Hindawi
Complexity
Volume 2022, Article ID 3445583, 14 pages
https://doi.org/10.1155/2022/3445583

Research Article

Audio and Video Matching Zero-Watermarking Algorithm Based on NSCT

Di Fan, Wenxue Sun, Huiyuan Zhao, Wenshuo Kang, and Changzhi Lv

Shandong University of Science and Technology, Qingdao, Shandong 266590, China

Correspondence should be addressed to Changzhi Lv; lvchangzhi@126.com

Received 2 April 2022; Revised 15 June 2022; Accepted 29 July 2022; Published 24 August 2022

Academic Editor: Wenjie Lu

In the Internet age, information security is threatened anytime and anywhere and the copyright protection of audio and video as well as the need for matching detection is increasingly strong. In view of this need, this paper proposes a zero-watermarking algorithm for audio and video matching based on NSCT. The algorithm uses NSCT, DCT, SVD, and Schur decomposition to extract video features and audio features and generates zero-watermark stream through synthesis, which is stored in a third-party organization for detection and identification. The detection algorithm can obtain zero watermark from the audio and video to be tested and judge and locate tampering by comparing with the zero watermark of the third party. From the experimental results, this algorithm can not only detect whether the audio and video are mismatched due to tampering attacks but also locate the mismatched audio and video segments and protect the copyright.

1. Introduction

With the development of global networking, digital media is fast and convenient. While bringing convenience, security issues are increasingly prominent. Digital watermarking technology can protect the copyright of audio or video to a certain extent and is a hot research field of data security. But at present, the matching of audio and video cannot be detected and located by digital watermarking technology, which is a blind area of security protection and detection. Therefore, the research on audio and video matching detection and location is urgent.

At present, there are very few watermarking algorithms for audio and video matching detection. Most of the digital watermarking algorithms are image watermarking, audio watermarking, video watermarking, etc. The media attached to the algorithm are single. Image digital watermarking mainly includes spatial domain method [1], transform domain method [2–6], and deep learning-based method [7]. Transform domain method commonly uses DCT (Discrete Cosine Transform), NSCT (Nonsubsampled Contourlet Transform) [4, 5], DWT (Discrete Wavelet Transform) [6], and so on. As a new direction, watermarking algorithm based on deep learning appears on the way of watermarking technology, but it still needs to be improved in terms of watermark capacity and algorithm complexity. Video watermarking can be divided into original video watermarking algorithm and video watermarking algorithm based on compression domain. The former can refer to the existing image watermarking algorithm [8–10]. The latter is a watermarking technology combined with specific video encoding methods, such as MPEG [11], H.264 [12], and H.265 [13] video watermarking algorithms. Audio watermarking algorithms mainly include time domain and transform domain algorithms, and time domain audio algorithms include least significant bit algorithm [14], echo hiding algorithm [15], and phase coding algorithm. In order to improve the robustness of watermarking algorithm, more scholars begin to pay attention to the research of watermarking algorithm in transform domain and transfer the embedding position of watermark from time domain to transform domain. For example, [16] proposed the audio watermarking technology based on DWT, [17] proposed the audio watermarking technology based on SVD (singular value decomposition) and fractional Fourier transform, and [18] proposed the audio watermarking technology based on

DWT and SVD. At present, most of the watermarking algorithms of audio and video are designed separately, but multimedia data is composed of audio and video together, so it is not enough to protect only one of them. Tamper protection or even matching protection is needed for both audio and video. Dittmann et al. [19] proposed the earliest cross-watermarking algorithm in 1999, which can verify the synchronization between audio and video. Although this cross-watermarking algorithm can be easily implemented, the watermark cannot resist various attacks [19]. In order to improve the robustness of watermarking, Wang and Pan [20] proposed an audio-video cross-watermarking algorithm combined with a visual saliency model, which embedded the watermark into the DC coefficient of DCT through quantitative index modulation [20]. Esmaeilbeig and Ghaemmaghami [21] proposed an audio and video watermarking algorithm based on compressed domain. The algorithm generates hash bits in the audio part and embeds them as watermarks in the QDCT coefficients of video Immc1 frames [21]. The above audio and video cross-watermarking does not provide copyright protection for audio and video at the same time but only generates watermarks based on the whole multimedia stream, which can only judge whether the whole audio and video match and cannot locate the tampering of small segments in audio and video streams. Sun et al. [22] proposed a video zero-watermarking algorithm based on NSCT, DCT, DWT, and SVD. The algorithm generates zero-watermarking frame by combining audio watermark with video frame feature matrix, which can be utilized locating the attacks for the video besides verifying its copyright [22].

