



# Robust Video Watermarking Using Normalized Zernike Moments

Shiyi Chen<sup>1</sup>, Yi Chen<sup>2</sup>, Yanli Chen<sup>3</sup>, Limengnan Zhou<sup>4</sup>, and Hanzhou Wu<sup>1</sup>(✉)

<sup>1</sup> School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China

[h.wu.phd@ieee.org](mailto:h.wu.phd@ieee.org)

<sup>2</sup> School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China

<sup>3</sup> School of Big Data and Computer Science, Guizhou Normal University, Guiyang 550001, China

<sup>4</sup> School of Electronic and Information Engineering, University of Electronic Science and Technology of China, Zhongshan Institute, Zhongshan 528402, China

**Abstract.** Digital video watermarking has become a hot research topic in recent years due to the increasing demand of protecting the intellectual property of video data. Even though many conventional video watermarking methods have been reported in past years, few of them are resistant to high-intensity geometric attacks, which motivates the authors in this paper to propose a video watermarking technique that is robust against high-intensity geometric distortion. To this purpose, the proposed method embeds the watermark information into the normalized Zernike moments of the target frames of the cover video sequence. The advantage is that the normalized Zernike moment preserves a strong invariance to geometric distortions such as rotation and scaling attacks. During data embedding, secret bits are embedded into adaptively selected moments with slight modifications to provide good robustness while maintaining the imperceptibility. The chrominance channel of the video data rather than the luminance one is used in our algorithm, as distortion in the former channel is less sensitive to the human visual system. Experimental results show that, compared with the existing scheme, the PSNR values of the proposed method gain about 7 dB averagely, meaning that the proposed method achieves high imperceptibility. Moreover, it is demonstrated that the proposed method is more robust against geometric distortions such as rotation and upscaling, which has verified the applicability and superiority of the proposed work.

**Keywords:** Video watermarking · Robust · Zernike moments · Information hiding · Copyright protection · Multimedia security

## 1 Introduction

Video watermarking is a technique for protecting digital video data from piracy. As illegal distribution of copyrighted digital video is ever-growing, video water-

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

X. Sun et al. (Eds.): ICAIS 2022, LNCS 13340, pp. 323–336, 2022.

[https://doi.org/10.1007/978-3-031-06791-4\\_26](https://doi.org/10.1007/978-3-031-06791-4_26)

marking attracts an increasing amount of attention within the information security community. Over the last decade, various watermarking techniques have been introduced for copyright protection and data authentication. Based on the domain where the watermark information is embedded, this technique can be divided into three main classes: compressed, spatial and transform domain [1].

Among the above-mentioned three categories, transform domain algorithm is widely used due to its effectiveness in maintaining robustness against various attacks. The most commonly used transforms are singular value decomposition (SVD), discrete Fourier transform (DFT), discrete Cosine transform (DCT), discrete wavelet transform (DWT) and dual-tree complex wavelet transform (DT CWT) [1]. In [2], Huan *et al.* introduce an algorithm applying SVD on the DT CWT domain. In [3], Bhaskar *et al.* proposed a robust video watermarking scheme with squirrel search algorithm. For lack of a strong rotational invariance proved by mathematical principles, these methods are not robust to rotation attacks with a large angle, while this property is possessed by Zernike moment. In the image watermarking field, methods based on Zernike moment has been widely used, for example, in [4–6]. In [7], the author proposed a video watermarking algorithm based on Zernike moment. However, this algorithm only resists against rotation attacks rather than scaling attacks. Therefore, although contributions expended by predecessors have exceeded the development of robust watermarking techniques, problems like resistance to geometric attacks are still challenging in the video watermarking community and need further research.

In this paper, we propose a robust video watermarking algorithm based on normalized Zernike moments to resist against geometric distortions. Since different video compression algorithms are used on the Internet, the proposed algorithm is designed in the uncompressed domain for suiting any video compression standard. Because of the geometric-invariant property proved by mathematical principles, normalized Zernike moments are employed in our method as invariant features. For watermark embedding and extraction, Dither Modulation-Quantization Index Modulation (DM-QIM) is employed in the algorithm by using dither vectors and modulating the Zernike moments into different clusters to make an adequate trade-off between robustness and distortion [8]. To achieve high visual quality, we embed the watermark information into the U channel of the cover video sequence because distortion in luminance is more noticeable than that in chrominance as for human visual system [9]. The experimental results show that our approach maintains good visual quality and achieves great robustness to geometric attacks with high intensity comparing to the prior work.

