AIMS Mathematics, 7(5): 8390–8407.
DOI: 10.3934/math.2022468
Received: 04 November 2021
Revised: 26 January 2022
Accepted: 06 February 2022
Published: 28 February 2022

http://www.aimspress.com/journal/Math

Research article

Zero-watermarking Algorithm for Audio and Video Matching Verification

Wenxue Sun[1], Huiyuan Zhao[1], Xiao Zhang[1], Yuchao Sun[1], Xiaoxin Liu[1], Xueling Lv[2] and Di Fan[1,*]

[1] Shandong University of Science and Technology, Qingdao, 266590, Shandong, China
[2] Department of Management and Economics, Tianjin University, Tianjin, 300072, China

* **Correspondence:** Email: fandi_93@126.com.

Abstract: For the needs of tamper-proof detection and copyright identification of audio and video matching, this paper proposes a zero-watermark algorithm that can be used for audio and video matching verification. The algorithm segments audio and video in smaller time units, generates a video frame feature matrix based on NSCT, DCT, and SVD, and generates a sound watermark based on methods such as DWT and K-means. The zero watermark combines video, audio and copyright information. The experimental results show that the zero watermark generated by this algorithm can not only realize highly accurate matching detection and positioning of audio and video, but also well resist common single attack and combination attacks such as noise, scaling, rotation, frame attack and format conversion, which has good robustness.

Keywords: audio and video matching verification; zero watermark; copyright recognition; robustness
Mathematics Subject Classification: 00A69

1. Introduction

With the development of multimedia technology and AI technology, videos can be easily accessed, copied, tampered and disseminated, and some may cause interest infringement and personal attacks on copyright holders, and some may cause political disturbances [1,2]. The authenticity of videos on the Internet is often questioned, and the crisis of trust spreads in the society [3–5]. Although digital watermarking technology can confirm and protect the copyright of audio or video, it can not be used to detect whether audio and video match or not, and can not indicate the mismatched part of media where audio or video tamper occurs. Therefore, the research on audio and video matching and tamper-proof is of great significance in the verification of information authenticity.

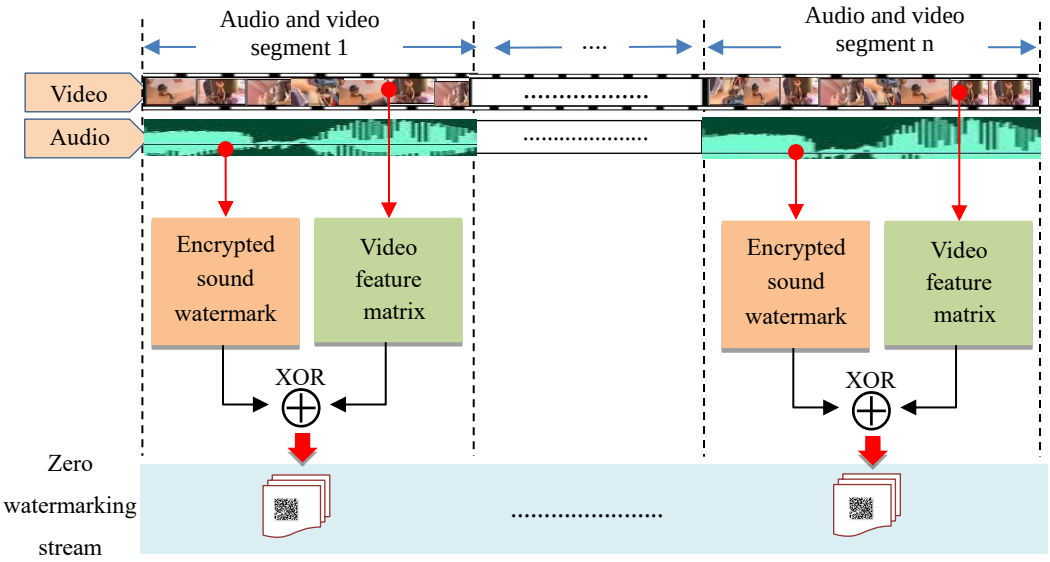
At present, there are relatively few watermarking technologies whose main purpose is to detect the matching of audio and video. Most digital watermarking algorithms are aimed at a single type of carrier such as image, audio, and video. The image digital watermarking algorithm can be divided into three categories: spatial domain methods [6,7], transform domain methods [8–14] and deep learning-based method [15–17]. The spatial image watermarking algorithm, such as modifying the lowest pixel level according to watermark information [6], has low computational complexity but weak anti-attack ability. The transform domain method uses DCT (Discrete Cosine Transform, DCT) [8,10], DWT (Discrete Wavelet Transform, DWT) [9], Contourlet [11] and other transformations to transform the image into a scope for watermark addition, which has better concealment and robustness. The deep learning-based methods are produced with the rise of deep learning [17], which is a new direction of watermarking technology. This method uses the deep learning-based model to embed and extract watermarks through learning, and its anti-attack performance is good, but it still needs further research in terms of watermark capacity and algorithm complexity. Similarly, audio watermarking algorithms can be divided into time domain and transform domain algorithms, such as least significant bit algorithm [18], echo hiding algorithm and phase coding algorithm [19], audio watermarking technology of DWT-DCT [20], audio watermarking technology of SVD (Singular Value Decomposition, SVD) decomposition and fractional Fourier transform [21]. Audio watermarking technology based on DWT and Schur decomposition [22]. Image watermark algorithm can be used for video [23–25], in addition, the video watermark also has a compression domain method, which is a watermark technology combining specific video coding methods, such as MPEG [26], H.264 [27,28], H.265 [29] video watermark algorithm.

Audio and video cross watermarking can detect the matching of the entire audio and video [30–34]. The earliest cross-watermarking algorithm was proposed by Jana Dittmann in 1999. The algorithm embeds a sequence that decreases by 1 into the audio carrier, and at the same time embeds a sequence that increases by 1 into the video carrier, so that the sum of the embedded sequences in the audio and video information is always equal to 0, verify the synchronization between audio and video by adding all the watermark data to see if the sum is 0. This algorithm has low computational complexity and is easy to implement, but the robustness of watermarking is poor [30]. The method proposed by Agrawal is to use the lens segmentation method to divide the input video into several non-overlapping lenses, generate a sound watermark based on the audio bit plane, and embed it in the blue component wavelet low-frequency subband of each video frame in the lens, while applying the particle group and genetic algorithm to optimize the embedding position [31]. The video watermark embedding scheme based on the optimal position proposed by Sundararajan is also based on the audio bit plane to generate the audio watermark. The difference from the literature [31] is that the cuckoo algorithm is used for optimal position analysis and the audio watermark data is embedded [32]. The audio and video cross-watermarking algorithm proposed by Wang Xinyuan combined with the visual saliency model is to decode the video key frame and the audio frame at the same time, use the low-pass amplitude characteristics of the audio to generate the watermark, and analyze the saliency component of the video key frame. Later, the watermark is embedded in the DC coefficients of the DCT transform of the non-salient area through quantization index modulation [33]. The compressed domain watermark proposed by Esmaeilbeig uses the hash bits generated by the hash algorithm in the audio part as a watermark, which is embedded in the quantized residual DCT coefficients of the video I frame [34]. Although the above audio and video cross-watermarking can protect audio or video copyright, it cannot locate the tampering of local small fragments in the media stream.