This paper presents an audio and video matching digital watermarking algorithm based on NSCT transform. The algorithm extracts video features and audio features of each segment, respectively, and generates a zero-watermark stream through synthesis. Experiments show that this algorithm can not only detect whether the audio and video are mismatched due to tampering attacks and locate the mismatched audio and video segments but also protect the copyright.

2. General Framework of Audio and Video Matching Zero-Watermarking Algorithm

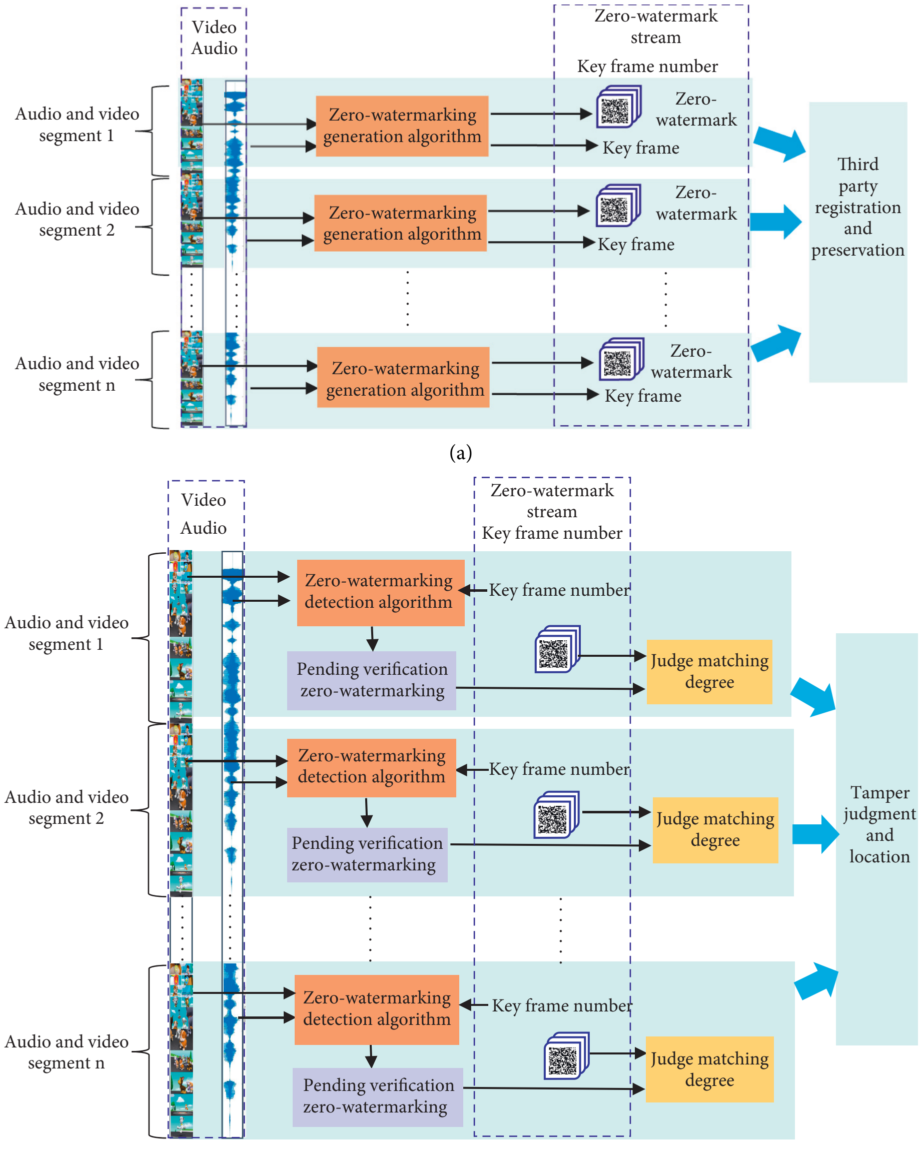
The difference between zero watermarking and traditional digital watermarking is that it is not really embedded into the carrier, but it is obtained by extracting the stable features of the carrier to construct the feature moment and performing XOR operation with the watermark information. This paper can not only generate zero watermark but also realize the matching detection of audio and video, and its generation and detection framework is shown in Figure 1. The generation algorithm first preprocesses the audio and video and segments them by 1s, and the audio and video segments are synchronized and corresponding in time. Then, the audio stable features are extracted from the audio segment to construct the audio feature matrix, and the key frames and their features are extracted from the visual frequency band to generate the encrypted video watermark. XOR is performed between the encrypted video watermark and the audio feature matrix to obtain the zero watermark of the segment. The zero watermark generated by each segment is integrated with its audio and video features. When the whole audio and video performs the same operation, a zero-watermark stream is formed, which is saved together with the key frame number and other information to a third party such as the copyright center. The matching detection process is to generate zero watermark for audio and video segments in the same way as that of the copyright center and detect the matching of audio and video by comparing with the zero watermark of the copyright center. In addition to detecting audio and video matching, this zero watermark can also be used for traditional copyright recognition.