The remainder of this paper is organized as follows. The preliminary knowledge related to the scheme is discussed in Sect. 2. In Sect. 3, we introduce the proposed video watermarking approach in detail. While in Sect. 4, experiments for imperceptibility and robustness evaluation of the proposed scheme is conducted. Finally, conclusions and future work are drawn in Sect. 5.

## 2 Preliminaries

In this section, we describe the preliminary knowledge of the proposed algorithm, which can be separated into four parts. In each part, we demonstrate the main contents and explain the reason why we use them.

### 2.1 Geometric Attacks

When watermarked videos are available online, some kinds of content-preserving attacks may be applied, which inevitably reduce the energy of the watermark inside the transmitted videos [6]. Among all these distortions, geometric attack is a relatively challenging one, since a slight geometric deformation often fails the watermark detection. In this paper, for practical applications, we mainly discuss the most common geometric attacks: rotation and scaling attacks.

Geometric attack is defined by a set of parameters that determines the operation performed over the target document, for example, scaling attack can be characterized by applying a scaling ratio to the sampling grid, and a similar conclusion can be given to rotation attacks. These common geometric attacks will cause two typical distortions in the document. One is the shifting of pixels in the spatial plane. The other is alteration of the pixel values due to interpolation [10]. Hence withstanding an arbitrary displacement of all or some of its pixels by a random amount is the main concern for resistance to geometric deformations.

### 2.2 Zernike Moments

In our method, we use normalized Zernike moments for data embedding due to its geometric invariance proved by mathematical principles, which is a modification of Zernike moments. Therefore, we first introduce Zernike moments in this part.

Zernike moments are orthogonal moments based on Zernike polynomial, which is a complete orthogonal set over the interior of the unit circle [11]. The set of these polynomials can be denoted in the following equation:

$$V_{nm}(x, y) = R_{nm}(\rho)e^{jm\theta} \quad (1)$$

where  $x, y$  denote the pixel position,  $\rho = \sqrt{x^2 + y^2}$ , and  $\theta = \tan^{-1}(y/x)$ .  $n$  is a non-negative integer which represents the order and  $m$  is the repetition designed to satisfy the fact that  $n - |m|$  is both non-negative and even.  $R_{nm}(\rho)$  are radial Zernike polynomials, which are given by the equation below:

$$R_{nm}(\rho) = \sum_{s=0}^{n-|m|/2} \frac{(-1)^s [(n-s)!] \rho^{n-2s}}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \quad (2)$$

After computing Zernike polynomials in Eq. (1), we can get the Zernike moments of order  $n$  with repetition  $m$  for a continuous image function:

$$A_{nm} = \frac{n+1}{\pi} \int \int_{x^2+y^2 \leq 1} f(x, y) V_{nm}^*(\rho, \theta) dx dy \quad (3)$$

where  $V_{nm}$  represents the Zernike polynomial, and  $*$  denotes complex conjugate.

For digital signal, the integrals are replaced by summations. Since the Zernike polynomial is a set over the interior of the unit circle, where each frame is reconstructed. By utilizing the properties of the Zernike polynomial set discussed above, frame image  $f(x, y)$  can be reconstructed to  $\hat{f}(x, y)$  in Eq. (4).

$$\hat{f}(x, y) = \sum_{n=0}^N \sum_m A_{nm} V_{nm}(\rho, \theta) \tag{4}$$

where  $A_{nm}$  represents the Zernike moments of order  $n$  with repetition  $m$ . A larger  $N$  results in a reconstruction result with more accuracy.

### 2.3 Invariant Properties of Normalized Zernike Moment

Based on the mathematical definition of Zernike moment, the amplitude of which can be used as a rotation-invariant feature. By utilizing the normalization technique, the normalized Zernike moments are invariant to both rotation and scaling attacks. The certification process is addressed in detail as follows.

**Rotation Invariance.** From Eq. (3),  $A_{nm}$  can be simplified as  $A_{nm} = |A_{nm}|e^{jm\theta}$ . After rotating each frame image clockwise by angle  $\alpha$ , the relationship between the original and rotated frames in the same polar coordinate becomes

$$A'_{nm} = A_{nm}e^{-jm\alpha} \tag{5}$$

which means after rotation, the amplitude of the Zernike moment remains the same. As a result, it can be used as a rotation-invariant feature of each frame.

