

# EXPLORING CONSISTENT SPATIO-TEMPORAL DISTORTION AND STABLE 3-D DCT COEFFICIENTS FOR ROBUST BLIND VIDEO WATERMARKING

Fei Zhang, Hongxia Wang<sup>†</sup>, Mingze He, and Ling Yang

School of Cyber Science and Engineering, Sichuan University, Chengdu, China;  
Key Laboratory of Data Protection and Intelligent Management (Sichuan University),  
Ministry of Education, China.

## ABSTRACT

With the rapid development of mobile Internet and video applications, robust video watermarking technology has become a focal area of research for protecting and tracking intellectual property rights in digital media. An important characteristic of video is that it has both spatial and temporal properties. Previous studies in video watermarking have primarily focused on either spatial or temporal distortions and did not uniformly consider all types of video distortion, which restricts the robustness of video watermarking. In this paper, a novel robust blind video watermarking is proposed by exploring consistent spatio-temporal distortion and stable 3-D DCT coefficients. Our method achieves stronger robustness by uniformly treating the spatial and temporal distortions of the video. The properties of the stable 3-D DCT coefficients are mathematically proved, which makes the scheme generalizable. Extensive experiments have demonstrated that our method is resistant to video compression (H.264/AVC, H.265/HEVC) attacks, geometric attacks, and temporal domain attacks, and outperforms the current state-of-the-art video watermarking schemes.

**Index Terms**— Video watermarking, spatio-temporal distortion, 3-D DCT, stable coefficient, robust.

## 1. INTRODUCTION

The rapid development of 5G wireless communication technology and popular video sharing platforms such as *Tiktok* and *Youtube* make it convenient for users to download and share video resources [1]. However, it also poses a great challenge to the copyright protection of video resources [2,3]. The copyright owner or distribution organization uses robust video watermarking technology to embed proof of ownership in an

imperceptible manner, meeting multimedia information security needs such as copyright protection and tracking. [4–6].

During the process of sharing in network distribution, videos not only suffer from spatial distortion such as compression, scaling, and cropping, but also suffer from temporal distortion such as frame rate conversion, video trimming, and video reversing [7]. However, current robust video watermarking methods generally focus on either spatial or temporal distortions. Huang *et al.* [8] embedded a watermark in the pseudo-3-D DCT domain of the video, which only considers compression distortion and lacks robustness to geometric attacks and temporal processing. Sahu *et al.* [9] used scale invariant feature transform (SIFT) features of the video side view as a watermark against temporal domain attacks, but the scheme is not resistant to geometric attacks. Madine *et al.* [10] used a multiplicative watermarking in 2-D discrete wavelet transform (DWT) domain to resist temporal and compression attacks, but the scheme was not robust to geometric distortions. In recent years, Yang *et al.* [11] embedded the complete watermark in the 2-D DCT high-frequency component of each frame to resist the temporal distortion, but this scheme was also not resistant to geometry attack. Huan *et al.* [12] incorporated the hybrid domain of dual-tree complex wavelet transform (DT CWT) and singular value decomposition (SVD) to embed a watermark, and achieved a combined robustness against compression attacks and geometric attacks. However, this scheme didn't address the issue of temporal distortion.

In summary, the existing robust video watermarking lacks a uniform strategy for dealing with spatio-temporal distortion, which limits the application of video watermarking. To this end, a novel robust blind video watermarking is proposed in this paper. Compared to previous similar 3-D video watermarking [8, 13–16], the main contributions are as below.

- We find the consistency law of spatio-temporal distortion to improve the robustness against both spatial and temporal attacks.
- We explore stable 3-D DCT coefficients to deal with spatial and temporal domain distortions, and mathematically prove them.