3. Zero-Watermarking Generation Algorithm for Audio and Video Matching

The zero-watermark generation algorithm for audio and video matching is shown in Figure 2. The audio and video are decoded and segmented in 1s to obtain several short audio and video pairs composed of video and audio segments. Each audio and video pair are matched and detected so as to realize audio and video tampering judgment and positioning in a small time period. Video watermarking is generated by NSCT, DCT, Schur decomposition, and other algorithms. DWT and SVD algorithms are used to extract audio features. The encrypted video watermarking is XOR operated with the extracted sound feature matrix to obtain the audio and video matching zero watermark. Zero watermark will be registered by the third-party copyright organization to save, when the audio and video need to be authenticated and detected out of the use.

3.1. Generation of Encrypted Video Watermark. Video watermark is composed of key frame features of video segment. First, the key frame image is extracted based on frame difference Euclide distance method, and the extracted key frame number is saved as the key, and the video frame image as the watermark is found by the key in the zero-watermarking detection of audio and video matching. Based on the key frame image, it is converted from RGB space to YCoCg color space. The Co component was decomposed by NSCT, DCT, Schur decomposition, and other methods to generate the video feature matrix, which was binarized and encrypted to obtain the encrypted video watermark. The detailed steps of generating encrypted video watermarks are shown in Figure 3.

3.1.1. Key Frame Extraction Algorithm Based on Euclidean Distance between Frames. This algorithm uses the method based on the Euclidean distance between frames to extract key frames [23]. The main idea of this method is to calculate the Euclidean distance of two consecutive frames of images and select the key nodes through the Euclidean distance of images. This method is simple and easy to operate. The definition of interframe Euclidean distance is shown in the following equations:


Video
Audio
Audio and video segment 1
Audio and video segment 2
Audio and video segment n
Zero-watermarking generation algorithm
Zero-watermark stream
Key frame number
Zero-watermark
Key frame
Third party registration and preservation
Zero-watermarking detection algorithm
Pending verification zero-watermarking
Key frame number
Judge matching degree
Tamper judgment and location

(a)

(b)

Figure 1: Total frame of zero-watermarking generation and detection based on the audio and video. (a) The overall framework of zero-watermarking generation algorithm for audio and video segments. (b) The overall framework of zero-watermarking detection algorithm.