**Scaling Invariance.** After scaling the size of an image, the nonlinear interpolation will convert the content of the unit circle from the original one, which means that Zernike moments are not robust to scaling deformations.

$$m_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} x^p y^q f(x, y) \tag{6}$$

To achieve scaling invariance, we can normalize each frame as shown in [12] before computing the Zernike moments. The normalization phase of which is concluded in a detailed way by the following four steps:

Step 1) Center the image by transforming  $f(x, y)$  to  $f_1(x, y) = f(x - \bar{x}, y - \bar{y})$ .  $(\bar{x}, \bar{y})$  is the centroid of  $f(x, y)$ , which can be calculated below.

$$\bar{x} = \frac{m_{10}}{m_{00}}, \bar{y} = \frac{m_{01}}{m_{00}} \tag{7}$$

where  $m_{10}, m_{01}$  and  $m_{00}$  are the moments of  $f(x, y)$  as defined in Eq. (6).

Step 2) Apply a shearing transform from  $f_1(x, y)$  to  $f_2(x, y)$  in the  $x$  direction using Eq. (8) with  $A_x = \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix}$  making sure that the  $\mu_{30}$  of the resulting image is zero, which stands for central moments and is described in Eq. (9).

$$g(x, y) = A \cdot f(x, y) \quad (8)$$

Step 3) Apply a shearing transform from  $f_2(x, y)$  to  $f_3(x, y)$  in the  $y$  direction with  $A_y = \begin{pmatrix} 1 & 0 \\ \gamma & 1 \end{pmatrix}$  so that the  $\mu_{11}$  of the resulting frame reaches zero.

$$\mu_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (9)$$

Step 4) Scale  $f_3(x, y)$  in both  $x$  and  $y$  directions with  $A_s = \begin{pmatrix} \alpha & 0 \\ 0 & \delta \end{pmatrix}$  to a prescribed standard size and achieve  $\mu_{50} > 0$  and  $\mu_{05} > 0$  from the outcome.

In [12], it is proved that image and its affine transforms have the same normalized image. Consequently, when it is employed in video algorithms, the same conclusion can be drawn. As a result, after normalization, the amplitude of the Zernike moments stays invariant to both rotation and scaling attacks.

## 2.4 Quantization Index Modulation

Quantization Index Modulation (QIM) [8] is an embedding operation used for information hiding, which preserves provably better rate distortion-robustness trade-offs than spread-spectrum and low-bit(s) modulation methods. In this paper, we use the modification of QIM: Dither Modulation (DM)-QIM algorithm. In this subsection, we introduce the basic theory of DM-QIM as follows:

**Embedding Procedure.** Suppose  $f(n, m)$  is an image, where  $n \in [1, N]$ ,  $m \in [1, M]$  and  $W(k)$ ,  $k \in [1, N \times M]$ , which is used as watermark. Let  $d(k)$  be an array of uniformly distributed pseudo-random integers chosen within  $[-\Delta/2, \Delta/2]$ , which is generated according to a secret key. Dither vectors  $d_0(k)$  and  $d_1(k)$  are used for embedding the ‘0’ and ‘1’ bits of the watermark respectively. For simplicity, we combine the two vectors into  $d_{W(k)}(k)$ .

$$d_0(k) = d(k) \quad (10)$$

$$d_1(k) = d_0(k) - \text{sign}(d_0(k)) \times \frac{\Delta}{2} \quad (11)$$

where  $f^w(n, m)$  denotes the watermarked image and  $\Delta$  represents the quantization step, which is the most important parameter of QIM. The watermark embedding operation is performed below in Eq. (12).

$$f^w(n, m) = Q(f(n, m) + d_{W(k)}(k), \Delta) - d_{W(k)}(k) \quad (12)$$

where  $Q(x, y)$  is defined below, and  $round(x)$  returns the nearest integer of  $x$ .

$$Q(x, \Delta) = \Delta \times round\left(\frac{x}{\Delta}\right) \tag{13}$$

**Extraction Procedure.** To extract the watermark data, we put the watermark bits ‘0’ and ‘1’ into (8) using the watermarked frame as an input instead of the original one, and then estimate the errors between the watermarked image and the results obtained above. By comparing the two errors, the one with the lower value represents the watermark bit. The extraction procedure is concluded below.

$$g^{W(k)}(n, m) = Q(\tilde{f}^w(n, m) + d_{W(k)}(k), \Delta) - d_{W(k)}(k) \tag{14}$$

$$\tilde{W}(k) = argmin_{p \in [0,1]} \left| \tilde{f}^w(n, m) - g^p(n, m) \right| \tag{15}$$

where  $\tilde{f}^w(n, m)$  denotes the frame we received, and  $g^{W(k)}(n, m)$  is used to calculate the watermark value in Eq. (15).  $d_{W(k)}(k)$  is a dither vector used for embedding the watermark bit,  $\Delta$  represents the quantization step, and both of them should be the same as the embedding procedure.  $argmin(x)$  means the independent variable that minimizes the value of  $x$ .

