}$.// **Embed the Reversible Watermark.**
9. $\left\{L_h | L_h \in X_{cd}^{'}\right\} \rightarrow \{0\}$, $X_{cd}^{'} \rightarrow X_{ia}^{'}$.
10. $H = Hash(X_{ia}^{'})$, $\{L_h | L_h = H\}$, $X_{ia}^{'} \rightarrow X_c^{''}$.// **Insert Integrity Authentication Sequence and generate Dual Watermarked Key Frames.**

Steps 1 to 5 delineate the process of crafting a de-robust watermark keyframe through the generation of a compensation keyframe X_{c2}. The fundamental objective of producing such a keyframe is to minimize the distinction between the robust watermark keyframe and the original keyframe. Both the embedding end and the extraction end can access the frame via the relevant variables, facilitating the use of the frame to aid in the restoration of the original keyframe. The normalized PHT moment M_i^{RrT} of the de-robust watermark keyframe is derived from the robust embedding stage's output d_q and the normalized PHT moment M_i^{Rwr} associated with the robust watermark keyframe, followed by inverse normalization of the moment. Subsequently, the de-robust watermark keyframe is generated as per Step 5, with detailed keyframe reconstruction elucidated in Section 3.5.

To ensure reversibility, a reversible watermark is essential for reinstating the original keyframe. This watermark encompasses the distortion introduced during the robust watermark embedding stage, encompassing the quantization error (output of the robust embedding stage), rounding error between the original keyframe and the robust watermark keyframe, and the LSBs employed for video attack detection and authentication. The determination of the rounding error d_r and the holistic construction of the reversible watermark are explicated in Steps 6 to 7. Verification of the received video's integrity by the legitimate user involves applying a hash operation to generate a hash sequence H of length L_h, with b_l denoting the predetermined pixel LSBs allocated for storing H in the keyframe, constituting part of the reversible watermark. Embedding the reversible watermark D into the PHT high-order domain corresponding to the keyframe, through recursive coding technology-based reversible watermarking techniques [30], that is, the key frame corresponds to the area outside the unit circle in polar coordinates. It should be noted that the embedded area does not include the first L_h pixels of the key frame, which are used to embed the hash sequence required for attack detection and authentication. Before embedding the hash sequence, it is necessary to set all the first L_h pixels to 0 to obtain the detection frame $X_{ia}^{'}$, and perform a Secure Hash Algorithm-256 (SHA-256) operation on $X_{ia}^{'}$ to obtain a hash sequence H of length L_h bits. Finally, H is directly inserted into L_h to complete the embedding of the final sequence and generate the final dual-domain watermark key frame X_c''.

To address capacity constraints, when the area outside the unit circle is insufficient for accommodating all D, the approach involves utilizing the remaining moments in the unit circle as the embedding domain of D as a compensatory measure. Given the lower embedding strength of reversible embedding in comparison to robust embedding, any excess embedded reversible watermark serves as high-order noise for the robust watermark, minimizing its impact.

3.4 Attack Detection and Authentication of Video Information

To enhance the accuracy of extracting the dual-domain watermark, we introduce an attack detection and authentication mechanism for digital video. This method aims to determine if the received dual-domain watermarked video has been subjected to any malicious tampering by unauthorized parties. The process, outlined in Algorithm 3, involves extracting the LSBs of the initial L_h pixels from all frames within the dual-domain watermark key frame set Y_c''. This extraction yields a bit sequence H' corresponding to each frame. Subsequently, the LSBs of the first L_h pixels in each frame are reset to 0, followed by a SHA-256 operation on the frames post-reset to generate the respective hash sequence H''. If H' perfectly matches H'', it indicates that Y_c'' has not been compromised. Any disparity between H' and H'' signifies potential malicious alterations to Y_c''. It is essential to highlight two rare scenarios: (1) complete equality between H' and H'' and H', and the original H; (2) total alignment between H' and H'', but not with the original H. Although the likelihood of these scenarios is minimal, they are acknowledged for completeness; their occurrence is not factored into the assessment process.

Algorithm 3: Watermark extraction and video reconstruction ()

Input: Dual Watermarked Key Frames Y_c'';

Output: Robust Watermark $\tilde{W}/W$, or Original Key Frame X;

1. $Y_c'' \rightarrow H'$.// **Extract the Hash sequence in the marked frame.**
2. $\{L_h | L_h \in Y_c''\} \rightarrow \{0\}$, $Y_c'' \rightarrow \widetilde{Y_c''}$.// **Set the first L_h pixel values in Y_c'' to 0.**
3. $H'' = Hash(\widetilde{Y_c''})$.// **Calculate the Hash sequence corresponding to $\widetilde{Y_c''}$.**
4. If $H'' \neq H'$, which means Y_c'' has been attacked:

   $\widetilde{M_{n,k}} = \sum_{s=0}^{H-1}\sum_{t=0}^{H-1} F_{\tilde{Y}}(x_s, y_t)[V_{n,k}(\widetilde{x_s, y_t})]\Delta x_s \Delta y_t$.// **Calculate PHT moments.** $\widetilde{M_{n,k}}^{w} = Key \otimes \left\{\widetilde{M_{n,k}}^{PCET}, \widetilde{M_{n,k}}^{PCT}, \widetilde{M_{n,k}}^{PST}\right\} = \{\widetilde{M_{n1,k1}}, \widetilde{M_{n1,k1}}, \ldots, \widetilde{M_{nLp,kL_P}}\}$.// **Select the watermarked PHT moments.**

   $\widetilde{M_{ni,ki}}^{Rw} = \dfrac{\widetilde{M_{ni,ki}}^{w}}{M_{0,0}} \times \dfrac{10^3}{WT_{ni,ki}}$.// **Adaptively normalize the watermarked PHT moments.**

   $\widetilde{w_i} = \arg\min_{\widetilde{w_i} \in \{0,1\}} \left| \widetilde{M_i^{RwT}} u - Q^w(\widetilde{M_i^{RwT}} u, \Delta) \right|$.// **Extract the robust watermark.**
5. If $H'' = H'$, which means Y_c'' has not been attacked:

   $Y_c'' \xrightarrow{\text{recursive code-based reversible extracting}} X_c' + D$.// **Extract the Reversible Watermark D and construct the Robust Watermarked Key Frames.**

   Extract the Robust Watermark W from X_c', follow the Step 4.

   $X_{-c}' = X_c' + X_{c2} = X_c' + [\sum_{i=1}^{Lp}(d_{qi} V_{ni,ki}{}^{r} + d_{qi}{}^{*} V_{ni,ki}{}^{r*})]$.// **Generate the robust-watermark-removed key frames.**

   $X = d_r + X_{-c}'$.// **Construct the original Key Frames.**

3.5 Dual-Domain Watermark Extraction

After completing the attack detection and authentication phase, the legitimate user proceeds with extracting the dual-domain watermark. This extraction process bifurcates based on the condition of the received dual-domain watermarked video Y_c''. In instances where Y_c'' has been compromised by malicious attacks, the reversible watermark distortion prevents accurate extraction, subsequently hindering the restoration of the original keyframe X and, consequently, the full restoration of the original video. In such scenarios, only the robust watermark $\widetilde{W}$ can be successfully extracted. Conversely, if the received video remains intact without any distortions, both the reversible watermark D and the robust watermark W can be extracted as per Step 5 of Algorithm 3, allowing for the restoration of the original video. Specifically, in the presence of an attack on Y_c'', the extraction of the reversible watermark becomes unfeasible, impeding the restoration of the original video while enabling direct extraction of the robust watermark. The detailed process involves calculating the PHT moment $\widetilde{M_{n,k}}^{w}$ of Y_c'', followed by the determination of the corresponding normalization $\widetilde{M_{n,k}}^{Rw}$, culminating in the extraction of the robust watermark $\tilde{W}$.

In contrast, when Y_c'' remains unaffected by attacks, both the reversible watermark and the robust watermark can be extracted without impediments, allowing for the successful restoration of the original video. The procedure involves extracting the reversible watermark D and replacing the initial L_h pixels of all key frames encoded by the dual-domain watermark with the bits from D to derive X_c'.

Subsequently, de-robust watermark key frames X'_{-c} are generated through Step 5 of Algorithm 3 to restore all original key frames, which are then combined with the unembedded frames to reconstruct the original video.

4 Simulation Experiment and Analysis

In this stage, simulation experiments are conducted to evaluate the performance of the proposed scheme. Initially, a series of experiments is carried out to simulate attacks in order to determine the optimal parameter settings. Subsequently, the benefits of the optimized Spread Transform Dither Modulation (STDM) and other innovative technologies in the scheme are assessed. The invisibility of the scheme is then analyzed by evaluating the visual quality of the key frames marked by the dual-domain watermark under the optimal parameter settings. Finally, the scheme's robustness is compared against the most advanced video robust watermarking algorithm.