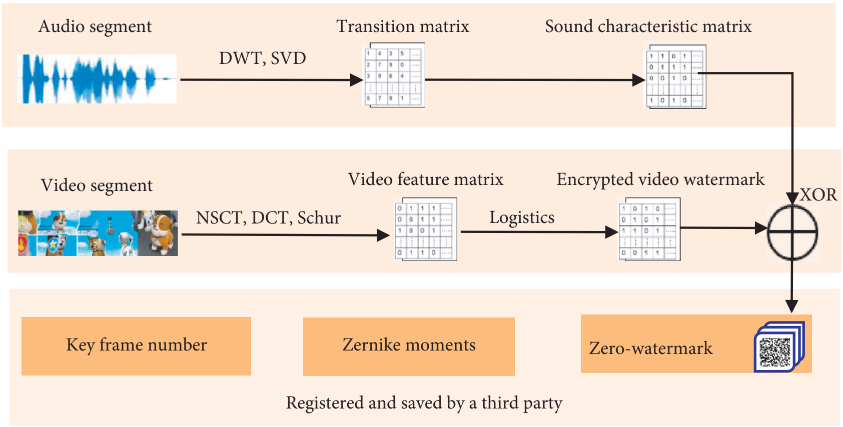


Figure 2: Zero-watermark generation process based on audio and video features.

```
Algorithm 1 name:  Generation of encrypted video watermark
Input:             Keyframe image I
Initialize:        D=zeros (32,32)
Begin:
    calculate $B_{nm}$ = Zernike ($l$)  /*Calculate the Zernike moment $B_{nm}$ of I*/
    Convert I from RGB to YCocg and extract Co component
    L=NSCT2 (Co)            /*The low-frequency subband L is obtained by two-layer
                            NSCT transformation of Co*/
    L is divided into non overlapping blocks, and block (i,j) is remarked as $n_{i,j}$ ($i,j$ = 1,2,...,32)
    $N_{i,j}$ = DCT2 (ni,j)      /*2-D DCT transformation of ni,j*/
    for i=1:32
        for j=1:32
            [$U_{i,j}$, $T_{i,j}$] = Schur ($N_{i,j,}$)    /*Schur decomposition of $N_{i,j}$*/
            $\lambda$ = diag ($T_{i,j}$)
            d (i,j)=max ($\lambda$)
        end for
    end for
calculate M=mean (d)
    a=find (d>M)   /*Look for points greater than M in the matrix D*/
    D (a)=1
    W=logistics (D)  /*Encrypt D with logistics to obtain the encrypted video watermark W*/
Output:            Encrypted video watermark W
```

Figure 3: Pseudocode of generating algorithm of encrypted video watermark.

$$t_k(i,j) = \left(g_{k+2}(i,j) - g_{k+1}(i,j)\right) - \left(g_{k+1}(i,j) - g_k(i,j)\right), \quad (1)$$

$$S_k = \sqrt{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} t_k^2(i,j)}, \quad (2)$$

where $g_k(i,j)$, $g_{k+1}(i,j)$, and $g_{k+2}(i,j)$ represent the gray value of the k frame image, $k+1$ frame image, and $k+2$ frame image at pixel point (i,j), respectively, k represents the number of frames of the video, and $k = 1, 2, 3, \ldots, J$. $t_k(i,j)$ represents the gray difference between the $k+2$ frame image and the $k+1$ frame image minus the gray difference between the $k+1$ frame image and the k frame image. The image size is $M \times N$.

The steps of extracting key frames based on the Euclidean distance between frames are as follows:

(1) Use (1) and (2) to calculate the Euclidean distance of each frame of image. If there are J frames of images, there are $J-2$ Euclidean distances.

(2) Calculate the extreme value of $J-2$ Euclidean distances.

(3) Find the maximum and minimum values of these extreme points and calculate their mean values.

(4) Compare each extreme point and the mean value. The image corresponding to the extreme point greater than the mean value is the key frame image.

3.1.2. NSCT Transform. NSCT has multiscale property and good anisotropy and translational invariance. NSCT transform is composed of NSP (Nonsubsampled Pyramid) and NSDFB (Nonsubsampled Directional Filter Bank). The nonsampling tower filter performs multiscale decomposition on the image first and then removes the low-frequency part. The nonsampling direction filter bank performs directional decomposition on the high-frequency part, making the NSCT transform multiscale and multidirectional anisotropy. The principle of three-stage NSCT transformation is shown in Figure 4. Its output is low-frequency $y1$ and three-stage high-frequency $y2$, $y3$, and $y4$, and its direction numbers are 2, 4, and 8, respectively.

After NSCT, the low-frequency part gathers the energy of the image and represents the contour information of the image, while the high-frequency part contains less energy of the image. The algorithm in this paper can ensure the embedding strength of watermark by taking advantage of the large energy value of the low-frequency part transformed by NSCT and the same size of the image as the original image, so the transformed low-frequency subband is selected as the object to construct zero watermark.