### 3 Proposed Method

In this section, we introduce the proposed video watermarking algorithm in terms of embedding and extraction, which is demonstrated below separately.

#### 3.1 Watermark Embedding

The watermark embedding procedure is demonstrated in Fig. 1, and some of the steps in the block diagram is explained in the following subsections.

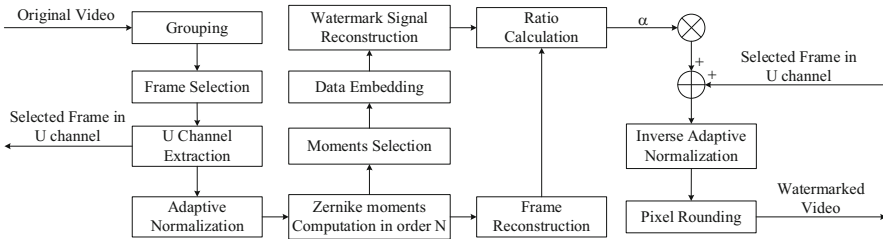


Fig. 1. Proposed robust watermarking framework

**U Channel Extraction.** In YUV format, Y represents the luminance channel and U, V are the two independent chrominance channels. As distortion in the chrominance channel is less sensitive to the human visual system than the luminance one [9], we extract the U channel in a YUV represented video for watermark embedding to enhance imperceptibility in the proposed method.

The following equation shows how to generate YUV signals from RGB sources:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0 \\ 127 \\ 127 \end{bmatrix} + \begin{bmatrix} 0.2989 & 0.5866 & 0.1145 \\ -0.1688 & -0.3312 & 0.5000 \\ 0.5000 & -0.4184 & -0.0816 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (16)$$

**Adaptive Normalization.** The adaptive normalization is almost identical as the procedure described in Sect. 2.3, except for step (4). After experiments, it is found that if videos in low resolution are normalized to a size much larger than the original one, more distortions will be produced.

To deal with this issue, the standard size  $M \times M$  mentioned in Sect. 2.3, step (4) is set adaptively based on the video size. For example, if the U channel of the input video sequence is in a resolution of  $176 \times 144$ . Empirically,  $M$  can be set as 256 for higher accuracy. For different requests,  $M$  is open for adjusting.

**Moments Selection.** In [13], it is said that Zernike moments with repetition  $m = 4j$ ,  $j$  integer, will deviate from orthogonality, meaning these moments cannot be computed accurately. In [14],  $|A_{00}|$  and  $|A_{11}|$  are independent of image, for that reason, they are not appropriate for watermark embedding. Given that conclusions can be drawn from Eq. (3) that  $|A_{n,m}| = |A_{n,-m}|$ , where  $|x|$  denotes the amplitude, we can dismiss the latter ones so as to eliminate the embedding modifications. From what has been discussed above, we remove these moments to maximize the applicability and superiority of our algorithm.

**Data Embedding.** After selection, we embed watermark data into the amplitude of all these selected moments using DM-QIM, which has already been extensively discussed in Sect. 2.3. In this paper, watermark data for each target frame contains 1 bit, this embedding operation can be described below.

$$|A_{n,m}^w| = Q(|A_{n,m}| + d_w, \Delta) - d_w \quad (17)$$

where the superscript  $w$  indicates that it is the value after embedding.  $Q(x, y)$  is a quantizer defined in Eq. (13), and  $w = 0, 1$  represents the watermark bit.  $\Delta$  is a quantization step, which is set based on the value of  $|A_{n,m}|$ , and  $d_w \in [-\Delta/2, \Delta/2]$  is dither vector used for embedding watermark bit  $w$ .