---

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 62272331 and 61972269 and Sichuan Science and Technology Program under Grant 2022YFG0320, the Key Laboratory of Data Protection and Intelligent Management, Ministry of Education, Sichuan University, and the Fundamental Research Funds for the Central Universities under Grant SCU2023D008. <sup>†</sup>Corresponding author: hxwang@scu.edu.cn. (Hongxia Wang)

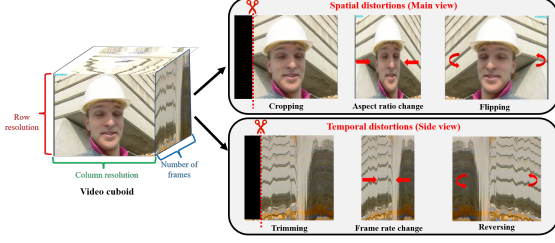


Fig. 1. Consistency of spatio-temporal distortion.

- We propose a novel robust blind watermarking framework to make this technique more practical. Compared with previous works, the proposed method has stronger robustness and better imperceptibility.

## 2. MOTIVATION

### 2.1. Consistency of Spatio-temporal Distortion

The video media is composed of a sequence of images, so it can be abstracted as a cuboid [17]. The height, width, and depth of the cuboid correspond to the number of rows, columns, and frames of the video, respectively. In order to be able to use a uniform strategy to deal with both spatial and temporal domain distortions, we investigate the consistency law between spatial and temporal distortions. Fig. 1 illustrates the consistency between the three spatial domain distortions and the three temporal domain distortions. It is evident that the distortions caused by them are alike, the only difference being the area in which the distortion occurs. In general, spatial domain distortions such as cropping, aspect ratio changes, and flipping affect the main view of the video cuboid, whereas temporal domain distortions like video trimming, frame rate conversion, and video reversing affect the side view. Hence, we put forward the consistency law that there is a consistency between spatial and temporal distortions. In designing the watermarking scheme, we should adopt a uniform strategy to deal with the possible spatial and temporal distortions of the video, rather than considering them separately.

### 2.2. Stable 3-D DCT Coefficients

In this subsection, we explore the stable 3-D DCT coefficients against spatio-temporal distortion and mathematically prove them. A  $N \times N \times N$  3-D DCT transform process is as follows:

$$F(u, v, w) = C(u)C(v)C(w) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \sum_{z=0}^{N-1} S(x, y, z) \times \cos\left(\frac{(2x+1)u\pi}{2N}\right) \cos\left(\frac{(2y+1)v\pi}{2N}\right) \cos\left(\frac{(2z+1)w\pi}{2N}\right), \quad (1)$$

where  $S(x, y, z)$  is the pixel and  $F(u, v, w)$  is the frequency coefficient, the  $C(\cdot)$  is defined as follows:

$$C(i) = \begin{cases} \sqrt{\frac{1}{N}}, & i = 0, \\ \sqrt{\frac{2}{N}}, & i = 1, 2, \dots, N-1. \end{cases} \quad (2)$$

Similar to the concept of basis images in 2-D transforms, we can also write the 3-D basis functions corresponding to the frequency domain coefficients of a 3-D DCT:

$$\phi_{u,v,w}(i, j, k) = C(u)C(v)C(w) \times \cos\left(\frac{(2i+1)u\pi}{2N}\right) \cos\left(\frac{(2j+1)v\pi}{2N}\right) \cos\left(\frac{(2k+1)w\pi}{2N}\right) \quad (3)$$

where the function shape of  $\phi_{u,v,w}$  determines the frequency characteristics of the frequency domain coefficients  $F(u, v, w)$ . We hope that the shape of  $\phi_{u,v,w}$  can be held constant under 3-D flipping, rotation, and transposition attacks, in which case we can obtain 3-D DCT coefficients that are stable against spatio-temporal distortions. Therefore,  $\phi_{u,v,w}$  needs to be satisfied:

$$\begin{cases} \phi_{u,v,w}(i, j, k) = \phi_{u,v,w}(N-1-i, j, k) = \phi_{u,v,w}(i, N-1-j, k) = \phi_{u,v,w}(i, j, N-1-k), & \text{Flip} \\ \phi_{u,v,w}(i, j, k) = \phi_{u,v,w}(j, i, k) = \phi_{u,v,w}(k, j, i) = \phi_{u,v,w}(i, k, j), & \text{Transpose} \\ \phi_{u,v,w}(i, j, k) = \phi_{u,v,w}(N-1-j, i, k) = \phi_{u,v,w}(N-1-k, j, i) = \phi_{u,v,w}(N-1-i, k, j) = \\ \phi_{u,v,w}(j, N-1-i, k) = \phi_{u,v,w}(k, N-1-j, i) = \phi_{u,v,w}(i, N-1-k, j) = \\ \phi_{u,v,w}(j, i, N-1-k) = \phi_{u,v,w}(k, j, N-1-i) = \phi_{u,v,w}(i, k, N-1-j). \end{cases} \quad (4)$$

After exploring, we find that the above equation is satisfied when  $u, v, w$  are equal and they are all even. The specific proof process is as follows.

It is satisfied when  $u, v, w$  is even:

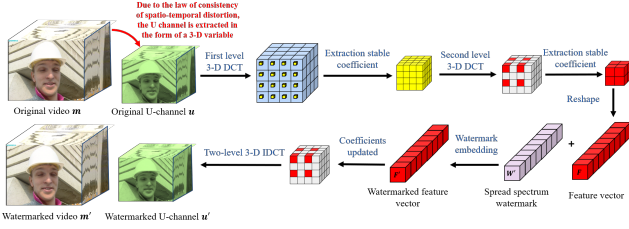
$$\begin{cases} \cos\left(\frac{(2(N-1-i)+1)u\pi}{2N}\right) = \cos\left(u\pi - \frac{(2i+1)u\pi}{2N}\right) = \cos\left(-\frac{(2i+1)u\pi}{2N}\right) = \cos\left(\frac{(2i+1)u\pi}{2N}\right) \\ \cos\left(\frac{(2(N-1-j)+1)v\pi}{2N}\right) = \cos\left(v\pi - \frac{(2j+1)v\pi}{2N}\right) = \cos\left(-\frac{(2j+1)v\pi}{2N}\right) = \cos\left(\frac{(2j+1)v\pi}{2N}\right) \\ \cos\left(\frac{(2(N-1-k)+1)w\pi}{2N}\right) = \cos\left(w\pi - \frac{(2k+1)w\pi}{2N}\right) = \cos\left(-\frac{(2k+1)w\pi}{2N}\right) = \cos\left(\frac{(2k+1)w\pi}{2N}\right). \end{cases} \quad (5)$$

Obviously, combining with the definition of the basis functions in Eq. (3), the ‘Flip’ term in Eq. (4) is satisfied. Further, it is satisfied when  $u = v = w$ :

$$\begin{cases} \cos\left(\frac{(2i+1)u\pi}{2N}\right) \cos\left(\frac{(2j+1)v\pi}{2N}\right) \cos\left(\frac{(2k+1)w\pi}{2N}\right) = \cos\left(\frac{(2i+1)v\pi}{2N}\right) \cos\left(\frac{(2j+1)u\pi}{2N}\right) \cos\left(\frac{(2k+1)u\pi}{2N}\right) \\ \cos\left(\frac{(2i+1)u\pi}{2N}\right) \cos\left(\frac{(2j+1)v\pi}{2N}\right) \cos\left(\frac{(2k+1)u\pi}{2N}\right) = \cos\left(\frac{(2i+1)u\pi}{2N}\right) \cos\left(\frac{(2j+1)v\pi}{2N}\right) \cos\left(\frac{(2k+1)u\pi}{2N}\right) \\ \cos\left(\frac{(2i+1)u\pi}{2N}\right) \cos\left(\frac{(2j+1)v\pi}{2N}\right) \cos\left(\frac{(2k+1)u\pi}{2N}\right) = \cos\left(\frac{(2i+1)u\pi}{2N}\right) \cos\left(\frac{(2j+1)v\pi}{2N}\right) \cos\left(\frac{(2k+1)u\pi}{2N}\right). \end{cases} \quad (6)$$

Combining with the definition of the basis functions in Eq. (3), the ‘Transpose’ term in Eq. (4) is satisfied.