3.1.3. DCT Transform. DCT is a kind of orthogonal real transform, which has strong information concentration ability and is widely used in digital watermarking technology because of its strong robustness and good concealment [24]. For the two-dimensional image $f(x, y)$, its DCT and its inverse transform are shown as follows:

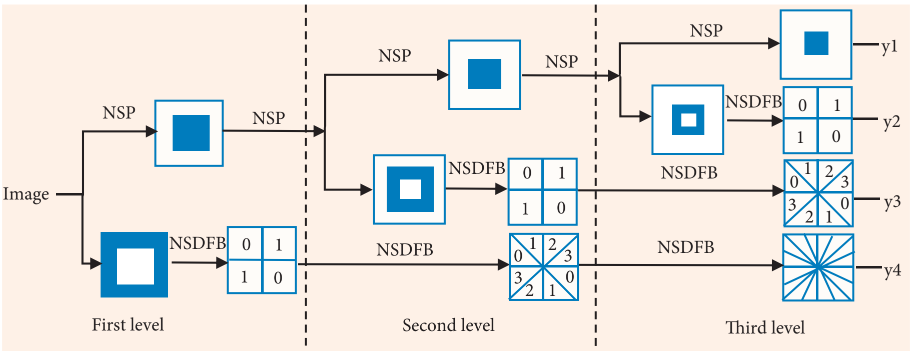


Figure 4: NSCT decomposition block diagram.

$$\begin{aligned} F(u,v) &= \frac{2}{\sqrt{MN}} c(u)c(v) \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\cos\frac{(2x+1)u\pi}{2M}\cos\frac{(2y+1)\pi}{2N}, \\ f(x,y) &= \frac{2}{\sqrt{MN}} \sum_{u=0}^{M-1}\sum_{v=0}^{N-1} c(u)c(v)F(u,v)\cos\frac{(2x+1)u\pi}{2M}\cos\frac{(2y+1)v\pi}{2N}, \end{aligned} \tag{3}$$

where $c(u) = c(v) = \left\{ \begin{matrix} 1/\sqrt{2}yu & v = 0 \\ 1 & \text{others} \end{matrix} \right\}$, u and v are the horizontal and vertical frequency, respectively, x, y are the pixel coordinates, $x, u = 0, 1, 2, \cdots, M-1$,and $y, v = 0, 1, 2, \cdots, N-1$.

3.1.4. Schur Decomposition. Schur decomposition decomposes a matrix X into the unit orthogonal matrix Y and the upper triangular matrix U such as $X = YUY^H$, and Y^H is the conjugate transpose of Y [25]. Then, the Schur of any n-order square matrix X can be decomposed into

$$X = YUY^H = \begin{bmatrix} \lambda_1 & r_{12} & \cdots & r_{1n} \\ & \lambda_2 & \ldots & r_{2n} \\ & & \ddots & \vdots \\ & & & \lambda_n \end{bmatrix}. \tag{4}$$

Schur decomposition is widely used in digital watermarking because of its scaling invariance and low computational complexity. When the matrix is scaled by a certain multiple, only the eigenvalues change by a multiple. The scaling invariance of Schur decomposition can deal with scaling attack well and improve the robustness of watermarking. In addition, Schur decomposition is a step of singular value decomposition; it does not need to transform the upper triangular matrix into diagonal matrix, so the calculation is less.

3.2. Generation of the Sound Feature Matrix. The sound feature matrix is generated from the features of the audio segment. The algorithm performs DWT and SVD on the segmented decoded audio to obtain stable audio features. Based on this, the feature matrix is formed and binarization is carried out. After that, XOR generates zero watermark for the encrypted video watermarking. The detailed steps of sound feature matrix generation are shown in Figure 5.

4. Audio and Video Matching Detection Algorithm

The video matching detection algorithm and the audio-video matching zero-watermark generation algorithm are inverse processes to each other, as shown in Figure 6. Supported by the key frame number, Zernike moment, and other information saved by the third party, the zero-watermarking generation algorithm is used to obtain the zero watermark of the audio and video to be tested. The similarity between the zero watermark to be detected and the zero-watermark stream saved by the third party is judged, and whether the audio and video segment has been tampered is determined according to the similarity threshold. The Zernike moment can better resist rotation attack. The detailed steps of audio and video matching detection algorithm are shown in Figure 7.