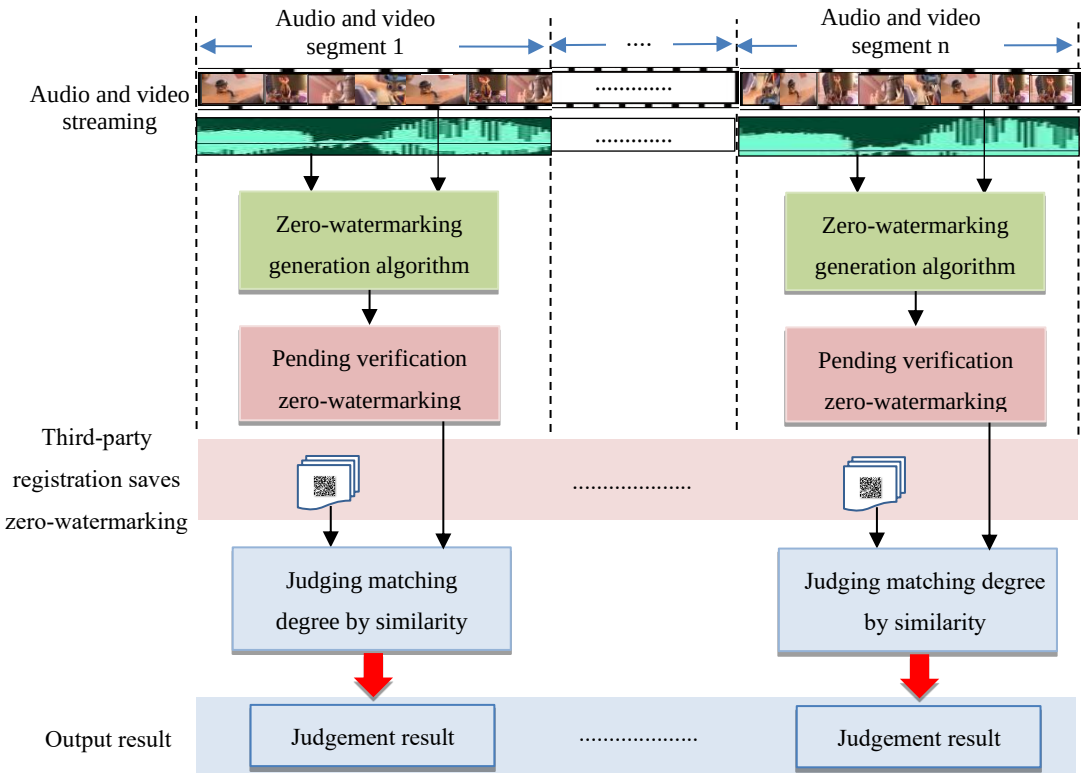
This paper presents a zero-watermarking algorithm which can be used for audio and video matching and local location. The algorithm generates a zero-watermarking stream based on audio and video segments, which can not only detect audio and video matching tamper but also realize tamper location. A variety of attacks and tampering experiments show that the proposed algorithm can detect and locate audio and video tampering, and the algorithm has high robustness.

2. Audio and video matching detection framework based on zero-watermarking

In this paper, the method of detecting audio and video matching is based on the generated zero watermark, and its framework is shown in Figure 1. The algorithm firstly segmented audio and video, and each segment generated a zero-watermarking integrated with audio and video features. The whole audio and video formed a zero-watermark stream, as shown in Figure 1(a). The matching detection process is shown in Figure 1(b). The audio and video segments are formed with the same method as Figure 1(a) to form a zero-watermarking, and compared with the zero- watermarking of the copyright center. If the correlation coefficient between the two is greater than the threshold, then the audio segment and the video segment match; if it is less than the threshold, the audio segment and the video segment do not match, and one of them is tampered with. Because the zero-watermarking comparison is performed segment by segment, it can realize the location of tampering, and the accuracy of the location is the segment length. In addition to detecting the matching of positioning audio and video, the method in this paper can also be used for traditional copyright determination.



(a) Zero-watermarking stream generation process integrating audio and video features.



(b) Audio and video matching detection and positioning process.

Figure 1. Audio and video matching detection framework based on zero watermarking.

3. Matching zero-watermarking generation

The zero-watermarking generation algorithm for a certain segment of audio and video in Figure 1(a) is shown in Figure 2. Extract key frames from video segments, and generate feature matrix of key frame images based on NSCT (Nonsubsampled Contourlet Transform, NSCT), DCT, SVD; for audio segments, incorporate copyright watermarks to generate sound watermarks. The feature matrix and the sound watermark are XOR to form the zero watermark of this paragraph. The zero-watermarking not only contains copyright information, but also incorporates the audio and video features of this paragraph, which can play a role in traditional copyright protection, as well as match detection and positioning. Among the four modules of the algorithm, key frame extraction, video feature matrix generation, encrypted sound watermarking generation and zero-watermarking generation, the key frame extraction is at the front end, which is used to reduce the redundancy of video, simplify the zero-watermarking process and reduce the number of zero- watermarking. Video feature matrix generation is to extract representative stable features from frame images to increase the robustness of watermarking. The generated part of encrypted sound watermark not only contains the sound information, but also hides the information, which increases the security of watermark on the one hand, and provides a basis for the realization of matching detection on the other hand. The zero-watermarking generation part combines image features and audio features to ensure the realization of the final algorithm function and excellent performance.

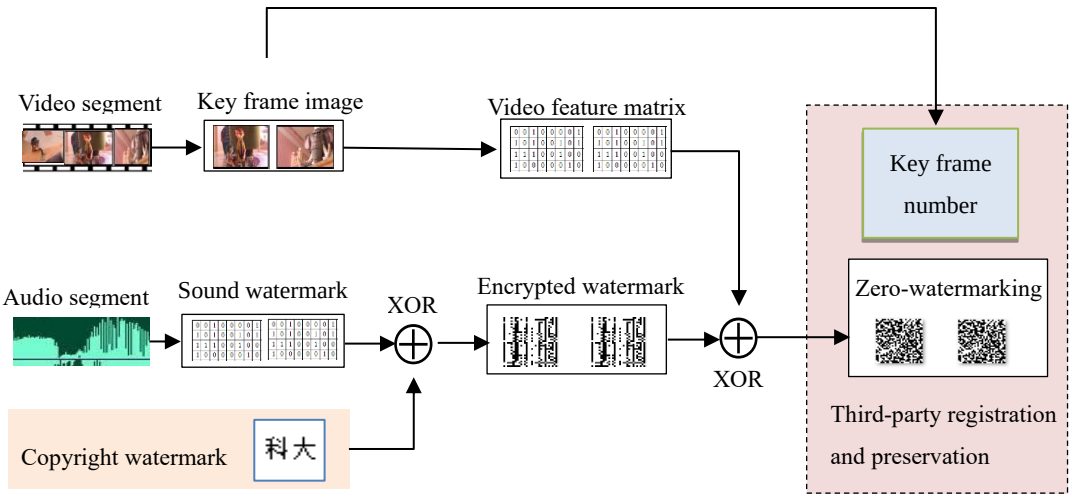


Figure 2. Zero-watermarking generation algorithm for audio and video segments.

3.1. Key frame feature matrix generation

There is a lot of redundancy between video frames. Selecting key frames can reduce redundancy and the number of zero watermarks. This paper uses the Euclidean distance of the frame difference to select the key frame, and saves the frame number of the key frame as the secret key. Based on the key frame, the feature matrix of the video segment is constructed by using NSCT, DCT, SVD and other methods. The features contained in the matrix have good stability and robustness, and combined with Zernike invariant moments, it increases the ability of extracting features to resist rotation attacks ability. The detailed steps are as follows:

1. Calculate the Zernike moment of the key frame image and save it for rotation correction.

2. After normalizing the size of the key frame image, it is transferred from the RGB space to the YCoCg space, and the three components of Y, Co and Cg are decomposed. It is found that the loss of Co component in image compression is less than that of Y component and Cg component. In order to improve the anti-compression attack ability and overall robustness of watermark, this paper proposes to embed watermark in Co component of image.