When Eq. (5) and Eq. (6) are satisfied simultaneously, i.e.,  $u = v = w$  and they are both even. Obviously combined with the definition in Eq. (3), the ‘Rotation’ term in Eq. (4) is satisfied. We refer to these 3-D DCT coefficients that satisfy Eq. (4) as stable coefficients. In our scheme, the watermarking message is embedded into these stable coefficients in the form of spread spectrum, which greatly improves the robustness of the watermarking scheme.



**Fig. 2.** The embedding process of proposed video watermarking framework.

### 3. PROPOSED VIDEO WATERMARKING FRAMEWORK

Through exploring consistent spatio-temporal distortion and stable 3-D DCT coefficients, we develop a video watermarking framework. In this section, we will describe the details of the our scheme in terms of the embedding process and the extraction process.

#### 3.1. Watermark Embedding

Fig. 2 gives the overall embedding process diagram. Since the human eye is the least sensitive to the blue component, we choose the U-channel to embed the watermark, as in the mainstream schemes [3, 11, 12]. The steps are as follows.

1) *Extracting Video U-channel:* We embed the watermark in the U channel. A YUV 4:2:0 video can be written as:

$$m = \{y, u, v\}, \quad (7)$$

where  $y, u, v$  denote the Y, U, V components of the video, respectively. Thus the first step is to extract the U channel  $u$  of the original video  $m$ , which is a three-dimensional variable.

2) *Performing First Level 3-D DCT Based on Cubic Splits:* In the three-dimensional space,  $u$  is uniformly divided into  $N \times N \times N$  cubes without overlapping. A 3-D DCT is performed for each cube and a stable coefficient is selected from each transformed cube (three indexes are equal and even). The stable coefficients extracted from all cubes are formed into a new cube, which has the size  $N \times N \times N$ . It should be emphasized that only one stable coefficient is extracted from each cube and the frequency components of the coefficients extracted from all cubes should be consistent. The stable coefficients of the first level 3-D DCT are robust to the spatio-temporal distortion within the split cube, and in order to be robust to the global spatio-temporal distortion of the video, we need to perform the second level 3-D DCT, from which we further extract the stable coefficients that are robust to the global spatio-temporal distortion.

3) *Performing Second Level 3-D DCT on the Stable Coefficient:* We perform an entire 3-D DCT transform of the stable coefficient cube obtained in the second step. Following this, stable DCT coefficients are selected from the transformed coefficients (three indexes are even), resulting in a cube of size  $(N/2) \times (N/2) \times (N/2)$ .

4) *Obtain the Feature Vector:* We reshape the cube of size  $(N/2) \times (N/2) \times (N/2)$  obtained in the third step into a one-dimensional vector, which we call the feature vector, i.e., the carrier of our watermark embedding, denoted as  $F = [f_0, f_1, \dots, f_{L-1}]$ .  $L$  denotes the length of the feature vector.

5) *Generating the Spread Spectrum Watermark:* Assume that the original watermark is  $W = [w_0, w_1, \dots, w_{K-1}]$  with  $K$  components and  $w_i \in \{0, 1\}$ . We use the Schmidt orthogonalization method to generate  $K$  orthogonal coded vectors,  $S_i = [s_{i0}, s_{i1}, \dots, s_{iL}]$ , each of length  $L$ . Generally speaking  $L > K$ . The spread spectrum watermark is denoted as  $W' = [w'_0, w'_1, \dots, w'_{L-1}]$  with  $L$  elements:

$$w'_j = \sum_{i=0}^{K-1} b_i \cdot s_{ij}, b_i = \begin{cases} 1, & w_i = 1 \\ -1, & w_i = 0 \end{cases}, \quad (8)$$

where  $i = 0, 1, \dots, K$  and  $j = 0, 1, \dots, L$ .