5. Experimental Results and Analysis

The experiment is carried out on MATLAB R2018b. The watermark is encrypted by logistic chaos, and its initial value x_0 and parameter u are used as the key. Only by knowing the zero-watermark algorithm, encryption method, and its key can the watermark information be decrypted correctly. Considering the security of the algorithm and watermark, the parameter of Logistic chaotic encryption is set as $x_0 = 0.1, u = 4$. For the length of audio and video segments, this paper determines that the audio and video segmentation unit is 1s through comprehensive analysis and experiments from the aspects of the stability of audio and video features, the rapidity of generating zero watermark, the minimization of occupied resources, the

```
Algorithm 2 name:  Generation of sound characteristic matrix
Input:             Audio segment Q, Vector length of each segment n
Initialize:        F=zeros (32,32)
Begin:
    [Ca2,L]=wavedec (Q,2,'haar')    /*Extract the approximate component Ca2 of wavelet transform of audio Q*/
    l=floor (sqrt (n))
    for k=1:1024   do
        A=Ca2 ((k-1)n+1 :kn)        /*Divide Ca2 into 1024 segments, each with a length of n*/
        J=reshape (A, l, l)         /*Upgrade A dimension to a matrix of 32*32 size*/
        [U,S,V]=SVD (J)             /*The diagonal matrix s is obtained by SVD decomposition of J*/
        e (k)=S (1,1)
    end for
    e1 = reshape (e,32,32)          /* Upgrade e dimension to a matrix of 32 * 32 size*/
    calculate m=mean (e1)
    b=find (e1>m)                   /* Look for the point in E1 greater than m*/
    F (b)=1
Output:            Sound characteristic matrix F
```

Figure 5: Pseudocode of sound feature matrix generation algorithm.

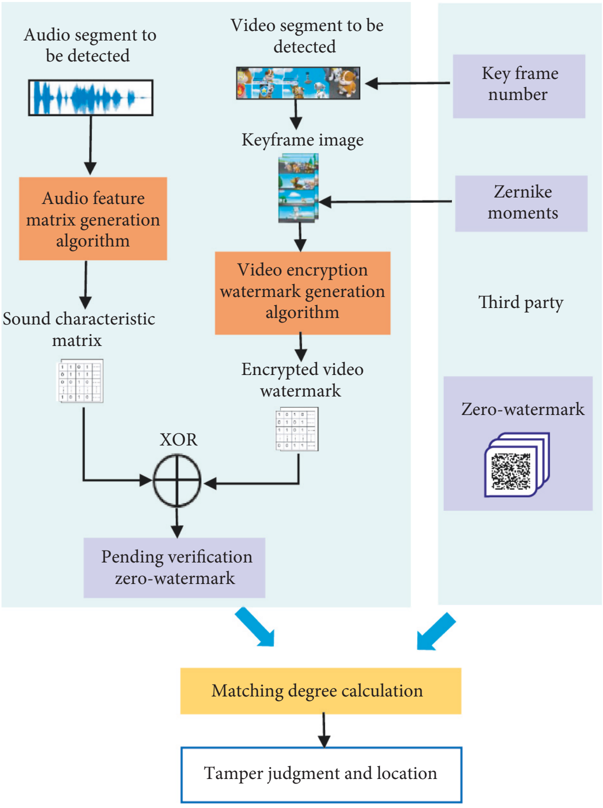

Audio segment to be detected
Video segment to be detected
Key frame number
Keyframe image
Zernike moments
Audio feature matrix generation algorithm
Video encryption watermark generation algorithm
Third party
Sound characteristic matrix
Encrypted video watermark
Zero-watermark
XOR
Pending verification zero-watermark
Matching degree calculation
Tamper judgment and location

Figure 6: Audio and video matching detection and location process.