3. Perform two-layer NSCT on the Co component, perform DCT transformation on its low-frequency subband, and divide the result into 32*32 non-overlapping sub-blocks. Each non-overlapping sub-block is marked as $T_{i,j}\,(i, j = 1 \cdots 32)$.

4. Perform SVD decomposition on each non-overlapping sub-block $T_{i,j}$ to obtain a diagonal matrix, as shown in formula (1).

$$[U_{i,j}, S_{i,j}, V_{i,j}] = SVD(T_{i,j}) \tag{1}$$

Among them, $U_{i,j}$ and $V_{i,j}$ are orthogonal matrices, and $S_{i,j}$ is a diagonal matrix.

5. Extract the maximum value of the diagonal elements in the diagonal matrix $S_{i,j}$, record it as $S_{i,j}$, and construct a 32*32 transition matrix $t(i, j) = S_{\max_{i,j}}\,(i, j = 1 \cdots 32)$.

6. Using the mean value M of the elements in the transition matrix $t(i, j)$ as the threshold, the transition matrix $t(i, j)$ is binarized to form the video feature matrix F,

$$F(i,j)=\begin{cases}1, t(i,j)>M\\ 0, t(i,j)\le M\end{cases}, i,j=1,2\cdots 32 \tag{2}$$

3.2. Copyrighted sound watermark generation

For audio segments, DWT and K-means algorithms are used to extract sound features, and at the same time, copyright watermarks containing copyright information are incorporated into sound features to obtain copyrighted sound watermarks. The detailed steps are as follows:

Step 1: Perform two-level DWT transformation on the audio segment to obtain its low-frequency wavelet coefficient LL.

Step 2: For section LL, calculate the first moment μ and the second moment σ of each wavelet coefficient.

Step 3: Perform K-means coding with μ and σ as features to obtain a one-dimensional two-valued feature matrix of sound, and obtain a two-dimensional voice feature matrix V by increasing the dimension.

Step 4: The copyright watermark B is incorporated into the sound feature matrix V to obtain the copyrighted sound watermark W.

Step 5: Use the Logistics chaotic encryption algorithm to encrypt W, and get the encrypted sound watermark w with copyright.

3.3. Zernike moments of key frame images

The Zernike Moment is the radial moment and is based on the orthogonal function of Zernike polynomials, and ZM (Zernike Moment, ZM) whose amplitude remains constant for only phase change during image rotation is widely applied to image rotation, feature extraction, and excellent [35].

Assuming that the polar coordinate of the image is expressed as $f(\rho,\theta)$, the Zernike moment in the N-order M heavy polar coordinate system is:

$$A_{nm}=\frac{n+1}{\pi}\int_0^{2\pi}\int_0^1 \mathrm{f}(\rho,\theta)V_{nm}^*(\rho,\theta)\rho d\rho d\theta \tag{3}$$

Among them, n is a non-negative integer, m is an integer that satisfies $|m|\le n$, and $|m|-n$ is an even number; V_{nm} is an m-fold Zernike moment polynomial of order n, and V_{nm}^* is the conjugate of V_{nm}. V_{nm} can be expressed as:

$$V_{nm}(\rho,\theta)=R_{nm}(\rho)e^{jm\theta} \tag{4}$$

$$R_{nm}(\rho)=\sum_{s=0}^{[n-|m|]/2}\frac{(-1)^s[(n-s)!]\rho^{n-2s}}{s!([n+|m|]/2-s)!([n-|m|]/2-s)!} \tag{5}$$

The Zernike moment of an image can be expressed as:

$$A_{nm}=|A_{nm}|\arg(A_{nm}) \tag{6}$$

Among them, arg() represents the calculation to find the argument.

Assuming that A_{nm} is the Zernike moment before the image is rotated, and A'_{nm} is the Zernike moment after the image is rotated, the rotation angle A'_{nm} is solved as:

$$\alpha = [\arg(A'_{nm}) - \arg(A_{nm})]/m, \ m \neq 0 \tag{7}$$

The Zernike moments of the key frame images of the algorithm are saved and used for rotation correction in the detection and identification stage to improve the anti-rotation performance of the algorithm. Figure 3 shows the Normalized Cross-Correlation of the watermark extracted without the Zernike rotation correction algorithm and the watermark NC (Normalized Cross-Correlation, NC) value extracted by the algorithm after adding the Zernike moment rotation correction. It can be seen that the NC value of the watermark obtained after the Zernike moment rotation correction is higher than the original watermark NC value, which effectively improves the anti-rotation attack ability of the algorithm, and the NC value increases more obviously under a large rotation attack.

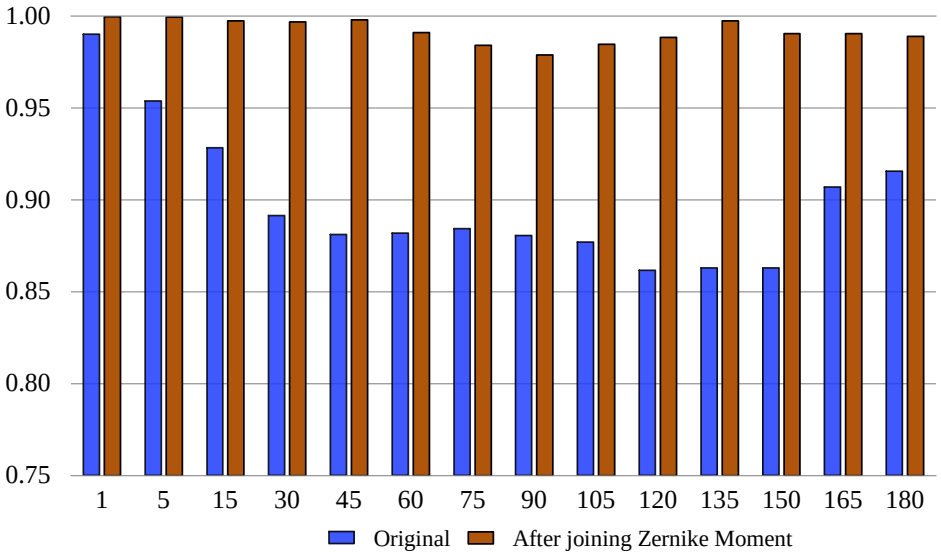


Figure 3. Comparison of experimental results of rotation correction.

4. Matching detection process based on zero-watermarking

The zero-watermarking detection algorithm is shown in Figure 4, which is the inverse process of the audio and video matching zero-watermarking generation algorithm. When performing audio and video matching detection, the audio and video stream to be verified is decoded and segmented, and a zero-watermarking is generated for each segment with the support of the third-party stored information. Calculate the similarity between the zero-watermarking of the audio and video to be verified and the zero-watermarking extracted from the third party. If it is greater than the threshold, the audio segment and video segment are matched; if it is less than the threshold, the audio segment and video segment do not match, and one of them is replaced or tampered with.