It is important to highlight that in the adaptive video dual-domain watermarking scheme utilizing the Phase Harmonic Transform (PHT) moment and optimized STDM, the embedding of the dual-domain watermark relies on the PHT moment. The PHT moment, which encompasses the composite moment of PCET, PCT, and PST moments, is essential for simulating and testing the performance of the watermark embedded in each of these moments. For clarity and ease of classification, the experimental section represents the dual-domain watermarking algorithms of the aforementioned three moments as DW-PCET, DW-PCT, and DW-PST, respectively.

4.1 Dataset, Parameters and Comparison Algorithm Settings

To determine the best parameter settings, we randomly selected 10 Quarter Common Intermediate Format (QCIF) videos with a resolution of 176 × 144 and Common Intermediate Format (CIF) videos with the same video content (resolution of 352 × 288) from Derf's collection (Xiph.org: Derf's Test Media Collection, accessed on 15 July 2024) for testing, where the first 150 frames of each video were selected. Through the test experiment, the practicality of the innovative technology in the proposed scheme was verified. In addition, we also selected 10 different QCIF videos and 10 CIF videos in Derf's collection as experimental videos in the robustness analysis comparison experiment. Before actually executing and analyzing the scheme, it is necessary to first use the H.264 video coding standard to encode the selected test videos and experimental videos in H.264. The specific operation is to run the simulated H.264 video encoding based on the MatlabR2021a platform, where the computer processor is an i5 processor, and connect x.264 through the MEX interface. The parameters that need to be set in the proposed scheme include:

1. The maximum order N corresponding to the PHT moment for dual-domain watermark embedding. This determines the maximum bit threshold for dual-domain watermark embedding.

2. Optimize the quantization step size Δ in STDM. This parameter is determined to optimally balance robustness and invisibility.

3. Expansion factor P. The optimal value of this parameter is determined to achieve the best balance between robustness and invisibility when 1-bit robust watermark bit is embedded in P-bit PHT moment.

4. Maximum embedding strength level Z. The robustness of the robust watermark is ensured by adjusting the embedding strength step size between adjacent levels.

5. Adaptive normalization weight $WT_{ni,ki}$. The most stable PHT moment is obtained by obtaining the optimal normalization weight.

The first two parameters are determined by the length of the robust watermark bit string, and the last three parameters are obtained through test experiments. The specific calculation process will be described in detail in the following Sections 3.2 and 3.3. $WT_{ni,ki}$ is different for each robust watermark bit and is therefore not shown here. The results are shown in Table 1.

Table 1: Parameter settings for two different lengths of robust watermark

Methods	Parameter settings	
	1024-bit robust watermark	4096-bit watermark
DW-PCET	$N = 27; \Delta = 60; P = 3; Z = 5;$	$N = 38; \Delta = 44; P = 3; Z = 5;$
DW-PCT	$N = 38; \Delta = 50; P = 3; Z = 5;$	$N = 53; \Delta = 36; P = 3; Z = 5;$
DW-PST	$N = 38; \Delta = 70; P = 3; Z = 5;$	$N = 53; \Delta = 48; P = 3; Z = 5;$

For the invisibility analysis and robustness analysis experiments, two lengths of robust watermark bit strings (1024 bits and 4096 bits) were employed to assess the scheme's performance across different embedding capacities. These bit strings were generated using a pseudo-random number generator [37].

To further assess the robustness of the scheme, comparisons were made with three leading video robust watermarking algorithms, namely Fan et al. [38], Singh et al. [39], and Chen et al. [27], in terms of invisibility and robustness. The experimental videos and robust watermark settings of these comparative methods were specified using the data sets and parameter configurations in this section. It is important to note that Chen-Zernike et al. utilized Zernike moments with limited numerical stability, resulting in constrained embedding capacity. Consequently, in the subsequent robustness comparison experiment involving 4096-bit robust watermark bits, Chen-Zernike was excluded from the analysis.

4.2 Adaptive Normalization Weights

To ensure the robustness and sufficient embedding capacity of the dual-domain watermark, it is crucial to establish strong stability in the adaptive normalized Phase Harmonic Transform (PHT) moment. This is achieved by fitting the normalized moment through the determination of weight $WT_{n,k}$. The polynomial coefficients in Eq. (16) are fitted via attack simulation to optimize the embedding strength. Using a combination of 10,176 × 144 QCIF videos and 10,352 × 288 CIF videos randomly selected from Derf's collection, attacks such as AWGN with variance $\sigma^2 = 0.03$, JPEG2000 with compression ratio $C_r = 100$, and mean filtering with window size $W_s = 5 \times 5$ were applied to determine the best embedding strength. The fitting values of the coefficients in Eq. (16) were obtained using the bi-square polynomial fitting method [40].

PCET Moments:

$$a_1 = 0.04, a_2 = -4.24e-05, a_3 = 5.82e-07,$$

$$a_4 = 3.34e-03, a_5 = -3.15e-05, a_6 = -1.98e-04,$$

$$a_7 = -4.87e-03, a_8 = 5.03e-05, a_9 = 2.99e-09, a_{10} = -3.81e-03.$$

PCT Moments:

$$a_1 = 0.06, a_2 = -1.91e-05, a_3 = -8.79e-04,$$

$a_4 = 6.12e-07, a_5 = 5.83e-05, a_6 = 2.08e-05,$

$a_7 = -5.12e-04, a_8 = -7.57e-03, a_9 = -5.01e-07, a_{10} = -1.42e-07.$

PST Moments:

$a_1 = 0.06, a_2 = -1.64e-05, a_3 = -2.10e-05,$

$a_4 = 4.89e-07, a_5 = 6.31e-07, a_6 = 2.19e-03,$

$a_7 = -4.72e-05, a_8 = -7.18e-09, a_9 = -5.32e-05, a_{10} = -2.94e-07.$

Upon the fitting values of each coefficient in the three moments, the adaptive normalized weights corresponding to each moment were computed using Eq. (16), resulting in the corresponding normalized fitting values. Subsequently, the goodness of fit was assessed by calculating the R-Squared values between these fitting values and the original values of the three moments, which yielded R-Squared values of 0.938, 0.945, and 0.955, respectively. R-Squared serves as a statistical indicator to evaluate the regression model's fit to the observed data: a value closer to 1 indicates a better fit, while a value closer to 0 suggests a poorer fit. These results demonstrate the effectiveness of the proposed scheme in attack simulation fitting, accurately capturing the substantial changes of the three moments during simulated attacks and ensuring good numerical stability. Therefore, the adaptive normalization of the PHT moment based on this scheme is both reasonable and effective. Calculation of the weight $WT_{n,k}$ can be carried out according to the order and number of repetitions of the PHT moment as outlined in Eq. (16).

4.3 Advantages Analysis of the Proposed Scheme

The proposed scheme integrates several cutting-edge technologies and strategies, including optimized STDM, attack simulation fitting strategy, and adaptive normalization of PHT moments. To comprehensively assess the enhancements and benefits brought about by this scheme, we have devised distinct scenarios to evaluate the strengths of these innovative technologies against their traditional counterparts. By comparing the robustness of the innovative technologies with conventional methods under various attack scenarios, we aim to demonstrate the advantages of the proposed scheme. A robust watermark bit string of 1024 bits has been selected for embedding in this study.

4.3.1 Advantage Analysis of Optimized STDM

The proposed scheme designs and uses optimized STDM to embed robust watermarks. The spreading factor P involved is different from QIM. It can be said that when $P = 1$, QIM=STDM. In order to evaluate the robustness advantage of the optimized STDM in the proposed scheme over QIM, we conducted an experimental analysis using DW-PCET as an example. The experimental settings of DW-PCT and DW-PST also refer to this experimental analysis. In the experiment, the spreading factor P is set to three levels, namely 1, 3, and 5. When $P = 1$, it represents QIM technology. When $P = 3$ or 5, it represents the variable spreading technology of optimized STDM. In the experiment, the representative AWGN attack, JPEG attack, JPEG2000 attack, H.264 attack, mean filtering attack and median filtering attack in CG attack are selected to attack the video embedded with dual-domain watermark. The parameter settings are shown in Table 2.