**Watermark Signal Reconstruction.** The watermark signal  $I_w(x, y)$  is reconstructed using Eq. (4) and multiplied with a coefficient based on the amplitude of both the original and the watermarked moment, which is demonstrated below.

$$I_w(x, y) = \sum_n^{N_{max}} \sum_m \left( \frac{|A_{nm}^w|}{|A_{nm}|} - 1 \right) \times A_{nm} V_{nm}(\rho, \theta) \quad (18)$$

where  $x, y$  denote the pixel position, and  $A_{nm}$  represents the Zernike moment in order  $n$  and repetition  $m$ , while the one with a superscript  $w$  indicates that it is the value after embedding the watermark.  $V_{nm}(\rho, \theta)$  denotes the Zernike polynomial, with  $\rho = \sqrt{x^2 + y^2}$ , and  $\theta = \tan^{-1}(y/x)$ .

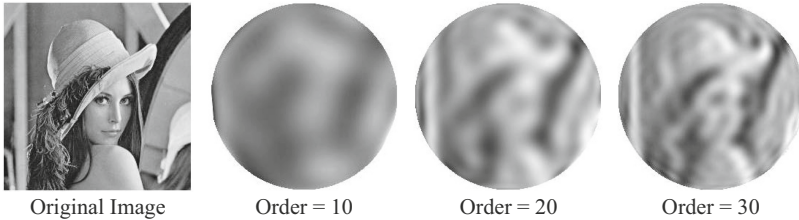
Finally, the watermark signal is added to the target frame with a coefficient  $\alpha$  designed for controlling the embedding strength of the watermark and ensures the imperceptibility. We calculate the value of  $\alpha$  with the following equation.

$$\alpha = \frac{\Theta I(x, y)}{\Theta I_r(x, y)} \tag{19}$$

where  $x^2 + y^2 \leq 1$  and  $\Theta(x)$  returns the mean value of  $x$ .  $I(x, y)$  is the original frame image where all the data is in the unit circle.  $I_r$  is demonstrated in Eq. (20), which represents the reconstructed frame of the original one without selecting the appropriate moments for watermark embedding.

$$I_r(x, y) = \sum_n^{N_{max}} \sum_m A_{nm} V_{nm}(\rho, \theta) \tag{20}$$

In order to ensure visual quality, we choose to add the reconstruction to the original frame instead of replacement, since it is limited in the unit circle and the reconstruction effect is far from satisfactory even with a high order. This conclusion is verified in Fig. 2, which takes a  $256 \times 256$  image of ‘Lena’ as an example to illustrate the reconstruction results with different orders. Furthermore, the reconstruction phase is rather time-consuming, for instance, when order is 30, it takes over 10s to reconstruct only one image. So it is not a brilliant choice to replace the watermarked signal with the original one for data embedding.



**Fig. 2.** Reconstruction of ‘Lena’ with different orders

To sum up, the embedding procedure can be concluded as follows:

- Step 1) Divide the input video into groups and select the target frames.
- Step 2) Perform adaptive normalization for calculating Zernike moments.
- Step 3) Calculate the Zernike moments from the normalized frame and select the appropriate ones as invariant features for watermark embedding.
- Step 4) Compute the amplitude of the selected moments and embed the same watermark bit into all of them, using DM-QIM to embed watermark bit.

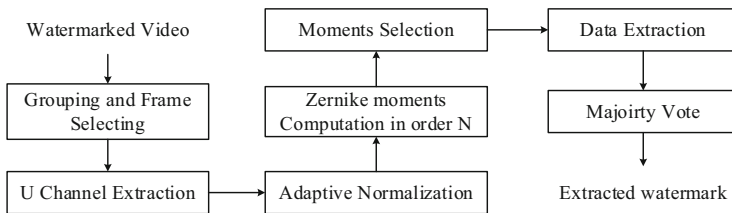


Step 5) Reconstruct the watermarked moments as the watermark signal and add it to the original frame with a coefficient  $\alpha$  defined in Eq. (19).

### 3.2 Watermark Extraction

In Fig. 3, the process of watermark extraction is introduced, which is similar to the embedding procedure. After computing the Zernike moments and selecting the appropriate ones of each frame, we extract the watermark bit from each moment using the extraction step in DM-QIM by Eq. (14) and (15).

**Majority Vote.** From all the watermark bits extracted from the selected moments in one frame, to dismiss the mutations, we choose the one with the highest frequency as the extracted watermark bit of each frame for more accuracy.



**Fig. 3.** Watermark extraction from Zernike Moments of the watermarked frame

The extraction procedure can be concluded in the following five steps:

Step 1) Divide the input video into groups and select the target frames.

Step 2) Perform adaptive normalization for calculating Zernike moments.

Step 3) Calculate the Zernike moments from the normalized frame and select the same moments used in watermark embedding procedure for extraction.