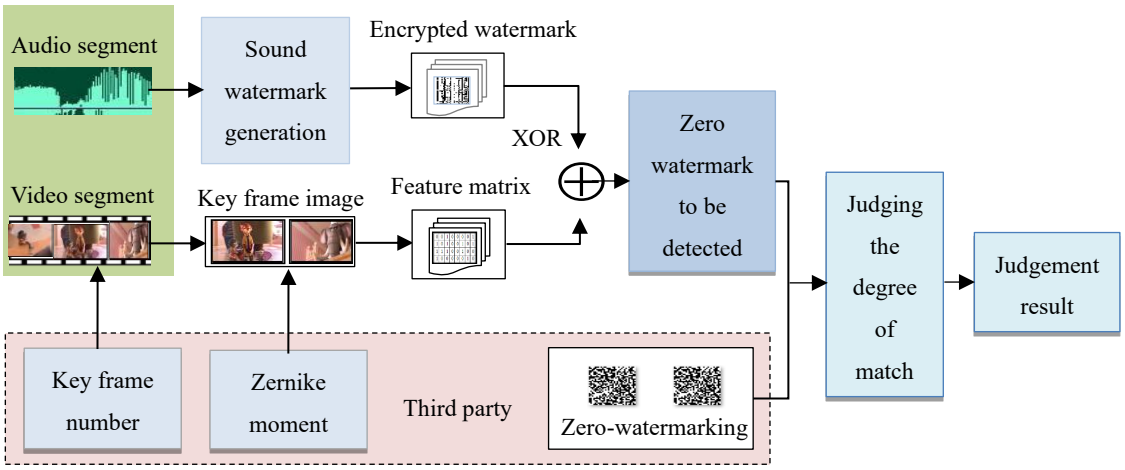


Figure 4. Flow chart of zero-watermark detection algorithm.

5. Experimental results and analysis

The experiment in this article is carried out on the Matlab R2014b platform. The copyright watermark used is the words "Technology University" and its size is 32*32. In the experiment, the parameter of Logistic chaotic encryption is $x_0 = 0.1, u = 4$, audio and video segments are segmented in 1s. For the question of audio and video segment length, from the perspectives of the stability of audio and video features, the rapidity of generating zero-watermarking, the minimization of occupied resources, and the accuracy of matching detection, through comprehensive analysis and experiments, we determined to segment audio and video in 1s as the time unit. In this way, on the one hand, stable features of audio and video segments can be effectively extracted to quickly construct optimized zero-watermarking, and on the other hand, tampering of small audio or video segments in the whole audio and video stream can also be accurately detected.

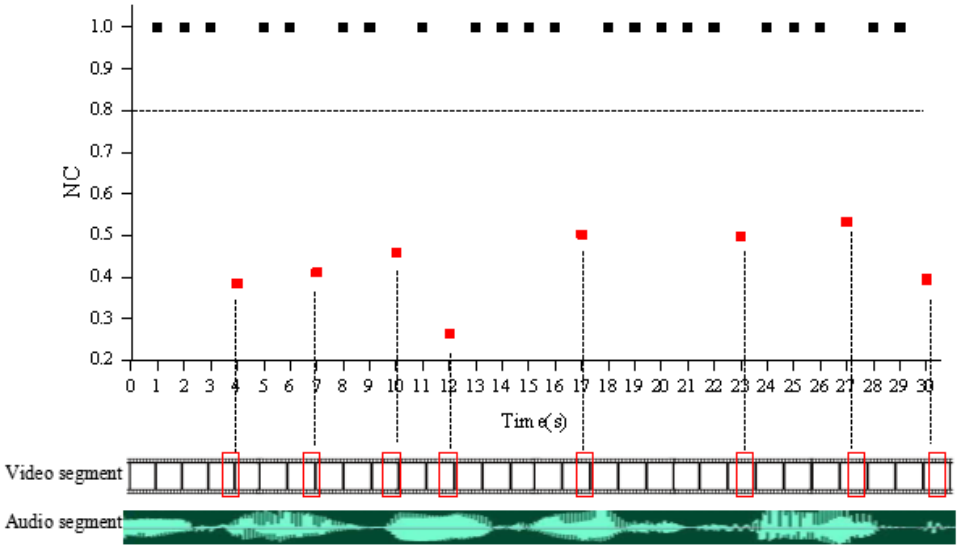
The video used in the following experiment is video in H.264 encoding format (including audio), the video frame size is 1080*1920, the duration is 30 seconds, the frame rate is 25fps, the audio stream sampling rate is 32KHZ, 16-bit quantization bit, dual sound road. According to the algorithm of this paper, it is divided into 30 audio and video segments, and 49 key frames are extracted in total. For the audio and video experiment in this paper, the algorithm takes less than 180s to generate the zero-watermarking stream, and less than 230s to detect the zero-watermarking matching.

In the experiment, the NC value is used as the objective evaluation criterion of the robustness of the watermark, and the Peak Signal-to-Noise Ratio is used as the difference measurement index of the two images. The larger the NC value, the better the extracted watermark effect and the stronger the algorithm robustness. If the PSNR (Peak Signal-to-Noise Ratio, PSNR) value is smaller, it means that the attack intensity is greater and the damage degree is stronger. The audio and video matching is determined by the normalized NC value, and the threshold is set to 0.8, that is, when the NC is greater than or equal to 0.8, the audio and video are matched; when the NC is less than 0.8, the audio or video is tampered with.

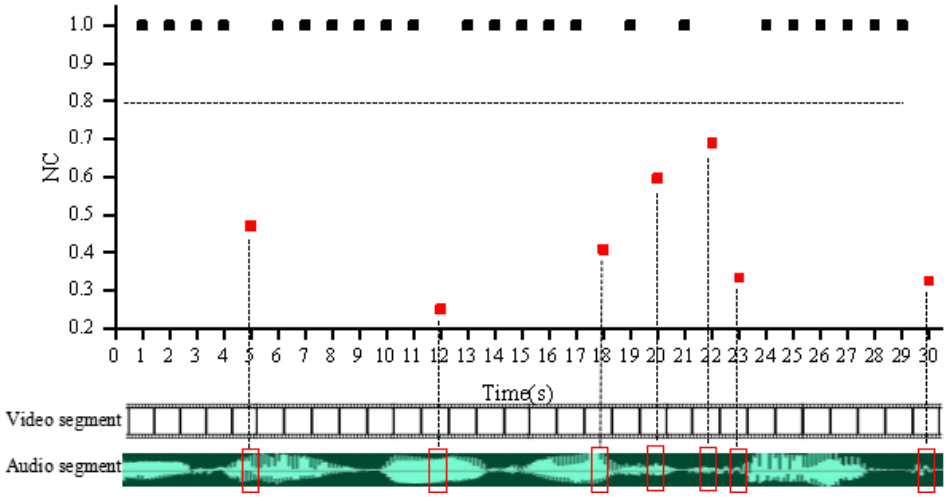
5.1. Audio and video matching and anti-tampering test

One of the functions of zero-watermarking in this paper is to detect and locate the matching of

audio and video. Based on the above video, we randomly extract video frames and audio in different periods for replacement, and then use this method to detect and locate the matching of audio and video respectively. The results are shown in Figure 5. It can be seen from the figure that for the tampered video segments 4, 7, 10, 12, 17, 23, 27 and 30, the detected zero-watermarking NC values are less than the set threshold, which is determined to be mismatched; for the tampered 5, 12, 18, 20, 22, 23 and 30 audio segments, the detected zero-watermarking NC value is also less than the set threshold, and the detection result is mismatch; the results of both tests are correct. Therefore, the main function of the algorithm in this paper, detection of audio and video matching and tamper-proof, is realized through the cooperation of each module of the system and the generated zero-watermarking, and the algorithm accurately gives the conclusion of "mismatch" no matter the tamper of small fragments in video or audio.



(a) Test results of matching detection after video tampering.



(b) Test results of matching detection after audio tampering.

Figure 5. Test results of matching detection and positioning of audio and video.

5.2. Algorithm security test

In this paper, the watermark image is scrambled by Logistic [36] chaos to hide the watermark information. For the watermark extraction party, only know the zero-watermarking algorithm and encryption methods and parameters, can correctly extract the watermark information, so as to increase the security of the algorithm and watermark.