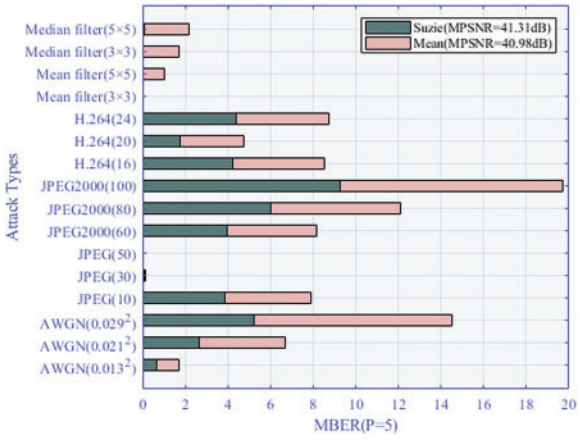
Table 2: Parameter settings for attacks

Attacks	Parameter range	Step
AWGN (σ^2)	0.013∼0.029	0.008
JPEG (Q_f)	10∼50	20
JPEG2000 (C_r)	60∼100	20
H.264 (QP)	16∼24	4
Mean filtering (W_s)	3 × 3, 5 × 5	–
Median filtering (W_s)	3 × 3, 5 × 5	–

Following the attack simulation experiment, the outcomes are summarized in Fig. 5, which presents the Mean Bit Error Rate (MBER) of the robust watermark bit string extracted from the QCIF video "Suzie" when subjected to various attacks, as well as the Mean Peak Signal-to-Noise Ratio (MPSNR) of the video "Suzie" under these conditions. Additionally, it includes the average MBER of the robust watermark bit string extracted from 20 videos, comprising 10 QCIF and 10 CIF videos, tested under the same attack, along with the average MPSNR values of the videos. MBER is computed as the Bit Error Rate (BER) of all key frames bearing robust watermarks across each experimental video, averaged out, while MPSNR is derived by analogy.

(a) P=1 (b) P=3

Figure 5: (Continued)



(c) P=5

Figure 5: Robustness performance of the proposed DW-PCET method with different P in terms of MBER against attacks. The parameter settings for each attack denote AWGN variance, JPEG quality factor, JPEG2000 compression ratio, H.264 quantification step and filtering window size, respectively

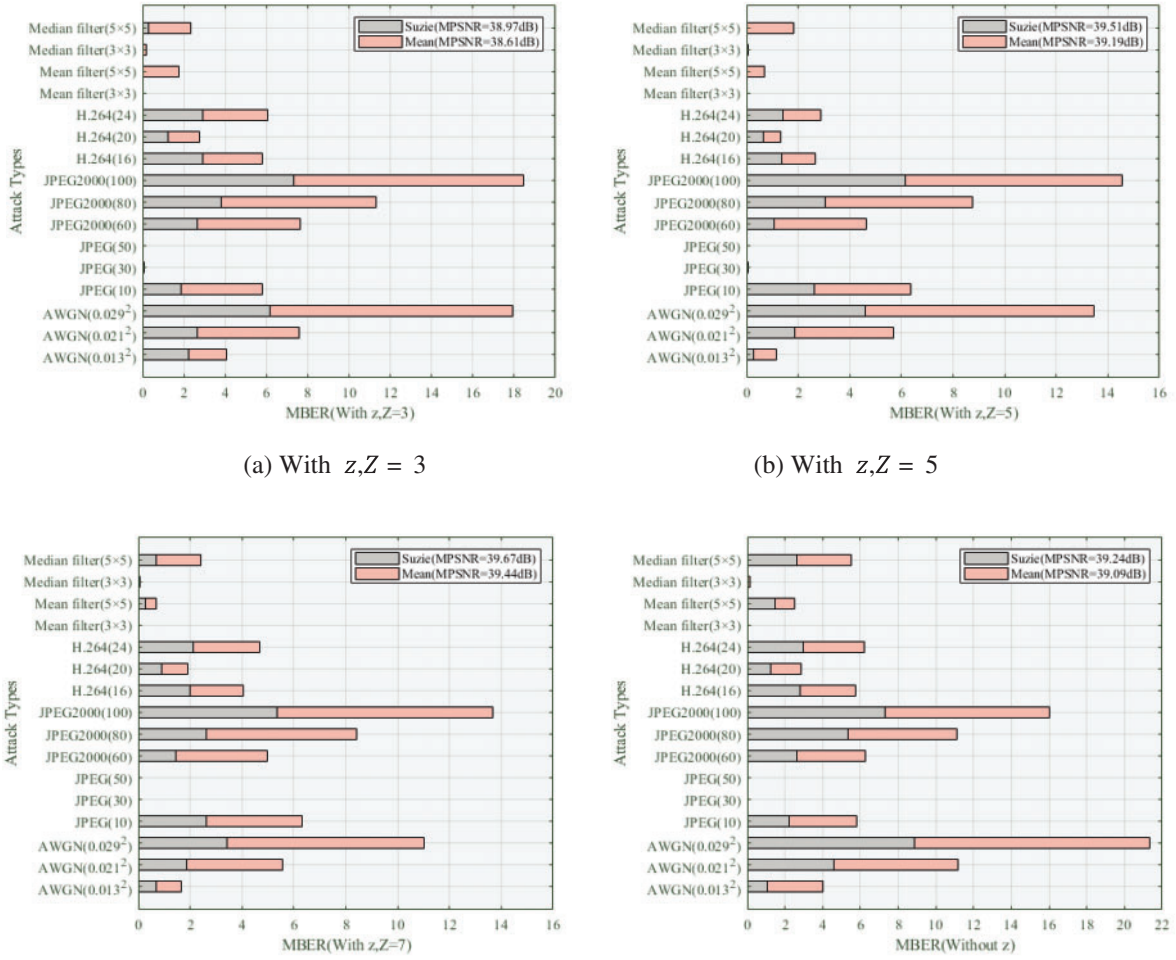
The data presented in Fig. 5 indicates that for $P = 3$, the optimized STDM strikes the optimal balance between invisibility and robustness. It is observed that when the MPSNR values are comparable, the lowest MBER value is achieved, showcasing the effectiveness of the optimized STDM's variable spread spectrum technology. Unlike QIM, the optimized STDM embeds a resilient watermark bit into the three-dimensional Packet Cosine Energy Transformation (PCET) moment, significantly enhancing robustness at a slight cost to invisibility. However, at $P = 5$, robustness diminishes due to an excess of spreading bits leading to inadequate low-order PCET moments. To mitigate this, the remaining spread spectrum watermark bits are embedded into high-order PCET moments, representing complex content susceptible to compression attacks.

The optimized STDM distinguishes itself from traditional STDM through multi-level embedding strength, allowing adaptive robust watermark embedding with varying intensities. Traditional STDM confines watermarks to fixed embedding intensities, whereas the optimized STDM achieves adaptability through multi-level quantization. To evaluate the advantage of optimized STDM over traditional STDM in the proposed scheme, a robustness analysis is conducted under two conditions: with and without an adaptive embedding scheme utilizing variable embedding strength z. By setting the maximum embedding strength Z at three levels (3, 5, and 7), the experiment regulates the upper limit of the embedding strength. The quantization step size Δ is adjusted to ensure similar MPSNR values across all experimental videos in both environments.

The summarized data results are presented in Fig. 6, showcasing the comparative outcomes of the robustness analysis under varying embedding strength levels in the optimized STDM scheme.

The findings in Fig. 6 demonstrate that the adaptive embedding scheme with variable embedding strength exhibits significantly improved robustness and advantages compared to the fixed embedding strength scheme across various attacks. Adaptive embedding allows for the adjustment of each watermark bit's embedding strength to match the robustness of the PCET moment under diverse attacks, resulting in enhanced robustness within the same visual quality range. With an increase in the maximum embedding strength Z, the robustness shows a gradual rise. However, a notable observation

is that as Z transitions from 5 to 7, the decrease in MBER diminishes notably, even showing a rebound. This highlights the need to consider the impact of the enhanced robustness in subsequent analyses.



(a) With z,$Z = 3$ (b) With z,$Z = 5$

(c) With z,$Z = 7$ (d) Without z

Figure 6: Robustness performance of the proposed DW-PCET method with different embedding strength strategies in terms of MBER against attacks

In the subsequent robustness comparison analysis, the optimal embedding strength is set at $Z = 5$, establishing it as the optimal configuration for the proposed scheme in comparison to other video robust watermark algorithms. Noteworthy is that the adaptive embedding scheme with variable embedding strength not only enhances the robustness of the watermark but also mitigates quantization errors stemming from watermark embedding. Eq. (17), compared to Eq. (8), only quantizes the integer part of the normalized PHT moment amplitude, preserving the decimal part for embedding. This approach significantly reduces quantization errors, minimizes visual quality impact, and decreases reversible watermark volume, further enhancing visual quality preservation.

Table 3 presents the lengths of quantization errors generated by traditional STDM and optimized STDM, aggregating the average error values across all key frames in the 20 experimental videos.