6) *Embedding the Spread Spectrum Watermark:* The original feature vector and the spread spectrum watermark are added to obtain the feature vector after embedding the watermark, denoted as  $F' = [f'_0, f'_1, \dots, f'_{L-1}]$ , and the calculation process is as follows:

$$F' = F + \delta \cdot W', \quad (9)$$

where  $\delta \in (0, +\infty]$  is the embedding strength of the watermarking.

7) *Generating the Watermarked Video:* The corresponding DCT coefficients are updated utilizing the obtained watermarked feature vector  $F'$ . A two-level 3-D IDCT is performed to obtain the watermarked U channel  $u'$ . Finally, we combine the Y-channel and V-channel and obtain the watermarked video  $m'$  as

$$m' = \{y, u', v\}. \quad (10)$$

#### 3.2. Watermark Extraction

The extraction process is similar to the embedding process. Thus, the extraction process requires the use of an orthogonal coding vector  $S_i$  that is consistent with the embedding process. We extract the feature vector  $F^*$  from the watermarked video.  $F'$  and  $F^*$  will not be consistent exactly because the watermarked video is under unknown attack. The extracted watermark can be expressed as  $W^* = [w_0^*, w_1^*, \dots, w_K^*]$ , and the corresponding components of each  $w_i^*$  can be obtained by

$$w_i^* = \begin{cases} 1, & I(F^*, S_i) \geq 0 \\ 0, & I(F^*, S_i) < 0 \end{cases} \quad (11)$$

where  $I(\cdot)$  represents the inner product.

## 4. EXPERIMENTAL EVALUATION

### 4.1. Experimental Setup

**Data.** We evaluate our method on eight CIF videos with a resolution of  $288 \times 352$ , i.e., *Akiyo*, *Coastguard*, *Container*,

**Table 1.** Comparison of average PSNR and SSIM for four video watermarking schemes

Schemes	[12]	[8]	[11]	Our
PSNR(dB)	41.80	40.17	39.61	<b>44.09</b>
SSIM	0.9461	0.9261	0.9058	<b>0.9712</b>

**Table 2.** The robustness against different attacks

Sources of Distortions	Attack Types (Parameter)	BER	Authentication
Spatial Distortions	H.264/AVC (QP=28)	0.0000	16/16
	H.264/AVC (QP=36)	0.0020	16/16
	H.265/HEVC(QP=28)	0.0000	16/16
	H.265/HEVC (QP=36)	0.0293	13/16
	Cropping (10%)	0.0000	16/16
	Cropping (50%)	0.0000	16/16
	Flipping (Horizontal)	0.0000	16/16
	Aspect Ratio Change (0.5×0.6)	0.0000	16/16
	Aspect Ratio Change (1.3×1.4)	0.0000	16/16
	Scaling (0.5)	0.0000	16/16
	Rotation 90°	0.0000	16/16
Rotation 180°	0.0000	16/16	
Temporal Distortions	Trimming(50%)	0.0000	16/16
	Reversing	0.0000	16/16
	Frame Rate Conversion(fps=16)	0.0000	16/16
	Frame Rate Conversion(fps=40)	0.0000	16/16
Spatio-temporal Distortions	Cropping(50%)+Aspect Ratio Change(0.5×0.6)	0.0251	14/16
	Scaling(0.5)+Frame Rate Conversion(fps=16)	0.0000	16/16

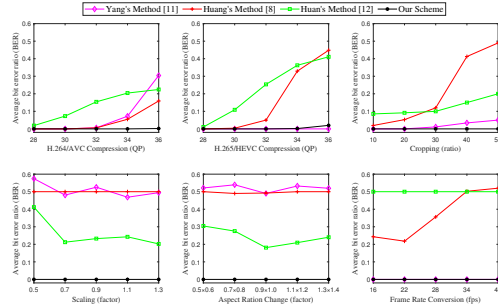
*Foreman, Hall, Mobile, Mother-daughter, Silent*, and eight HD videos with a resolution of  $720 \times 1280$ , i.e., *Vidyo1, Vidyo4, Kristen and Sara, In to Tree, Stockholm, Mobcal, Shields, Park Joy*. Each video contains 300 frames and has a frame rate of 25 fps [3, 12].