5.3. Algorithm robustness test

The following experiments respectively test the robustness of the watermarking algorithm under single attack and combined attack. Single attacks include Gaussian noise, salt and pepper noise, cropping, scaling, rotation, frame averaging, frame reorganization, format conversion, etc. Combination attacks are situations where more than single attacks work simultaneously. The robustness of the algorithm is judged by the PSNR and NC values. The experimental results of a single attack are shown in Table 1. The attack is performed on each frame of the video. The PSNR mean and PSNR variance are the mean and variance of the PSNR after all frames are attacked. The NC mean and NC variance are the mean and variance of the NC after all watermarking attacks. From the results in the table, the algorithm in this paper can still effectively extract the watermark after the attack. Even if the PSNR of the image is about 20dB, the NC value under most attacks is above 0.98, and the variance is also very low, indicating that the algorithm is robust sex. Experimental data show that, with the cooperation of all modules, the algorithm has certain advantages in robustness and anti-attack ability, especially for noise, filtering and compression attacks, and for rotation and clipping attacks, the NC value is above 0.96, which has good anti-attack ability.

Table 1. Robustness experimental results under a single attack.

Attack type	Attack parameter	PSNR mean	PSNR variance	NC mean	NC variance
Gaussian noise	0.01	20.8070	2.18×10^{-3}	0.9929	3.81×10^{-5}
	0.05	14.3482	1.70×10^{-2}	0.9805	2.26×10^{-4}
Salt and pepper noise	0.01	24.4024	3.62×10^{-3}	0.9952	3.33×10^{-5}
	0.05	17.4239	8.23×10^{-3}	0.9880	1.05×10^{-4}
Gaussian filtering	7*7 sigma=1	31.9114	0.5397	0.9954	1.69×10^{-5}
	7*7 sigma=5	26.1747	0.2591	0.9879	8.09×10^{-5}
Median filtering	3*3	32.0116	0.5945	0.9964	1.17×10^{-5}
JPEG compression	50	34.8562	0.3218	0.9844	6.40×10^{-5}
	20	31.1777	0.2277	0.9649	2.69×10^{-4}
Rotate	5°	21.0921	0.1819	0.9915	2.06×10^{-5}
	15°	18.2187	0.1990	0.9771	9.33×10^{-5}
	30°	15.8933	0.4587	0.9556	3.54×10^{-4}

Continued on next page

Attack type	Attack parameter	PSNR mean	PSNR variance	NC mean	NC variance
Scaling	1/2	33.9616	1.1238	0.9983	2.62×10^{-6}
	2	45.9223	2.1806	0.9995	4.66×10^{-7}
Cutting	Lower right 1/16	14.7301	0.2582	0.9888	1.86×10^{-4}
	Lower right 1/8	11.2353	0.2030	0.9679	3.12×10^{-4}
Frame attack	Frame average	23.0641	9.7437	0.9906	8.91×10^{-5}
	Frame reorganization	33.4309	14.2672	0.9949	4.28×10^{-5}
Format conversion	H.264	33.0963	14.4316	0.9859	9.09×10^{-5}
	MPEG4	33.6141	8.5036	0.9872	4.28×10^{-5}

We also changed the attack parameters in a relatively large range for several kinds of attacks and conducted repeated experiments to grasp the characteristics of the algorithm more comprehensively. The experimental results are as follows.

1) Noise attack. Gaussian noise and salt and pepper noise of different intensities are selected for attack test, and the noise intensity is increased by 0.01 in the range of 0–0.1, and the result is shown in Figure 6. Figure 6 shows the NC value of each extracted watermark in the form of a scatter diagram, in which the horizontal line represents its mean position and the mean and variance of the NC under two kinds of noise intensities. It can be seen that with the increase of noise intensity, the fluctuation range of NC value becomes larger, but the mean NC value of watermarking is above 0.97, indicating that the algorithm has strong anti-noise ability, especially under the attack of salt-and-pepper noise. The mean NC value of watermarking is above 0.98, indicating that the algorithm has stronger robustness to salt-and-pepper noise.

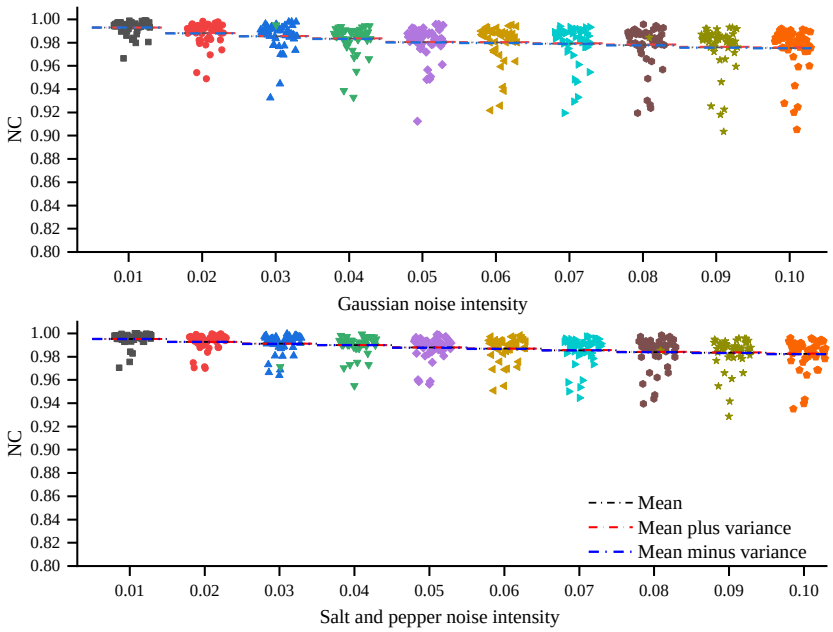


Figure 6. NC value, mean and variance under different noise attacks.

2) JPEG compression attack. The quality factor reflects the quality of the image after JPEG compression. If the value is larger, it means that the image quality is higher and the image suffers less attack. This paper has experimented with the robustness of the algorithm when the quality factor is increased by 10 in the range of 10–90. The result is shown in Figure 7. It can be seen that with the increase of the quality factor, the distribution of the NC values of the watermark extracted from multiple key frames becomes concentrated, and the fluctuation range of the NC values gradually decreases, indicating that the algorithm can resist JPEG compression attacks very well.

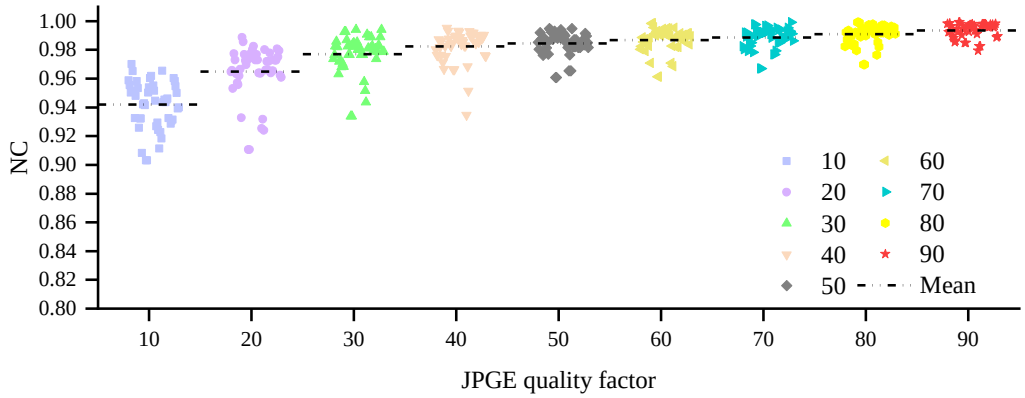


Figure 7. NC values under JPEG compression attacks with different quality factors.

3) Scaling attack. In this paper, the video is scaled by 1/8, 1/4, 1/2, 2, 4, and 6 times respectively. The results are shown in Figure 8. Experimental results show that the algorithm in this paper has strong robustness to scaling attacks. The extracted watermark NC value is about 0.98, and the NC value under the amplification attack is close to 1.