Table 3: The average number of bits of quantization error of different quantizers

Quantizer	Bit number of Suzie	Mean bit number
Traditional STDM	21,873	22,097.5
Optimized STDM	2619	2608.4

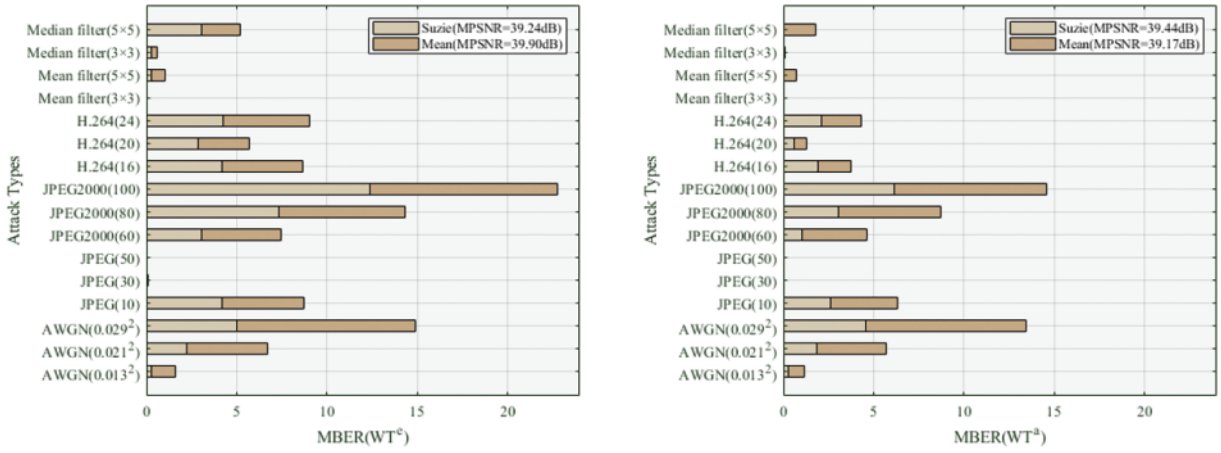
The results in Table 3 illustrate that the optimized STDM effectively reduces the quantization error d_q compared to traditional STDM, showcasing a significant improvement in error reduction. In conclusion, the optimized STDM implemented in the proposed scheme offers notable advantages.

4.3.2 Advantages Analysis of Adaptive Normalization

In a similar manner, we assess the advantages of adaptive normalization for attack simulation within the proposed scheme using two distinct normalization strategies: one employing a constant normalization weight WT^e, and the other utilizing the adaptive normalization weight WT^a. To ensure a fair comparison, we carefully set the appropriate WT^e to align the MPSNR values of the robustly watermarked video key frames against attack simulations under both normalization approaches. The experimental video selection mirrors that of Section 3.3.1, and the attack simulation encompasses the attacks and intensities outlined in Table 2.

During the experiment, we initially derive the adaptive normalization weight WT^a using Eq. (16) on a PCET moment-by-moment basis. Subsequently, WT^e is adjusted according to the amplitude to achieve comparable MPSNR values under both strategies.

The summarized outcomes are detailed in Fig. 7.



(a) With WT^e (b) With WT^a

Figure 7: Robustness performance of the proposed DW-PCET method with different normalization strategies in terms of MBER against attacks

Based on the findings presented in Fig. 7, it is evident that under comparable MPSNR values, the robust watermark embedding utilizing adaptive normalization weight WT^a exhibits a lower MBER value than its counterpart employing the constant normalization weight WT^e. This outcome

conclusively demonstrates the superior advantage of the adaptive normalization mechanism integrated within this scheme.

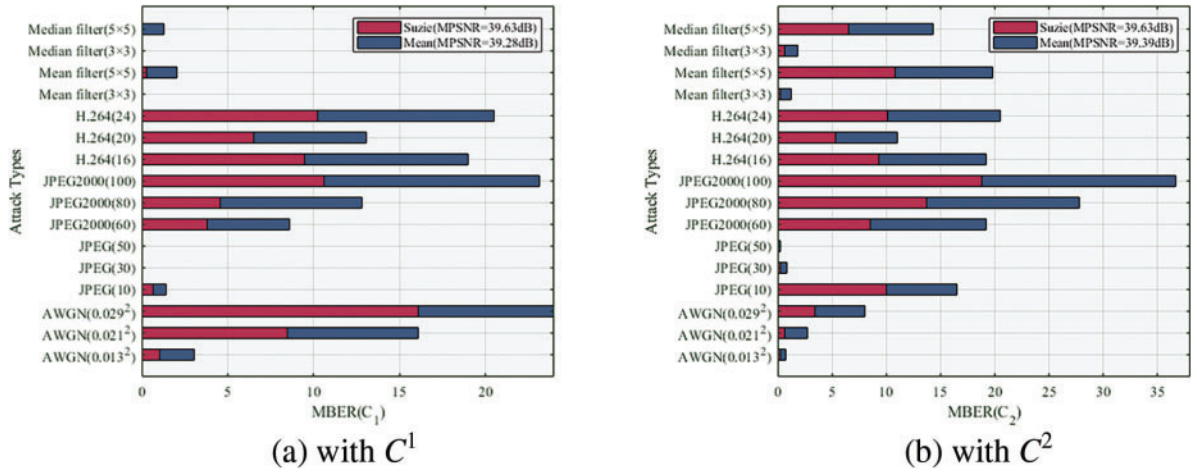
4.3.3 Advantages Analysis of Attack Simulation Scheme

In the proposed scheme, we implement an attack simulation strategy tailored to fit the values of adaptive normalization in PHT moments and the adaptive embedding of robust watermarks. To comprehensively analyze and assess the effectiveness of this strategy, we meticulously curate a collection of various attacks listed in Table 2, amalgamating them to form 7 distinct attack combinations denoted as C^1, C^2, C^3, C^4, C^5, C^6, C^7. Subsequently, these 7 attack combinations are applied within the attack simulation strategy, as outlined in Table 4.

Table 4: Attack combinations for attack simulation

Attack combinations	Attack types
C^1	None
C^2	AWGN
C^3	AWGN, JPEG
C^4	AWGN, JPEG, Mean filtering
C^5	AWGN, JPEG, Mean filtering, JPEG2000
C^6	AWGN, JPEG, Mean filtering, JPEG2000, Median filtering
C^7	AWGN, JPEG, Mean filtering, JPEG2000, Median filtering, H.264

To ensure a fair comparison of the fitting effects across different combinations, we adjust the quantization step size Δ under each combination strategy to align the MPSNR values of all key frames embedded with robust watermarks within the 20 experimental videos. Similar to the approach taken in evaluating the advantages of various innovative technologies discussed previously, we employ the same set of experimental videos to assess the combination that yields the highest level of robustness among the 7 attack combinations within the proposed scheme. The resulting experimental data is succinctly presented in Fig. 8.



(a) with C^1 (b) with C^2

Figure 8: (Continued)