Step 4) Compute the amplitude of the selected moments and extract all the watermark bits using DM-QIM extraction method described in Sect. 2.4.

Step 5) Using Majority Vote to select the watermark bit with the highest frequency as the final extraction watermark bit for each target frame.

## 4 Experimental Results and Analysis

In this section, in order to evaluate the effectiveness of the proposed method, we conceive experiments to analyze the imperceptibility and robustness against geometric attacks by comparing our scheme with the existing approach [2].

## 4.1 Experimental Setup

All the experiments in this paper is implemented in the environment of Matlab R2016a on a PC with 8 GB RAM and 2.3GHz Intel Core i5 CPU, running on 64-bit Windows 10. To evaluate our method fairly, we selected six standard video sequences in CIF format ( $352 \times 288$ ), i.e., *Akiyo*, *Foreman*, *Hall*, *Mother and Daughter*, *Paris* and *Silent* [15], and each testing video contains 300 frames.

For simulation, we normalize the frame image of each video in U channel to  $256 \times 256$  and set the GOP (Group of Picture) length as 6, while the watermark length is 50, which is generated pseudo-randomly using a key, so that each GOP carries one watermark bit. After preliminary experiment in Sect. 4.2, we set the step length  $\Delta$  to 30000, with  $d_0 = 0$ ,  $d_1 = 15000$ . For simplicity, we embed each watermark bit into the first frame of one GOP, which represents the index frame. For fair comparison, the GOP in prior work [2] is also set to 6. The embedding strength  $T$  in [2] is set by 400 as the recommended value.

## 4.2 Parameter Setting

In this section, we conduct an experiment to find the optimal setting of  $\Delta$ , which is the most important parameter in our scheme. To evaluate the accuracy of the extracted watermark, Normalized Cross Correlation (NCC) is exploited as a standard, which is demonstrated below by Eq. (21).

$$NCC(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X) \cdot Var(Y)}} \quad (21)$$

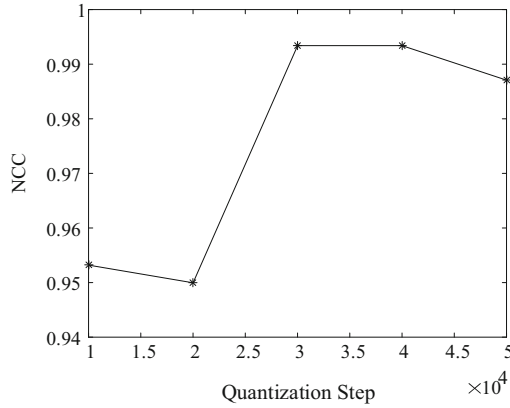
where  $Cov(X, Y)$  denotes the covariance between image  $X$  and  $Y$ ,  $Var(X)$  means the variance of  $X$ . The value range of NCC is  $[-1, 1]$  where ‘1’ means complete match and ‘-1’ indicates that the two images are exactly the opposite.

To evaluate the performance of different quantization step values, we use the six standard test video sequences mentioned above in Sect. 4.1, and the other parameters remain unchanged. In Fig. 4, it can be seen that the NCC value of our proposed method changes with the increase of quantization step and a maximum NCC value is reached by setting  $\Delta$  to 30000 and 40000. Since the accuracy of the extracted watermark increases with the NCC value, so both can be chosen as the best quantization step. In the following discussions, we set  $\Delta = 30000$  for experiment unless specific statement is presented.

## 4.3 Imperceptibility

For practical application, watermark imperceptibility is a very important requirement of a digital video watermarking algorithm. In this subsection, we adopt the peek signal-to-noise ratio (PSNR) as the standard to measure the visual quality of the final watermarked video, which is demonstrated below.

$$PSNR = 10 \cdot \log_{10} \frac{max^2}{E}, E = \frac{1}{m \cdot n} \sum_{i=1}^m \sum_{j=1}^n |I(i, j) - K(i, j)|^2 \quad (22)$$



**Fig. 4.** NCC of the watermarked video for our scheme in different quantization steps

where  $I(i, j)$  and  $K(i, j)$  represent two different images, and  $|x|$  represents the absolute value of  $x$ .  $m, n$  denotes the height and width of each frame, and  $max$  indicates the upper limit value of the pixel in each frame image.

The experimental results on the test video sequences are listed in Table 1. It can be concluded from the table that, the PSNR of watermarked videos in this paper are controlled around 37 dB, while the counterpart in [2] is 30 dB. Therefore, the PSNR values in our method gain about 7 dB averagely comparing to the prior work. Consequently, in terms of PSNR, our scheme outperforms the existing scheme [2] in imperceptibility, which results in better visual effects.