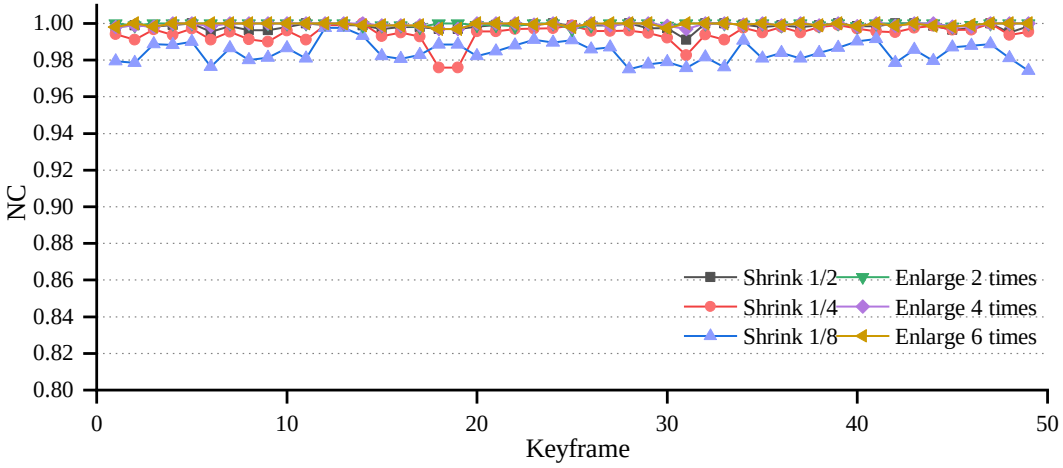


Figure 8. NC value under scaling attack.

4) Rotating attack. In this paper, the rotation attack experiment is carried out in the range of 0–180°with an interval of 15. The results are shown in Figure 9. It can be seen from Figure 9 that even under a large rotation attack, the average value of the extracted watermark NC is above 0.9.

5) Filter attacks. In this paper, Gaussian filters with different window sizes and different scales are selected for attack experiments, and the results are shown in Figure 9. It can be seen from Figure 9 that as the filter window size and surround scale increase, the NC of the watermark decreases, but the average value of the watermark is still around 0.96. This algorithm is robust to Gaussian filtering.
6) Cutting attack. In this paper, four corners cropping 1/20, 1/16, 1/8, and center cropping 1/16 attack experiments were performed on the video, and the results are shown in Figure 9. The experimental results show that the algorithm in this paper extracts the features of key frames when generating watermarks. Large-scale cropping attacks cause a large number of key frame features to be lost, which seriously affects the NC value of the watermark. Although the algorithm in this paper is slightly less robust against clipping attacks, the average value of NC is above the matching detection threshold, and there is no problem in detecting matching.

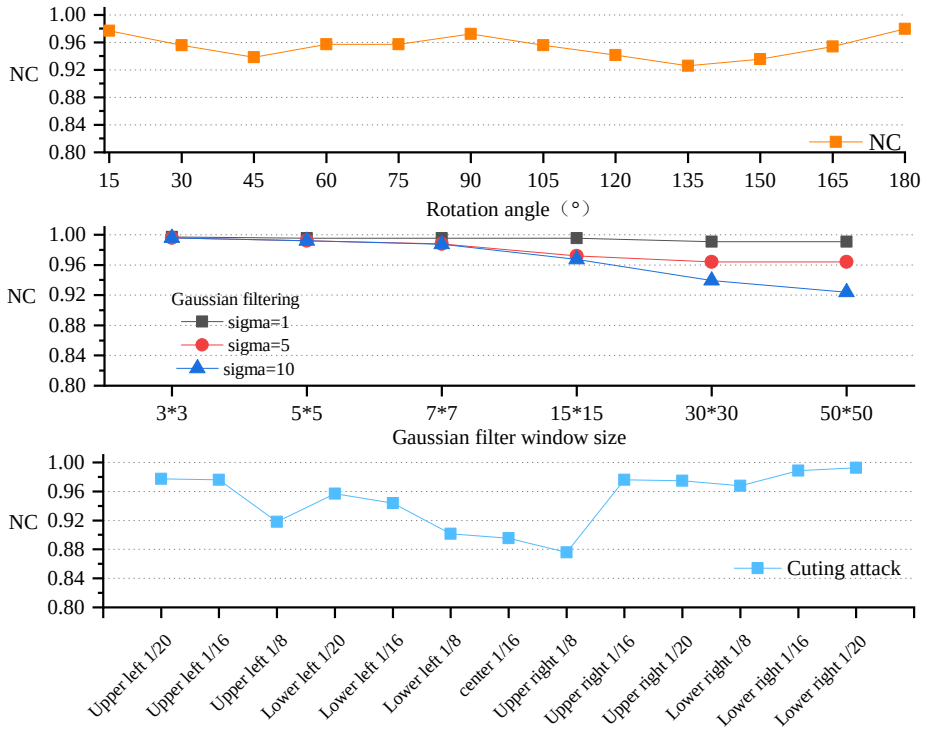
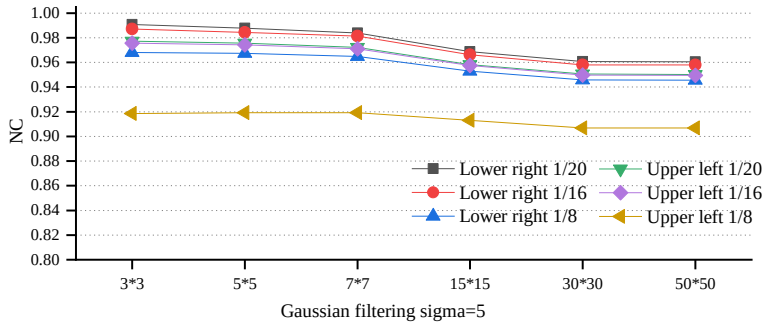


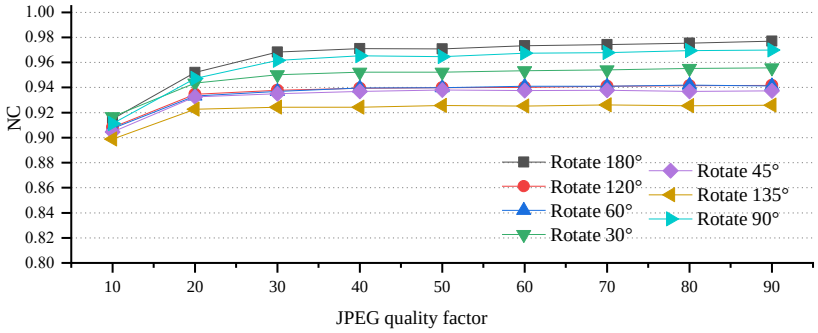
Figure 9. NC mean value under rotation, Gaussian filtering, and citing attacks.

7) Combination attack. In this paper, combination attack experiments are conducted on the algorithm, mainly including JPEG compression + cropping, rotation + Gaussian filtering, format conversion + other three combination attacks. The results are shown in Figure 10.