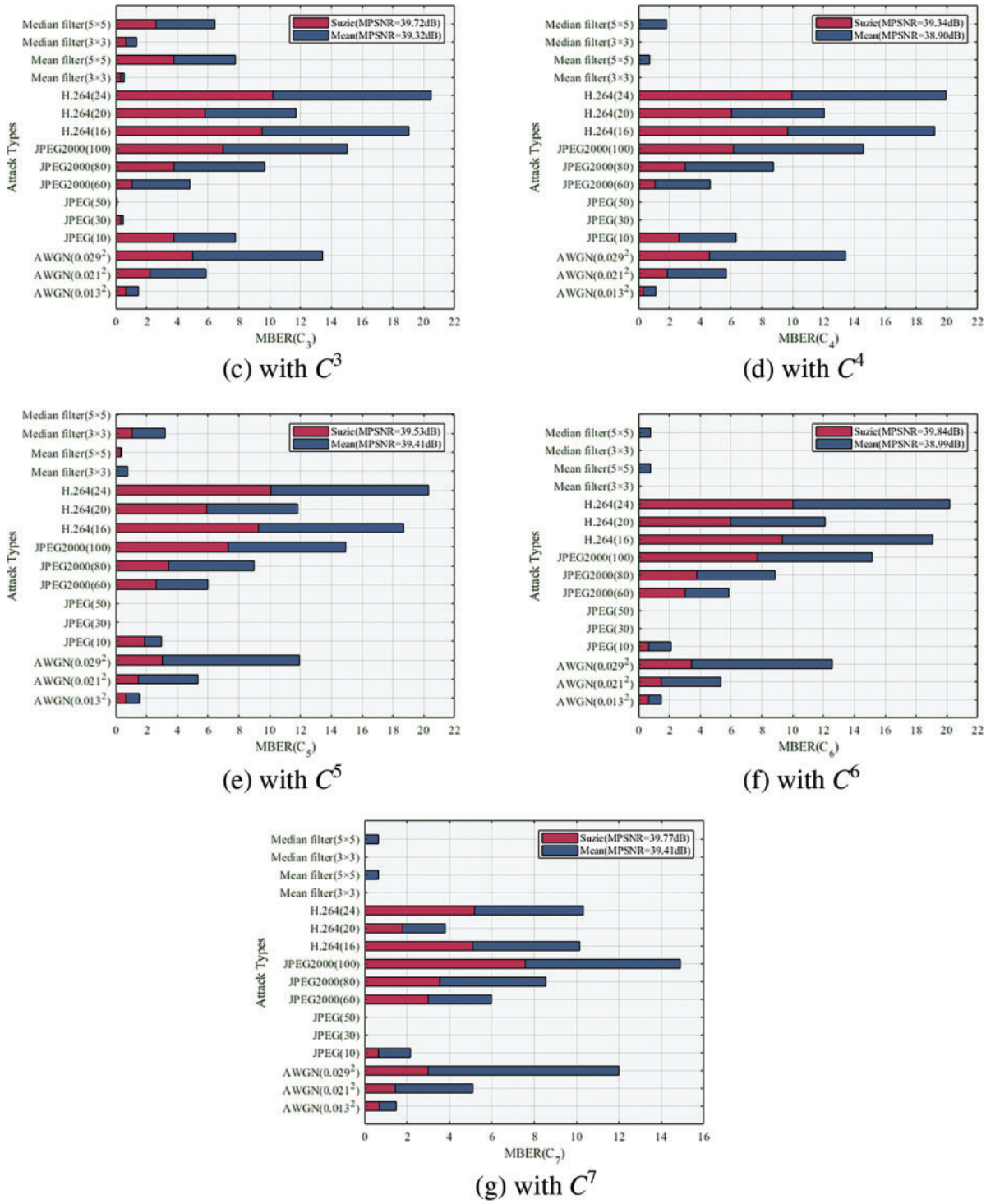


Figure 8: Robustness performance of the proposed DW-PCET method with different attack combinations in terms of MBER against attacks

As illustrated in Fig. 8, attack combination 1, which does not utilize attack simulation, demonstrates the lowest robustness compared to scenarios employing attack simulation strategies. This outcome is attributable to the absence of any attack in combination 1, indicating that neither the adaptive normalization of PHT moments nor the adaptive watermark embedding techniques were applied. The analysis highlights that neglecting these two innovative technologies significantly compromises robustness, as the normalized PHT moment and variable adaptive embedding strength are crucial for enhancing resilience.

Moreover, it is evident that using a simulation strategy that incorporates multiple attack types improves the overall robustness of the proposed scheme against various attack vectors. However, a careful examination of the data reveals two key issues: first, while increasing the number of simulated attack types may enhance general robustness, it can inadvertently weaken resilience against specific attacks. This is because the adaptive selection of embedding strength may not align with the optimal strength for any given attack type, resulting in reduced targeting effectiveness. Second, this approach can lead to increased computational complexity.

To address these challenges, we propose a representative attack compensation mechanism aimed at mitigating the associated costs while maintaining robustness. We compared C^2 and C^3 and found that there is no JPEG2000 attack in C^3, but compared with the scheme using C^2, the scheme using C^3 is also more robust to JPEG2000 attacks. Comparing C^2 and C^4, we found that there is no median filtering attack in C^4, but compared with the scheme using C^2, the scheme using C^4 is also more robust to median filtering attacks. This shows that the JPEG attack used in the attack simulation can represent the JPEG2000 attack to a certain extent, and the mean filtering attack can represent the median filtering attack to a certain extent. Based on this phenomenon, we designed a representative attack compensation mechanism with the same principle: selecting a representative attack of the same type to represent this type of attack. This mechanism can better balance the computational complexity and robustness, and can alleviate the contradiction between overall robustness and specific robustness. Here, we select AWGN attack, JPEG attack, mean filtering attack and H.264 attack as the actual best attack simulation combination, and use the attacks as representatives of the corresponding types of attacks to improve the robustness against different types of attacks. In summary, applying the attack simulation fitting numerical strategy to the proposed scheme can actually enhance the robustness against different types of attacks, proving that the scheme using the attack simulation fitting numerical strategy is more robust than the scheme without the attack simulation strategy.

4.4 Invisibility Analysis of the Proposed Scheme

This section presents an experimental evaluation of the invisibility of the proposed scheme, utilizing two key metrics: visual quality and PSNR values. Visual quality serves as an intuitive indicator, while the PSNR value between the key frame with the dual-domain watermark and the original key frame acts as an objective measure. The experimental parameters for the invisibility assessment are set based on the optimal parameters identified in the preceding sections.

In this experiment, a 1024-bit robust watermark is embedded into 20 test videos, following which the reversible watermark D is computed and embedded into the video. Subsequently, robust watermarked videos and dual-domain watermarked videos are generated, and the key frames containing the watermarks are identified. The PSNR values between these marked key frames and the original key frames are calculated and averaged. As an illustration, the visual quality performance for the video Tennis is depicted in Fig. 9.

(a) (b) (c)

Figure 9: Comparison between the marked keyframe generated by dual-domain watermarking of a keyframe in Tennis using DW-PCET and the corresponding original keyframe. (a) The dual-domain watermarked keyframe ($PSNR = 41.2$); (b) the corresponding original keyframe; (c) the difference between (b) and the keyframe recovered from (a)

In Fig. 9, the visual comparison is presented between a key frame marked with a dual-domain watermark and the corresponding original key frame, along with the disparity between the original key frame and its reconstructed counterpart in the video Tennis. Through visual assessment, it is observed that (a) exhibits high visual quality, making it challenging for the human eye to discern any noticeable differences directly. Additionally, (c) demonstrates that DW-PCET can faithfully restore the original key frame without succumbing to the influence of malicious attacks. Similar effects of restoration are achieved by DW-PCT and DW-PST, not only on the remaining key frames within the video but also across key frames in the other videos.

To enhance the evaluation of the proposed scheme's invisibility with precision, we computed the MPSNR values for all key frames bearing robust watermarks and dual-domain watermarks in the video Tennis. This objective analysis facilitates a comprehensive comparison of the average MPSNR values for these key frames across the 20 experimental videos. The consolidated data is tabulated in Table 5 for reference and detailed examination.

Table 5: MPSNRs in terms of dB of the robustly watermarked keyframe and the dual-domainly watermarked keyframe calculated by the RRW-PCET method

Watermarked keyframe	X_c'	X_c''
MPSNR for Tennis	41.03	39.25
Mean MPSNR	40.47	38.87

The MPSNR values for key frames bearing robust watermarks and dual-domain watermarks in the video 'Tennis' are recorded as 41.03 and 39.25 dB, respectively. Notably, the MPSNR value for key frames with dual-domain watermarks is lower compared to those with robust watermarks, attributed to the additional distortion introduced by embedding the reversible watermark D. Across the 20 experimental videos, the average MPSNR discrepancy between these key frames does not exceed 1.6 dB, indicating an acceptable level for practical implementation. This observation is consistent in the outcomes of DW-PCT and DW-PST, highlighting the reliability of the scheme across various scenarios.

4.5 Robustness Analysis of the Proposed Scheme

The proposed scheme includes DW-PCET, DW-PCT, and DW-PST. In order to effectively evaluate the robustness of these three methods, we compare these three methods with three state-of-the-art video robust watermarking algorithms (Fan et al. [38], Singh et al. [39], and Chen et al. [27]) in this section. As described in Section 3.1, we selected 20 experimental videos (10 QCIF videos and CIF videos with the same video content) from Derf's collection and used the three methods in the proposed scheme to embed two robust watermark bit strings of different lengths (1024 bits and 4096 bits) and combined them with the best parameter settings obtained in the above experiments for experiments. It should be noted that the three comparison algorithms are all video watermark algorithms that use digital images or feature matrices of digital images as robust watermarks. To solve the problem of inconsistent watermark types, we use the two-bit strings of 1024 bits and 4096 bits to form 32×32 and 64×64 robust watermark images respectively through zig-zag scanning in order to ensure a consistent environment of the input robust watermark bits. Based on the above settings, Tables 6 and 7 respectively show the MPSNR values of embedding two different lengths of robust watermarks, where the MPSNR value of each algorithm is calculated by calculating the average PSNR value of all video frames used for embedding in each algorithm, and finally taking the average of the PNSR average of each video to obtain the MPSNR value of each algorithm.