**Parameters.** The actual watermark payload is 36 bits. To allow the extractor to authenticate the integrity of the watermark, we use CRC-8 and BCH(45,63) to encode the watermark sequence. To make it easier for the comparison experiments, we fill one more bit, i.e. the length of the original watermark  $\mathbf{W}$ ,  $K = 64$ . When performing the first level of 3-D DCT based on a cube split, the size of the cube is  $16 \times 16 \times 16$ , i.e.  $N = 16$ . Thus, the length of the feature vector  $\mathbf{F}$  and the spread spectrum watermark  $\mathbf{W}'$  is  $(N/2) \times (N/2) \times (N/2) = 512$ . To better balance imperceptibility and robustness, after the first level of 3-D DCT, the frequency coefficients selected from each cube are  $AC_{2,2,2}$ . Besides, we set the embedding strength  $\delta = 2000$ .

**Comparison.** We compare our scheme with three existing state-of-the-art schemes including DT CWT-based video watermarking [12], pseudo-3-D DCT-based video watermarking [8], and DCT high-frequency-based video watermarking [11]. For a fair comparison, we embed the same 64 bits watermark message without any error correction code (ECC) encoded for all schemes.

## 4.2. Imperceptibility

The average peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) are given in Table 1. It can be clearly seen that our method achieves the highest PSNR value compared to the existing state-of-the-art schemes. the PSNR value of our scheme reaches 44.09 dB, which is an



**Fig. 3.** The comparison of robustness against various attacks among the four watermarking schemes.

average of about 3.6 dB higher than the existing schemes. Besides, the SSIM value of our scheme is also the highest among the four schemes, achieving 0.9712. It is evident from the PSNR and SSIM metrics that our scheme achieves the best imperceptibility.

## 4.3. Robustness

We assess the robustness of our scheme using various types of compression attacks, spatial attacks, temporal attacks, and combined attacks as shown in Table 2. In Table 2, “13/16” means that the watermark extracted from 13 out of 16 test videos passed the CRC authentication. In practice, videos are distributed in an encapsulated format. Thus, after embedding the watermark, we first compress the video using the H.264/AVC (QP=28) encoder and perform the attacks in Table 2 on the encapsulated video via FFmpeg, and then finally perform the watermark extraction. From Table 2, we can observe that the proposed watermarking scheme has strong robustness in the presence of various types of attacks. To verify the superiority of our scheme in terms of robustness, we compare the bit error rate (BER) of our scheme with existing schemes under various types of attacks, and the results are shown in Fig. 3. The experimental results demonstrate that our scheme has stronger robustness compared to existing schemes, with a higher PSNR value and SSIM value.

## 5. CONCLUSION

In this paper, we propose a novel video watermarking scheme by exploring the consistency law of spatio-temporal distortion and stable 3-D DCT coefficients. We discover the consistency between spatial and temporal distortion and propose to solve the temporal distortion using the method of solving spatial distortion. Besides, we explore the 3-D DCT stable coefficients, which are robust to spatio-temporal distortion. Experimental results show that our method achieves better imperceptibility and stronger robustness compared to the existing state-of-the-art schemes. In the future, we will develop an content-adaptive mechanism in the video watermarking.