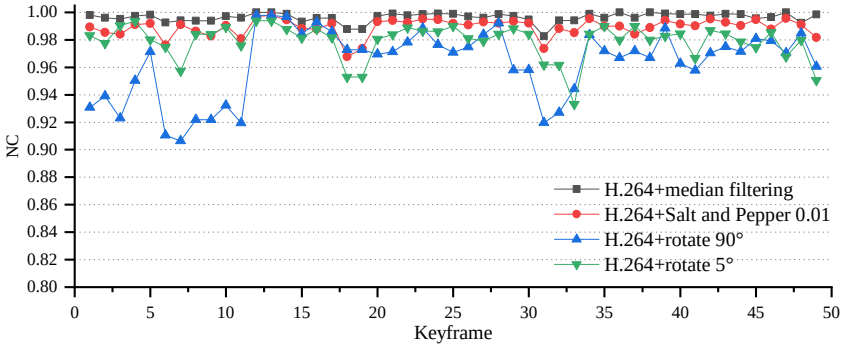
From the experimental results in the figure, it can be seen that the algorithm in this paper also has good robustness to combined attacks, with an average NC value of 0.9 or more; in contrast, the algorithm is more robust to Gaussian filtering attacks than cropping attacks, and it is more robust to JPEG compression. The attack is more robust than rotating attacks. In the combination of format conversion and other attacks, the results show that the sensitivity of different video frames to different attacks is different, so the anti-attack ability is related to the frame, but overall, the NC value is still relatively high, which can be used for matching detection.



(a) NC mean value under cropping + Gaussian filtering attack.



(b) NC mean value under rotation + JPEG compression attack.



(c) NC mean value under format conversion and other attacks.

Figure 10. The mean value of NC under combined attack.

5.4. Comparison and analysis with other algorithms

This paper also experimented with the comparison between the algorithm in this paper and the algorithm in literature [24] and literature [37]. Among them, the literature [24] proposed a robust video watermarking algorithm combining discrete cosine transform and discrete wavelet transform technology. Literature [37] is a video watermarking algorithm based on the Contourlet domain, using

SIFT (Scale Invariant Feature Transform, SIFT) feature extraction for correction. The main difference between the algorithms in this paper and them is the using of sound watermark and zero watermark. Literature [38] proposed a robust non-blind video watermarking technology based on DWT and QR decomposition. Literature [39] proposed a discrete wavelet based on redundancy.

In this part of the experiment, the classic test videos foreman_cif and bus_cif are selected. The resolution is 352*288, the frame rate is 29fps, the duration of bus_cif is 5 seconds, and the duration of foreman_cif is 10 seconds. The watermark used in literature [24,37–39] is a 32*32 binary image, and the watermark used in this algorithm is a 32*32 binary watermark based on audio and copyright images. The comparative experimental results are shown in Table 2. It can be seen that the algorithm in this paper has good robustness under rotation, scaling, and cutting attacks, especially in terms of rotation and cutting.

Table 2. Comparison of experimental results between the algorithm in this paper and the algorithm in the literature (NC mean).

Experiment video	Attack type	Literature [24] algorithm	Literature [37] algorithm	Literature [38] algorithm	Literature [39] algorithm	Proposed algorithm
foreman	Rotate (10°)	0.8209	0.8226	0.4890	0.9234	0.9416
	Rotate (30°)	0.8096	0.8591	0.4888	0.9196	0.9036
	Rotate (45°)	0.7992	0.8330	0.4810	0.8913	0.8949
	Scaling (1/2)	0.9290	0.9757	0.8663	0.9824	0.9984
	Scaling (2)	0.9041	0.6348	0.8852	0.9790	0.9988
	Rotate (10°) + Scaling (2)	0.8042	0.8591	0.4819	0.9201	0.9403
	Rotate(30°）+Scaling(1/2)	0.7924	0.8435	0.4805	0.8708	0.8691
	Cutting (1/8)	0.5292	0.9078	0.8831	0.9127	0.9204
	Cutting (1/4)	0.3936	0.8070	0.8748	0.8296	0.8428
	Cutting (1/2)	0.3235	0.6000	0.7012	0.6923	0.7093
	Median filtering	0.9295	0.9965	0.8645	0.9854	0.9956
bus	Rotate (10°)	0.8562	0.8887	0.4998	0.9465	0.9624
	Rotate (30°)	0.8208	0.7861	0.4870	0.9238	0.9258
	Rotate (45°)	0.7961	0.7078	0.4826	0.8820	0.8899
	Scaling (1/2)	0.9473	0.9843	0.9412	0.9884	0.9954
	Scaling (2)	0.9138	1	0.9560	0.9863	0.9984
	Rotate (10°) + Scaling (2)	0.8279	0.8904	0.4912	0.9490	0.9620
	Rotate(30°）+Scaling(1/2)	0.7947	0.7809	0.4863	0.9202	0.9249
	Cutting (1/8)	0.5945	0.9061	0.8692	0.9389	0.9449
	Cutting (1/4)	0.4789	0.8157	0.8590	0.8450	0.8626
	Cutting (1/2)	0.4560	0.5965	0.6812	0.7239	0.7467
	Median filtering	0.9469	0.9826	0.8982	0.9856	0.9934

6. Conclusions

In this paper, aiming at the anti-tamper detection of audio and video matching and the invisibility and robustness of watermarking, a zero-watermarking algorithm that can be used for audio and video matching verification and fine positioning is proposed. The algorithm generates a zero-watermarking

stream fused with audio and video features in units of audio and video segments. The zero-watermarking carries audio and video information at the same time. It can not only be used for traditional copyright determination, but also for audio and video tampering detection and positioning. Whether the video or audio segment has been tampered with, it can be detected by the zero-watermarking, which overcomes the watermark formed in an overall way can only be used for copyright identification and cannot accurately detect and locate the problem of small audio or video tampering. In this paper, a variety of attacks and tampering experiments are carried out on the proposed algorithm. The experimental results show that the proposed algorithm can detect and locate tampering of audio and video segments with high precision, and it also has high robustness and can resist most common single attack and combination attack.

Acknowledgement

The research was supported by the following projects: Scientific research project of National Language Commission (YB135-125); Key Research and Development Project of Shandong Province (2019GGX101008, 2016GGX105013).

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. H. Agarwal, F. Husain, Development of payload capacity enhanced robust video watermarking scheme based on symmetry of circle using lifting wavelet transform and SURF, *J. Inf. Secur. Appl.*, **59** (2021), 102846. https://doi.org/10.1016/j.jisa.2021.102846
2. S. Xuecheng, L. Zheming, W. Zhe, L. Yongliang, A geometrically robust multi-bit video watermarking algorithm based on 2-D DFT, *Multimed. Tools Appl.*, **80** (2021), 13491–13511. https://doi.org/10.1007/s11042-020-10392-9
3. G. Sandaruwan, L. Ranathunga, Robust and adaptive watermarking technique for digital images. IEEE International Conference on Industrial & Information Systems. IEEE, 2017, 1–6. https://doi.org/10.1109/ICIINFS.2017.8300387
4. M. Reza Keyvanpour, N. Khanbani, M. Boreiry, A secure method in digital video watermarking with transform domain algorithms, *Multimed. Tools Appl.*, **80** (2021), 20449–20476. https://doi.org/10.1007/s11042-021-10730-5
5. D. Ariatmanto, F. Ernawan, An improved robust image watermarking by using different embedding strengths, *Multimed. Tools Appl.*, **79** (2020), 12041–12067. https://doi.org/10.1007/s11042-019-08338-x
6. J. Abraham, V. Paul, An imperceptible spatial domain color image watermarking scheme, *J. King Saud Univ.-Comput. Inf. Sci.*, **31** (2019), 125–133. https://doi.org/10.1016/j.jksuci.2016.12.004
7. W. Zhou, G. Jiang, M. Yu, F. Shao, Z. Peng, Reduced-reference stereoscopic image quality assessment based on view and disparity zero-watermarks, *Signal Process. Image Commun.*, **29** (2014), 167–176. https://doi.org/10.1016/j.image.2013.10.005