Table 6: MPSNRs in terms of dB of video key frames after embedding 1024-bit watermark for all methods including the comparison algorithm

Methods	Fan et al. [38]	Singh et al. [39]	Chen et al. [27]
MPSNR for Tennis	39.27	39.24	39.51
Mean MPSNR	39.33	39.18	39.56
Methods	DW-PCET	DW-PCT	DW-PST
MPSNR for Tennis	39.77	39.89	39.90
Mean MPSNR	39.60	40.04	39.59

Table 7: MPSNRs in terms of dB of video key frames after embedding 4096-bit watermark for all methods including the comparison algorithm

Methods	Fan et al. [38]	Singh et al. [39]	–
MPSNR for Tennis	38.76	38.88	–
Mean MPSNR	38.66	38.41	–
Methods	DW-PCET	DW-PCT	DW-PST
MPSNR for Tennis	39.11	39.42	39.08
Mean MPSNR	39.37	39.07	38.97

As shown in Tables 6 and 7, the three video dual-domain watermarking algorithms (DW-PCET, DW-PCT, DW-PST) in the proposed scheme can obtain higher MPSNR values than the three comparison algorithms. If the three algorithms in the proposed scheme can obtain lower MBER values than the comparison algorithms under the same conditions, it means that the proposed scheme has more advantages in robustness against attacks. The MPSNR value of Chen-Zernike is missing in

Table 7, because the numerical stability of high-order Zernike moments is poor and it is impossible to maintain the stability of watermark bits under large bit embedding. Therefore, the 4096-bit robust watermark embedding of Chen-Zernike is not discussed in this section.

4.5.1 Robustness Analysis against CSP Attacks

In order to test the robustness of the three methods in the proposed scheme and other comparison algorithms against CSP attacks, we applied the CSP attacks with the corresponding attack strengths in Table 2 to the 20 experimental videos marked with the above robust watermarks. Fig. 10 shows the MBER values of the three methods in the proposed scheme and the other three comparison algorithms (only two comparison algorithms in (b), excluding Chen-Zernike) against the CSP attacks after embedding 1024-bit robust watermark bits and 4096-bit robust watermark bits. The MBER value of each algorithm is obtained by calculating the average BER value of all video frames used for embedding in each algorithm, and finally taking the average BER value of all experimental videos to obtain the MBER value of each algorithm.

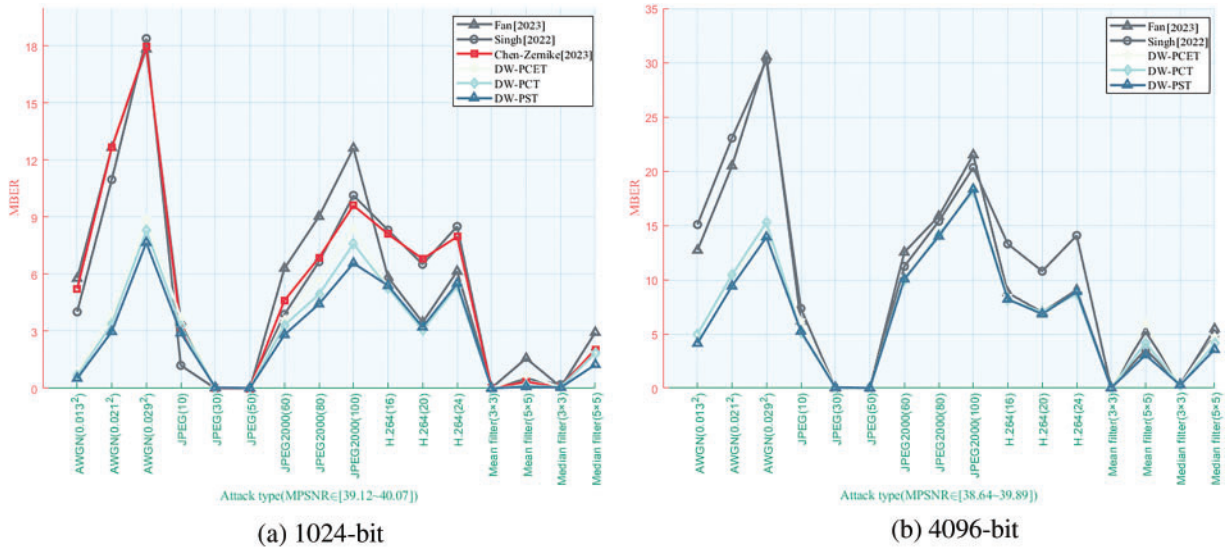


Figure 10: Robustness of all compared methods (Fan et al. [38], Singh et al. [39], and Chen et al. [27]) in terms of MBER against AWGN, JPEG, JPEG2000, H.264, mean filtering, and median filtering

The proposed scheme demonstrates comprehensive robustness against CSP attacks, particularly in the context of AWGN attacks, where the MBER values achieved by its three methods—DW-PCET, DW-PCT, and DW-PST—are significantly lower than those of the comparison algorithms (Fan, Singh, and Chen-Zernike). For instance, at $\sigma^2 = 0.029$, the MBER values are 8.88, 8.23, and 7.62 for the proposed methods, compared to 16.83, 17.38, and 16.95 for the comparatives. This indicates superior robustness in the proposed scheme.

However, in the case of JPEG2000 attacks, the MBER values of the proposed methods are only slightly better than those of the comparison algorithms. This is because the proposed scheme opted to focus on JPEG attacks to balance computational complexity with robustness and thus neglected JPEG2000 in its simulations. As a result, the proposed methods exhibit significantly higher MBER values against JPEG attacks, even under maximum attack intensity ($C_r = 100$), where they remain

below 7.41. These findings underscore the effectiveness of the adaptive normalization and variable embedding strength strategies employed in the proposed scheme.

For mean filtering attacks, all methods, including the comparison algorithms, show low MBER values below 1.58, indicating strong robustness. Notably, among the proposed methods, DW-PCT and DW-PST outperform DW-PCET due to their inherent robustness against filtering attacks.

In median filtering attacks, the proposed methods again yield the lowest MBER values. However, as median filtering was not included in the previous attack simulations to balance complexity and robustness, this reinforces the success of the adaptive strategy of the proposed scheme.

Regarding H.264 attacks, Fan's method exhibits the lowest MBER value outside the proposed schemes. Fan's approach achieves its robustness by selecting the watermark embedding sub-block based on chrominance and adjusting the embedding method according to the video and watermark characteristics, enhancing resistance to H.264 attacks. The distribution of MBER values for each method in (b) is similar to that in (a), with an overall increase observed across all methods. This increase is attributed to the need to appropriately reduce embedding intensity while maintaining a comparable level of MPSNR, leading to higher MBER values. Under AWGN attack, all methods exhibit a significant rise in MBER values; however, the three methods of the proposed scheme consistently demonstrate lower values than the comparison algorithms, highlighting the advantages of adaptive variable embedding strength.

For JPEG2000 and JPEG attacks, the proposed methods show lower MBER values and greater robustness compared to the comparison algorithms. Notably, the MBER value of DW-PCET is higher than that of DW-PCT and DW-PST due to the relative numerical instability of the PCET moments compared to the PCT and PST moments [34].

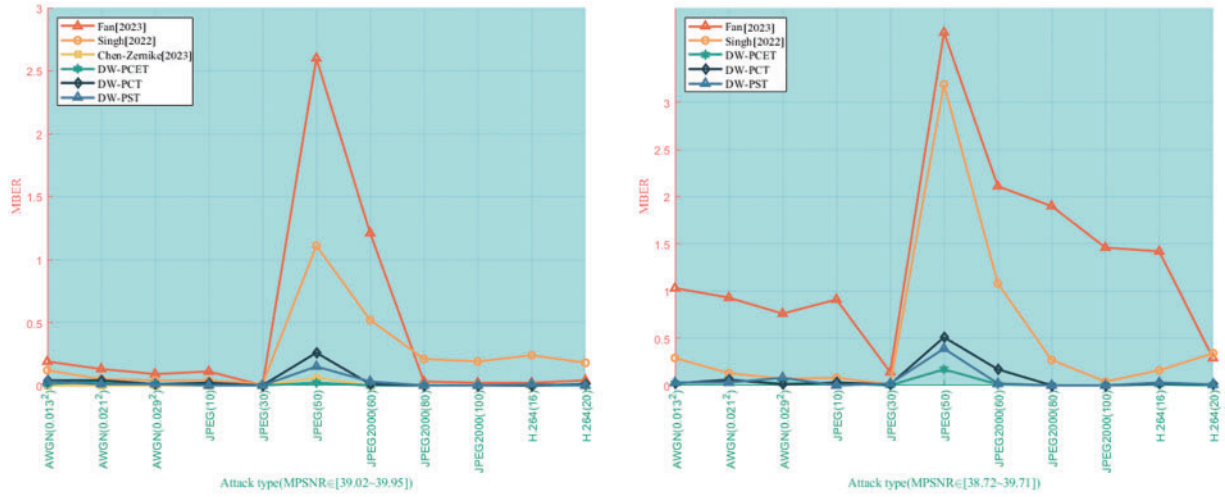
In the case of mean and median filtering attacks, the MBER values of the proposed methods decrease, except for Fan and Singh. This reduction is a result of the robust watermark embedding capability of the proposed methods, which capture geometric moments of global frame features, proving effective against filtering attacks.