## 6. REFERENCES

- [1] Mingze He, Hongxia Wang, Fei Zhang, Sani M Abdullahi, and Ling Yang, "Robust blind video watermarking against geometric deformations and online video sharing platform processing," *IEEE Transactions on Dependable and Secure Computing*, 2022.
- [2] Pierre Fernandez, Alexandre Sablayrolles, Teddy Furon, Hervé Jégou, and Matthijs Douze, "Watermarking images in self-supervised latent spaces," in *2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 3054–3058.
- [3] Md Asikuzzaman, Md Jahangir Alam, Andrew J Lambert, and Mark Richard Pickering, "Imperceptible and robust blind video watermarking using chrominance embedding: a set of approaches in the DT CWT domain," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 9, pp. 1502–1517, 2014.
- [4] Md Asikuzzaman, Md Jahangir Alam, Andrew J Lambert, and Mark R Pickering, "A blind watermarking scheme for depth-image-based rendered 3D video using the dual-tree complex wavelet transform," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 5497–5501.
- [5] Tae-Woo Oh, Min-Jeong Lee, Kyung-Su Kim, Young-Suk Yoon, and Heung-Kyu Lee, "Spatial self-synchronizing video watermarking technique," in *2009 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2009, pp. 4233–4236.
- [6] Ce Wang, Chao Zhang, and Pengwei Hao, "A blind video watermark detection method based on 3D-DWT transform," in *2010 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2010, pp. 3693–3696.
- [7] Fei Zhang, Hongxia Wang, Ling Yang, and Mingze He, "Robust blind video watermarking by constructing spread-spectrum matrix," in *2022 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, 2022, pp. 1–6.
- [8] Hui-Yu Huang, Cheng-Han Yang, and Wen-Hsing Hsu, "A video watermarking technique based on pseudo-3-D DCT and quantization index modulation," *IEEE Transactions on information forensics and security*, vol. 5, no. 4, pp. 625–637, 2010.
- [9] Nilkanta Sahu and Arijit Sur, "Sift based video watermarking resistant to temporal scaling," *Journal of Visual Communication and Image Representation*, vol. 45, pp. 77–86, 2017.
- [10] Faride Madine, Mohammad Ali Akhaee, and Nematollah Zarmehi, "A multiplicative video watermarking robust to H. 264/AVC compression standard," *Signal Processing: Image Communication*, vol. 68, pp. 229–240, 2018.
- [11] Ling Yang, Hongxia Wang, Yulin Zhang, Jinhe Li, Peisong He, and Sijiang Meng, "A robust DCT-based video watermarking scheme against recompression and synchronization attacks," in *Digital Forensics and Watermarking: 20th International Workshop, IWDW 2021, Beijing, China, November 20–22, 2021, Revised Selected Papers*. Springer, 2022, pp. 149–162.
- [12] Wennan Huan, Sheng Li, Zhenxing Qian, and Xinpeng Zhang, "Exploring stable coefficients on joint sub-bands for robust video watermarking in DT CWT domain," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 1955–1965, 2021.
- [13] Frederic Deguillaume, Gabriela Csurka, Joseph JK O'Ruanaidh, and Thierry Pun, "Robust 3D DFT video watermarking," in *Security and Watermarking of Multimedia Contents*. SPIE, 1999, vol. 3657, pp. 113–124.
- [14] Jae Hyuck Lim, Dae Jin Kim, Hyun Tae Kim, and Chee Sun Won, "Digital video watermarking using 3D-DCT and intracubic correlation," in *Security and Watermarking of Multimedia Contents III*. SPIE, 2001, vol. 4314, pp. 64–72.
- [15] Patrizio Campisi and Alessandro Neri, "Video watermarking in the 3D-DWT domain using perceptual masking," in *IEEE International Conference on Image Processing 2005*. IEEE, 2005, vol. 1, pp. I–997.
- [16] Ying Li, Xinbo Gao, and Hongbing Ji, "A 3D wavelet based spatial-temporal approach for video watermarking," in *Proceedings Fifth International Conference on Computational Intelligence and Multimedia Applications. ICCIMA 2003*. IEEE, 2003, pp. 260–265.
- [17] Md Asikuzzaman and Mark R Pickering, "An overview of digital video watermarking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2131–2153, 2017.