8. D. Fan, Y. Li, S. Gao, A novel zero watermark optimization algorithm based on Gabor transform and discrete cosine transform, *Concurr. Comput. Pract. Exp.*, **2** (2020). https://doi.org/10.1002/cpe.5689
9. C. Kumar, A. K. Singh, P. Kumar, et al, SPHIT-based multiple image watermarking in NSCT domain, *Concurr. Comput. Pract. Exp.*, **32** (2020), e4912.1-e4912.9. https://doi.org/10.1002/cpe.4912
10. M. Moosazadeh, G. Ekbatanifard, An improved robust image watermarking method using DCT and YCoCg-R color space, *Optik*, **140** (2017), 975–988. https://doi.org/10.1016/j.ijleo.2017.05.011
11. C. Zhu, Y. Li, W. Chi, S. Gao, D. Fan, Zero-watermarking algorithm for color image in contourlet domain based on Schur decomposition, *Inf. Technol. Inf. Technol.*, **227** (2019), 94–98.
12. D. Wei, Y. Li, Convolution and Multichannel Sampling for the Offset Linear Canonical Transform and Their Applications, IEEE T. Signal PR., **67** (2019), 6009–6024. https://doi.org/10.1109/TSP.2019.2951191
13. D. Wei, M. Jiang, A fast image encryption algorithm based on parallel compressive sensing and DNA sequence, *Optik*, **238** (2021), 166748. https://doi.org/10.1016/j.ijleo.2021.166748
14. Y. Li, D. Wei, L. Zhang, Double-encrypted watermarking algorithm based on cosine transform and fractional Fourier transform in invariant wavelet domain, *Inf. Sci.*, **551** (2021), 205–227. https://doi.org/10.1016/j.ins.2020.11.020
15. W. Zhou, C. Liu, J. Lei, L. Yu, T. Luo, HFNet: Hierarchical feedback network with multilevel atrous spatial pyramid pooling for RGB-D saliency detection, *Neurocomputing*, 2021. https://doi.org/10.1016/j.neucom.2021.11.100
16. W. Zhou, J. Liu, J. Lei, L. Yu, J. Hwang, GMNet: Graded-feature multilabel-learning network for RGB-thermal urban scene semantic segmentation, *IEEE T. Image Process.*, **30** (2021), 7790–7802. https://doi.org/10.1109/TIP.2021.3109518
17. M. Tancik, B. Mildenhall, R. NG, Invisible hyperlinks in physical photographs, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, 2117–2126. https://doi.org/10.1109/CVPR42600.2020.00219
18. S. Anguraj, P. S. Shantharajah, E J. Jeba, A steganographic method based on optimized audio embedding technique for secure data communication in the Internet of Things, *Comput. Intell.*, **36** (2019), 557–573. https://doi.org/10.1111/coin.12253
19. P. Hu, D. Peng, Z. Yi. Y. Xiang, Robust time-spread echo watermarking using characteristics of host signals, *Electron. Lett.*, **52** (2016), 5–6. https://doi.org/10.1049/el.2015.1508
20. M. Mosleh, S. Setayeshi, B. Barekatain, M. Mohammad, A novel audio watermarking scheme based on fuzzy inference system in DCT domain, *Multimed. Tools Appl.*, **80** (2021), 20423–20447. https://doi.org/10.1007/s11042-021-10686-6
21. M. Abdelwahab Khaled, M. Abd El-atty Saied, Wi. El-Shafa, S. El-Rabaie, F. E. Abd El-Samie, Efficient SVD-based audio watermarking technique in FRT domain, *Multimed. Tools Appl.*, **79** (2020), 5617–5648. https://doi.org/10.1007/s11042-019-08023-z
22. H. Karajeh, T. Khatib, L. Rajab, M. Maqableh, A robust digital audio watermarking scheme based on DWT and Schur decomposition, *Multimed. Tools Appl.*, **78** (2019), 18395–18418. https://doi.org/10.1007/s11042-019-7214-3
23. S. Bhargavi Latha, D. Venkata Reddy, A. Damodaram, Video watermarking using neural networks, *Int. J. Inf. Comput. Secur.*, **14** (2021), 40–59. https://doi.org/10.1504/IJICS.2021.112207

24. J. Sang, Q. Liu, C. Song, Robust video watermarking using a hybrid DCT-DWT approach, *J. Electron. Sci. Technol.*, **18** (2020), 179–189. https://doi.org/10.1016/j.jnlest.2020.100052
25. Kh. Manglem Singh, A robust rotation resilient video watermarking scheme based on the SIFT, *Multimed. Tools Appl.*, **77** (2018), 16419–16444. https://doi.org/10.1007/s11042-017-5213-9
26. A. Rakesh, B. Sarabjeet Singh, Video watermarking scheme based on IDR frames using MPEG-2 structure, *Int. J. Inf. Comput. Secur.*, **11** (2019), 585–603. https://doi.org/10.1504/IJICS.2019.103065
27. C. Li, Y. Yang, K. Liu, L. Tian, A Semi-Fragile video watermarking algorithm based on H.264/AVC, *Wirel. Commun. Mob. Comput.*, 2020. https://doi.org/10.1155/2020/8848553
28. F. Madine, M. A. Akhaee, N. Zarmehi, A multiplicative video watermarking robust to H.264/AVC compression standard, *Signal Process. Image Commun.,* **68** (2018), 229–240. https://doi.org/10.1016/j.image.2018.06.015
29. S. Gaj, A. Sur, PK. Bora, Prediction mode based H.265/HEVC video watermarking resisting re-compression attack, *Multimed. Tools Appl.*, **79** (2020), 18089–18119. https://doi.org/10.1007/s11042-019-08301-w
30. J. Dittmann, A. Steinmetz, R. Steinmetz, Content-based Digital Signature for Motion Pictures Authentication and Content-Fragile Watermarking, IEEE International Conference of the Multimedia Systems, **1** (1999), 574–579. https://doi.org/10.1109/MMCS.1999.779264
31. P. Agrawal, A. Khurshid, DWT and GA-PSO Based Novel Watermarking for Videos Using Audio Watermark, International conference on swarm intelligence. ICSI, 2014, 212–220. https://doi.org/10.1007/978-3-319-11897-0_25
32. M. Sun dararajan, G. Yamuna, CWT and CS algorithm based video watermarking using audio watermark, *Procedia Comput. Sci.*, **87** (2016), 93–98. https://doi.org/10.1016/j.procs.2016.05.132
33. X. Wang, Y. Pan, Audio and video cross watermarking algorithm based-on visual saliency model, *Electron. Meas. Technol.*, **40** (2017), 112–115.
34. Z. Esmaeilbeig, S. Ghaemmaghami, Compressed Video Watermarking for Authentication and Reconstruction of the Audio Part, 2018 15th International ISC (Iranian Society of Cryptology) Conference on Information Security and Cryptology. ISCISC. 2018, 1–6. https://doi.org/10.1109/ISCISC.2018.8546897
35. G. Sucharitha, R. Kumar Senapati, Biomedical image retrieval by using local directional edge binary patterns and Zernike moments, *Multimed. Tools Appl.*,**79** (2020), 1847–1864. https://doi.org/10.1007/s11042-019-08215-7
36. J. Qu, Image encryption algorithm based on Logistic chaotic scrambling, *Science and Technology Innovation*, 2020, 2.
37. Y. Jiang, M. Cai, C. Song, SIFT based video watermarking algorithm against manifold attacks in contourlet domain, *Comput. Simul.*, **35** (2018), 314–320.
38. C. Maiti, B. C. Dhara, Robust non-blind video watermarking using DWT and QR decomposition, *Adv. Intel. Syst. Comput.*, **999** (2020), 333–343. https://doi.org/10.1007/978-981-13-9042-5_28
39. R. Singh, A. Ashok, M. Saraswat, Robust Video Watermarking in Frequency Domain for Copyright Protection, ACM International Conference Proceeding Series, AICPS, 2021, 174–178. https://doi.org/10.1145/3474124.3474148

AIMS Press