Regarding H.264 attacks, Fan's method remains a strong competitor but experiences a larger increase in MBER values compared to the proposed methods. This is due to the fixed embedding strength in Fan's method, which fails to adjust for a higher embedding load effectively, impacting invisibility. To achieve similar MPSNR values, Fan's method compensates by selecting more video frames for longer robust watermarks, thereby increasing data dispersion and affecting robustness. Unlike (a), the comparative analysis of the robustness of the three methods in the proposed scheme and Chen is excluded from (b) due to the instability of Zernike moments with a larger watermark bit count. In summary, the three methods of the proposed scheme have the lowest MBER value overall compared with all compared algorithms, which illustrates the strong robustness advantage of the proposed scheme.

4.5.2 Robustness Analysis against GD Attacks

In order to test the robustness of the three methods and the comparison algorithms in the face of GD attacks, we adopt the same parameter settings as Section 4.5.1, and perform rotation attacks of different angles and scaling attacks of different ratios on each robust watermarked experimental video. Among them, the rotation angle RA range is defined as $[15^\circ, 90^\circ]$, and the step size is defined as 15°; the scaling factor is defined between $[0.5, 2]$, and the step size is defined as 0.25. Fig. 11 summarizes the robustness performance of these methods against rotation and scaling under two robust watermark

bit string lengths. Similarly, there are only two comparison algorithms in (b), and Chen-Zernike is not included. Among them, the MBER value of each algorithm is obtained by calculating the mean of the BER values of all video frames used for embedding in each algorithm, and finally taking the mean of the BER values of all experimental videos to obtain the MBER value of each algorithm.



(a) 1024-bit (b) 4096-bit

Figure 11: Robustness of all compared methods (Fan et al. [38], Singh et al. [39], and Chen et al. [27]) in terms of MBER against rotation and scaling

From the findings presented in Fig. 11, it is evident that all methods exhibit notable robustness. (a) showcases the MBER values of each method at varying rotation angles and scaling ratios post-embedding a 1024-bit robust watermark. Notably, Chen's method displays slightly inferior anti-rotation robustness compared to the three methods within the proposed scheme, attributed to the weaker numerical stability of integer-order Zernike moments in contrast to PHT moments. In terms of scaling attacks, the proposed methods showcase the highest robustness due to the scaling invariance of PHT moments [34], resulting in MBER values consistently below the threshold of 0.26.

Similarly, (b) outlines the MBER values of all methods following the embedding of a 4096-bit robust watermark at different rotation angles and scaling ratios. The distribution of MBER values across algorithms in (b) closely mirrors that of (a), with an overall increase in MBER values. The numerical instability associated with high-order Zernike moments with a 4096-bit robust watermark leads to the exclusion of Chen-Zernike from evaluation, akin to previous robustness comparison analyses.

In summary, the results highlight the strong robustness of the proposed methods across rotation angles and scaling ratios, underscoring their effectiveness in video watermarking applications.

5 Limitations of the Proposed Scheme

While our proposed dual-domain watermarking scheme for digital video offers promising advancements in copyright protection and watermark robustness, it is essential to acknowledge certain limitations and situations where the method may be less effective. The following sections outline the key shortcomings and potential risks associated with our approach:

1. Dependency on Parameter Optimization. The effectiveness of our algorithm is closely tied to the parameter optimization strategy based on attack simulation fitting. Although this method enables adaptive parameter adjustments, it may yield suboptimal performance across diverse video types or conditions, potentially introducing significant computational overhead and limiting its applicability in real-time scenarios.

2. Sensitivity to Video Content. The algorithm's efficacy varies with different video content. Highly dynamic scenes or complex backgrounds can challenge the watermark's invisibility and robustness, leading to perceptible artifacts or diminished integrity. Such variability necessitates careful consideration in the algorithm's deployment.

3. Risks in Reversibility. The use of recursive coding for ensuring reversibility introduces risks related to error propagation. Difficulties in the watermark recovery process or subsequent video manipulations may result in the loss of original information or unintended artifacts. Safeguarding the integrity of reversible watermarks under varied conditions remains a critical challenge.

6 Conclusion

The paper delves into video watermarking and introduces a novel video dual-domain watermarking scheme that leverages adaptive normalized PHT moment and optimized dithering spread transform dither modulation. The scheme comprises four primary components:

1. Key Frame Selection: Identifying smooth video frames from the footage as key frames, serving as the embedding domain for the dual-domain watermark.

2. Dual-Domain Watermark Embedding: Embedding a robust watermark for copyright protection and generating a reversible watermark for attack detection and authentication.

3. Attack Detection and Authentication: Employed to ascertain if the received video has undergone tampering.

4. Robust Watermark Extraction and Video Restoration: In case of an attack on the watermarked video, extracting the reversible watermark and robust watermark in sequence through the inverse embedding process to restore the original video. If an attack is detected, extraction focuses solely on the robust watermark.

Through extensive experiments, it is validated that embedding 1024-bit and 4096-bit robust watermarks renders the proposed scheme highly resilient against various attacks such as AWGN, JPEG, JPEG2000, H.264, CSP attacks like mean filtering and median filter, as well as geometric attacks encompassing rotation and scaling. The three geometric moment methods within the scheme exhibit superior robustness compared to cutting-edge video watermarking algorithms under similar conditions of invisibility and embedding capacity.

However, the analysis in the paper reveals certain shortcomings in the proposed scheme. The proposed dual-domain watermarking scheme effectively embeds robust watermarks and demonstrates resilience against various attacks, outperforming state-of-the-art algorithms in invisibility and capacity. Increased computational complexity due to moment-by-moment normalization and bit-by-bit embedding selection may hinder real-time applications and necessitates optimization for efficiency. Implementing this method in commercial settings may incur higher operational costs, highlighting the need to assess its cost-effectiveness compared to existing methods. These areas warrant focused attention in future research endeavors.

Acknowledgement: We thank all the members who have contributed to this work with us.

Funding Statement: This work was supported in part by the National Natural Science Foundation of China under Grant 62202496, 62272478 and the Basic Frontier Innovation Project of Engineering University of People Armed Police under Grant WJY202314, WJY202221.

Author Contributions: Conceptualization: Yucheng Liang, Ke Niu; Experimental operation and data proofreading: Yucheng Liang, Yingnan Zhang, Fangmeng Hu; Analysis and interpretation of results: Yucheng Liang, Ke Niu, Yingnan Zhang; Draft manuscript preparation: Yucheng Liang, Yifei Meng, Fangmeng Hu; Figure design and drawing: Yucheng Liang, Ke Niu, Yifei Meng. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Given that the source code and data contain research findings that have not yet been publicly disseminated by our experimental team, we currently face challenges in making them openly available. Moreover, our academic institution is governed by confidentiality protocols that necessitate the disclosure of the source code and data only after the stipulated decryption period has been met.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] R. Ogla, E. S. Mahmood, R. I. Ahmed, and A. M. S. Rahma, "New fragile watermarking technique to identify inserted video objects using H.264 and color features," *Comput. Mater. Contin.*, vol. 76, no. 3, pp. 1–22, 2023. doi: 10.32604/cmc.2023.039818.

[2] Y. Zhang, J. Ni, W. Su, and X. Liao, "A novel deep video watermarking framework with enhanced robustness to H.264/AVC compression," presented at the Proc. 31st ACM Int. Conf. Multimedia, Ottawa, ON, Canada, Oct. 29–Nov. 3, 2023. doi: 10.1145/3581783.3612270.

[3] M. Asikuzzaman, H. Mareen, N. Moustafa, K. -K. R. Choo, and M. R. Pickering, "Blind Camcording-resistant video watermarking in the DT CWT and SVD domain," *IEEE Access*, vol. 10, pp. 15681–15698, 2022. doi: 10.1109/ACCESS.2022.3146723.

[4] Q. Chang, L. Huang, S. Liu, H. Liu, T. Yang and Y. Wang, "Blind robust video watermarking based on adaptive region selection and channel reference," in *Proc. 30th ACM Int. Conf. Multimedia*, Lisboa, Portugal, Oct. 10–14, 2022. doi: 10.1145/3503161.3548168.

[5] M. He, H. Wang, F. Zhang, S. M. Abdullahi, and L. Yang, "Robust blind video watermarking against geometric deformations and online video sharing platform processing," *IEEE Trans. Dependable Secure Comput.*, vol. 20, no. 6, pp. 4702–4718, 2022. doi: 10.1109/TDSC.2022.3232484.

[6] S. Takale and A. Mulani, "DWT-PCA based video watermarking," *J. Electr., Comput. Network. Appl. Mathemat.*, pp. 2799–1156, 2022. doi: 10.55529/jecnam.26.1.7.

[7] K. Guo, Z. Xu, S. Luo, F. Wei, Y. Wang and Y. Zhang, "Invisible video watermark method based on maximum voting and probabilistic superposition," in *Proc. 31st ACM Int. Conf. Multimedia, MM*, New York, NY, USA, Oct. 29–Nov. 3, 2023. doi: 10.1145/3581783.3612842.