**Table 1.** PSNR of the proposed method and the existing work after watermarking

PSNR(dB)	Proposed method	Huan <i>et al.</i> [2]
Akiyo	38.9965	32.4720
Foreman	36.2325	30.1227
Hall	37.2345	29.4552
Mother and Daughter	37.1260	28.2724
Paris	36.5833	30.1578
Silent	36.9929	29.4093

#### 4.4 Geometric Robustness

In this subsection, experiments are conducted to analyze the robustness of our method against geometric attacks. The experimental data in Table 2 is obtained by averaging the results of the aforementioned six standard test video sequences.

**Table 2.** NCC of the proposed method and the existing work after Scaling attack

NCC	Without attacks	Scaling 150%	Scaling 200%	Scaling 300%	Scaling 400%
Proposed	0.9934	0.9868	0.9934	0.9934	0.9870
Huan <i>et al.</i> [2]	0.9796	0.8560	0.9265	0.6935	0.6102

From Table 2, it can be observed that, before attacks, the accuracy in our method is 2% higher than that in approach [2]. After scaling attacks, the NCC value in our method relatively maintains the same value as the one without attacks. While in approach [2], the result decrease remarkably as the scaling factor increases from 200% to 300% and 300% to 400%. When the scaling factor reaches 400%, the NCC value in [2] is lower than our approach by 40%. As a result, the proposed method outperforms [2] notably in scaling attacks.

**Table 3.** NCC of the proposed method and the existing work after Rotation Attack

NCC	Rotate 30°	Rotate 60°	Rotate 90°	Rotate 120°
Proposed	0.9415	0.9356	0.9747	0.9287
Huan <i>et al.</i> [2]	0.6040	0.4746	0.4045	0.1540

In Table 3, the NCC of the proposed method and the existing work after rotation attack is given. It can be concluded that when the rotation angle increases, the NCC value in our method can be maintained, which is slightly lower than the one without any attack. While in [2], the NCC value drops significantly as the rotation angle rises, especially from 90° to 120°, which is lower than the value in our method from 60% to 80%. After the analysis above, our method performs remarkably better than [2] in robustness against rotation attacks.

From all the conclusions discussed above, our method outperforms [2] in both imperceptibility and geometric robustness against scaling and rotation. Therefore, we can jump to the conclusion that our method has a relatively good visual effect and a great robustness against geometric attacks.

## 5 Conclusion

In this paper, we propose a novel video watermarking scheme, which combines the benefits of both Zernike moments and normalization to resist against geometric distortions. Zernike moments are employed for its special invariant properties against rotation attacks. Normalization is used to normalize the target frame, so that the normalized Zernike moment is robust to both scaling and rotation attacks. After calculating the Zernike moments, we select some of the appropriate ones for watermark embedding according to certain principles to improve robustness and reduce modifications. With the heavy computation load and low

accuracy for reconstructing Zernike moments, we use them to design watermark signal. The watermark embedded in each frame is obtained by taking the one with the highest frequency from all the candidate watermarks extracted from the amplitude of the selected Zernike moments to avoid errors. Based on the experimental results, our approach maintains good visual quality and achieves great robustness to rotation and scaling attacks comparing to the prior work.

In our method, it is shown that we apply a couple of forward and inverse normalizations, which produce inevitable distortions. To deal with this problem, it is necessary to design a new watermarking strategy to eliminate the loss. Meanwhile, as Zernike moment is computationally expensive, a watermark embedding algorithm with more efficiency needs to be explored. In future works, we will focus on the optimization for both preciseness and efficiency.

**Acknowledgements.** This work was partly supported by the National Natural Science Foundation of China (Grant Nos. 61901096, 62102112 and 61902235), and the Shanghai “Chen Guang” Program (Grant No. 19CG46).