[8] K. Sahu, "A logistic map based blind and fragile watermarking for tamper detection and localization in images," *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 8, pp. 3869–3881, 2022. doi: 10.1007/s12652-021-03365-9.

[9] D. Coltuc and J. M. Chassery, "Distortion-free robust watermarking: A case study," presented at the Secur. Steganograp. Watermarking Multimedia Contents IX, San Jose, CA, USA, SPIE, Feb. 27, 2007.

[10] R. Kumar and K. H. Jung, "Robust reversible data hiding scheme based on two-layer embedding strategy," *Inf. Sci.*, vol. 512, no. 5, pp. 96–107, 2020. doi: 10.1016/j.ins.2019.09.062.

[11] X. Liang, S. Xiang, L. Yang, and J. Li, "Robust and reversible image watermarking in homomorphic encrypted domain," *Signal Process.: Image Commun.*, vol. 99, 2021, Art. no. 116462. doi: 10.1016/j.image.2021.116462.
[12] E. Chrysochos, V. Fotopoulos, A. N. Skodras, and M. Xenos, "Reversible image watermarking based on histogram modification," in *Proc. 11th Panhellenic Conf. Inform. (PCI 2007)*, Patras, Greece, May 18–20, 2007. doi: 10.1109/TSSA.2014.7065948.
[13] C. C. Chang, P. Y. Lin, and J. S. Yeh, "Preserving robustness and removability for digital watermarks using subsampling and difference correlation," *Inf. Sci.*, vol. 179, no. 13, pp. 2283–2293, 2009. doi: 10.1016/j.ins.2009.03.003.
[14] R. Hu and S. Xiang, "Cover-lossless robust image watermarking against geometric deformations," *IEEE Trans. Image Process.*, vol. 30, pp. 318–331, 2021. doi: 10.1109/TIP.2020.3036727.
[15] R. Hu and S. Xiang, "Lossless robust image watermarking by using polar harmonic transform," *Signal Process.*, vol. 179, 2021, Art. no. 107833. doi: 10.1016/j.sigpro.2020.107833.
[16] Y. Tang, S. Wang, C. Wang, S. Xiang, and Y. -M. Cheung, "A highly robust reversible watermarking scheme using embedding optimization and rounded error compensation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 4, pp. 1593–1609, 2023. doi: 10.1109/TCSVT.2022.3216849.
[17] D. Fu, X. Zhou, L. Xu, K. Hou, and X. Chen, "Robust reversible watermarking by fractional order zernike moments and pseudo-zernike moments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 12, pp. 7310–7326, 2023. doi: 10.1109/TCSVT.2023.3279116.
[18] Y. Tang, K. Li, C. Wang, S. Bian, and Q. Huang, "A two-stage robust reversible watermarking using polar harmonic transform for high robustness and capacity," *Inf. Sci.*, vol. 654, 2024, Art. no. 119786. doi: 10.1016/j.ins.2023.119786.
[19] C. De Vleeschouwer, J. -F. Delaigle, and B. Macq, "Circular interpretation of bijective transformations in lossless watermarking for media asset management," *IEEE Trans. Multimedia*, vol. 5, no. 1, pp. 97–105, 2003. doi: 10.1109/TMM.2003.809729.
[20] Z. Ni, Y. Q. Shi, N. Ansari, W. Su, Q. Sun and X. Lin, "Robust lossless image data hiding designed for semi-fragile image authentication," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 4, pp. 497–509, 2008. doi: 10.1109/TCSVT.2008.918761.
[21] W. Wang, J. Ye, T. Wang, and W. Wang, "Reversible data hiding scheme based on significant-bit-difference expansion," *IET Image Process*, vol. 11, no. 11, pp. 1002–1014, 2017. doi: 10.1049/iet-ipr.2017.0151.
[22] S. Xiang and Y. Wang, "Distortion-free robust reversible watermarking by modifying and recording IWT means of image blocks," presented at the Dig.-Forens. Watermarking: 14th Int. Workshop, IWDW 2015, Tokyo, Japan, Oct. 7–10, 2016. doi: 10.1007/978-3-319-31960-5_28.
[23] F. Peng, W. -Y. Jiang, Y. Qi, Z. -X. Lin, and M. Long, "Separable robust reversible watermarking in encrypted 2D vector graphics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2391–2405, 2020. doi: 10.1109/TCSVT.2020.2986782.
[24] L. Xiong, X. Han, C. -N. Yang, and Y. -Q. Shi, "Robust reversible watermarking in encrypted image with secure multi-party based on lightweight cryptography," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 1, pp. 75–91, 2022. doi: 10.1109/TCSVT.2021.3055072.
[25] Y. Tang, C. Wang, S. Xiang, and Y. -M. Cheung, "A robust reversible watermarking scheme using attack-simulation-based adaptive normalization and embedding," *IEEE Trans. Inf. Forensics Secur.*, vol. 19, no. 1, pp. 4114–4129, 2024. doi: 10.1109/TIFS.2024.3372811.
[26] G. Gao, M. Wang, and B. Wu, "Efficient robust reversible watermarking based on ZMs and integer wavelet transform," *IEEE Trans. Ind. Inform.*, vol. 20, no. 3, pp. 4115–4123, 2024. doi: 10.1109/TII.2023.3321101.
[27] S. Chen, A. Malik, X. Zhang, G. Feng, and H. Wu, "A fast method for robust video watermarking based on zernike moments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 12, pp. 7342–7353, 2023. doi: 10.1109/TCSVT.2023.3281618.
[28] D. Shapiro, V. Sergeyev, and V. Fedoseev, "Improved ECC-based phase watermarking method for video copyright protection," in *2023 11th Int. Symp. Dig. Forens. Secur. (ISDFS)*, Chattanooga, TN, USA, 2023. doi: 10.1109/ISDFS58141.2023.10131785.

[29] C. Sharma, B. Amandeep, R. Sobti, T. K. Lohani, and M. Shabaz, "A secured frame selection based video watermarking technique to address quality loss of data: Combining graph based transform, singular valued decomposition, and hyperchaotic encryption," *Secur. Commun. Netw.*, vol. 2021, no. 4, pp. 5536170:1–5536170:19, 2021. doi: 10.1155/2021/5536170.

[30] D. Hou, W. Zhang, Y. Yang, and N. Yu, "Reversible data hiding under inconsistent distortion metrics," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5087–5099, 2018. doi: 10.1109/TIP.2018.2851074.

[31] D. Coltuc, "Towards distortion-free robust image authentication," *J. Phys.*, vol. 77, no. 1, 2007, Art. no. 012005. doi: 10.1088/1742-6596/77/1/012005.

[32] X. Wang, X. Li, and Q. Pei, "Independent embedding domain based two-stage robust reversible watermarking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2406–2417, 2019. doi: 10.1109/TCSVT.2019.2915116.

[33] Y. Li *et al.*, "A contrario detection of H.264 video double compression," in *2023 IEEE Int. Conf. Image Process. (ICIP)*, Kuala Lumpur, Malaysia, 2023. doi: 10.1109/ICIP49359.2023.10222775.

[34] P. -T. Yap, X. Jiang, and A. Chichung Kot, "Two-dimensional polar harmonic transforms for invariant image representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1259–1270, 2010. doi: 10.1109/TPAMI.2009.119.

[35] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theor.*, vol. 47, no. 4, pp. 1423–1443, 2001. doi: 10.1109/18.923725.

[36] B. Van Giffen, D. Herhausen, and T. Fahse, "Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods," *J. Bus. Res.*, vol. 144, no. 6, pp. 93–106, 2022. doi: 10.1016/j.jbusres.2022.01.076.

[37] B. A. Wichmann and I. D. Hill, "Generating good pseudo-random numbers," *Comput. Stat. Data Anal.*, vol. 51, no. 3, pp. 1614–1622, 2006. doi: 10.1016/j.csda.2006.05.019.

[38] D. Fan, H. Zhao, C. Zhang, H. Liu, and X. Wang, "Anti-recompression video watermarking algorithm based on H.264/AVC," *Mathematics*, vol. 11, no. 13, 2023, Art. no. 2913. doi: 10.3390/math11132913.

[39] R. Singh, H. Mittal, and R. Pal, "Optimal keyframe selection-based lossless video-watermarking technique using IGSA in LWT domain for copyright protection," *Complex Intell. Syst.*, vol. 8, no. 2, pp. 1047–1070, 2022. doi: 10.1007/s40747-021-00569-6.

[40] P. W. Holland and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," *Commun. Stat.-Theor. Methods*, vol. 6, no. 9, pp. 813–827, 1977. doi: 10.1080/03610927708827533.