## References

1. Asikuzzaman, M., Pickering, M.: An overview of digital video watermarking. *IEEE Trans. Circuits Syst. Video Technol.* **28**(9), 2131–2153 (2018)
2. Huan, W., Li, S., Qian, Z., Zhang, X.: Exploring stable coefficients on joint sub-bands for robust video watermarking in DT CWT domain. *IEEE Trans. Circ Syst. Video Technol.* **32**(4), 1955–1965 (2021)
3. Bhaskar, A., Sharma, C., Mohiuddin, K., Singh, A., Nasr, O.A.: A robust video watermarking scheme with squirrel search algorithm. *Comput. Mater. Continua* **71**(2), 3069–3089 (2022)
4. Kim, H., Lee, H.: Invariant image watermark using Zernike moments. *IEEE Trans. Circuits Syst. Video Technol.* **13**(8), 766–775 (2003)
5. Xiong, L., Han, X., Yang, C., Shi, Y.: Robust reversible watermarking in encrypted image with secure multi-party based on lightweight cryptography. *IEEE Trans. Circ. Syst. Video Technol.* **32**(1), 75–91 (2021)
6. Hu, R., Xiang, S.: Cover-lossless robust image watermarking against geometric deformations. *IEEE Trans. Image Process.* **30**, 318–331 (2021)
7. Xu, G., Wang, R.: A blind video watermarking algorithm resisting to rotation attack. In: *Proceeding of International Conference on Computer and Communications Security*, pp. 111–114 (2009)
8. Chen, B., Wornell, G.W.: Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. *IEEE Trans. Inf. Theory* **47**(4), 1423–1443 (2001)
9. Parraga, C.A., Brelstaff, G., Troscianko, T., Moorhead, I.R.: Color and luminance information in natural scenes. *J. Opt. Soc. Am. A Opt. Image Sci. Vision.* **15**(3), 563–569 (1998)
10. Xiang, S., Joong Kim, H., Huang, J.: Invariant image watermarking based on statistical features in the low-frequency domain. *IEEE Trans. Circ. Syst. Video Technol.* **18**(6), 777–790 (2008)
11. Khotanzad, A., Hong, Y.: Invariant image recognition by Zernike moments. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(5), 489–497 (1990)

12. Dong, P., Brankov, J., Galatsanos, N., Yang, Y., Davoine, F.: Digital watermarking robust to geometric distortions. *IEEE Trans. Image Process.* **14**(12), 2140–2150 (2005)
13. Xin, Y., Liao, S., Pawlak, M.: Geometrically robust image watermarking on a circular domain. *Pattern Recogn. Lett.* **40**(1), 3740–3752 (2007)
14. He, W., Sun, J., Yang, Z., Yang, D.: Video watermarking scheme based on normalization of pseudo-Zernike moment. In: *Proceeding of International Conference on Measuring Technology and Mechatronics Automation*, pp. 1080–1082 (2010)
15. Derf's Test Media Collection. <https://media.xiph.org/video/derf/>
16. Kamila, N., Mahapatra, S., Nanda, S.: RETRACTED: invariance image analysis using modified Zernike moments. *Pattern Recogn. Lett.* **26**(6), 747–753 (2005)
17. Asikuzzaman, M., Alam, M., Lambert, A., Pickering, M.: Imperceptible and robust blind video watermarking using chrominance embedding: a set of approaches in the DT CWT domain. *IEEE Trans. Inf. Forensics Secur.* **9**(9), 1502–1517 (2014)
18. Yuan, X., Pun, C.: Feature based video watermarking resistant to geometric distortions. In: *Proceeding of IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, pp. 763–767 (2013)
19. Lin, C., Wu, M., Bloom, J., Cox, I., Miller, M., Lui, Y.: Rotation, scale, and translation resilient watermarking for images. *IEEE Trans. Image Process.* **10**(5), 767–782 (2001)
20. Zhao, Y., Wang, S., Zhang, X., Yao, H.: Robust hashing for image authentication using Zernike moments and local features. *IEEE Trans. Inf. Forensics Secur.* **8**(1), 55–63 (2013)
21. Zhou, X., Wang, L.: SoRS: an effective SVD-DWT watermarking algorithm with SVD on the revised singular value. In: *Proceeding of IEEE International Conference on Software Engineering and Service Science*, pp. 997–1002 (2014)
22. Keyvanpour, M.R., Khanbani, N., Boreiry, M.: A secure method in digital video watermarking with transform domain algorithms. *Multimedia Tools Appli.* **80**(13), 20449–20476 (2021). <https://doi.org/10.1007/s11042-021-10730-5>
23. Mareen, H., Praeter, J., Wallendael, G., Lambert, P.: A scalable architecture for uncompressed-domain watermarked videos. *IEEE Trans. Inf. Forensics Secur.* **14**(6), 1432–1444 (2018)
24. Chen, L., Zhao, J.: Informed histogram-based watermarking. *Multimedia Tools Appli.* **77**(6), 7187–7204 (2018)