accuracy of matching detection, and so on. In this way, on the one hand, it can effectively extract the stable features of audio and video segments and quickly build an optimized zero watermark. On the other hand, it can also detect the tampering of small audio or video segments in the entire audio and video stream more accurately [22]. The following experiments use the video (including audio) in H.264 coding format, which is divided into 30 audio and video segments in the experiment. The video frame size is 1080 × 1920, the duration is 30 seconds, the frame rate is 27 fps, the audio stream sampling rate is 44.1 KHz, 16-bit quantization bits, and two channels.

```
Algorithm 3 name:  Zero-watermark detection algorithm
Input:             Audio and video streaming V, Zernike moment $A_{nm}$, Key frame number,
                   Zero-watermark W, Tamper judgement threshold T
Begin:             Decode the V segment top obtain the audio segment and video segment to be
                   verified
                     Extract key frame image G according to frame number
                     $\alpha = \frac{\arg(A'_{nm} - A_{nm})}{m}$  /* The Angle that the keyframe needs to be adjusted, where
                                          $m$ is the multiplicity of Zernike moments $n$ is the order of
                                          Zernike moment*/
                     The adjusted image $G$ is obtained by rotating $G$ angle $a$
                     Produce w for $G$ using Algorithm 1
                                          /*Generate encrypted video watermark*/
                     Produce F' for audio segment using Algorithm2
                                          /*Generate sound characteristic matrix*/
                     W=w'    F'
                     calculate NC         /*$NC(W',W) = \frac{\sum_i^N \sum_i^N W'(i,j)W(i,j)}{\sum_i^N \sum_i^N W(i,j))^2 \times \sum_i^N \sum_i^N W(i,j))^2}$*/
                     ifNC>T then          /* T is the initial set tamper judgement threshold*/
                       Flag=0             /* V is not tampered with and remarked as 0*/
                     else
                       Flag=1             /* V is tampered with and marked as 1*/
                     end if
Output:            Normalized correlation NC, Flag
```

Figure 7: Audio and video matching detection algorithm pseudocode.

In the experiment, the NC (Normalized Correlation) and BER (Bit Error Ratio) are used as the objective evaluation standard of watermark robustness. The NC experiment and analysis of the watermark image show that when the NC value is above 0.8, the correlation between the two watermark images is high [24]. Therefore, the tamper-proof threshold of audio and video is set as 0.8 in this paper for audio and video matching detection and identification; that is, when NC is greater than or equal to 0.8, audio and video are matched. When the value is less than 0.8, the audio or video is tampered with [22]. BER refers to the percentage of the extracted watermark error bits in the total bits. The PSNR (Peak Signal-to-Noise Ratio) is used as the difference measurement index of two images. The larger the value of PSNR, the better the invisibility of the watermark algorithm.

5.1. Audio and Video Matching and Tamper-Proof Test. For the above experimental audio and video, we replaced the video frames and audio segments in different time periods and then carried out the audio and video matching detection and positioning experiment. The experimental results are shown in Figure 8. The NC values of the zero watermark detected in Figure 8(a) are all less than the set threshold value of 0.8, so it is determined that they do not match. Therefore, segments 2, 5, 8, 11, 13, 16, 20, 23, 25, and 27 of the video are tampered with. The NC values in Figure 8(b) are all lower than the initially set threshold value, so it is determined that they do not match. Therefore, audio segments 2, 6, 8, 12, 15, 18, 21, 24, 26, 28, and 30 are tampered with. Experiments show that this method can detect whether the audio and video are mismatched due to tampering attacks and can locate the mismatched audio and video segments.

5.2. Algorithm Robustness Testing

5.2.1. Video Robustness Testing. In order to verify the robustness of this algorithm, common attacks such as Gaussian noise, salt and pepper noise, clipping, scaling, rotation, Gaussian filtering, median filtering, and frame attack are carried out on the video, as well as the combination of several one-way attacks, and the experimental results are shown in Table 1. From the whole experimental results, after the attack, even if some PSNR has reached below 10 dB, the NC value of the watermark of this algorithm is still above 0.9, indicating that the algorithm has good robustness:

(1) *Noise Attack.* Noise attack is one of the most common types of attacks. The algorithm in this paper has carried out Gaussian noise and salt and pepper noise attack experiments on the video. The results are shown in Figure 9. The range of noise attack intensity is 0.01–0.1, with 0.01 as an interval. The figure demonstrates that as noise level increases, the signal-to-noise ratio of the video key frame image is decreasing, but the NC mean of the watermark remains above 0.95, which shows that the algorithm in this paper has good robustness to noise attacks.

(2) *JPEG Compression Attack.* In this paper, the robustness of the algorithm is tested for JPEG

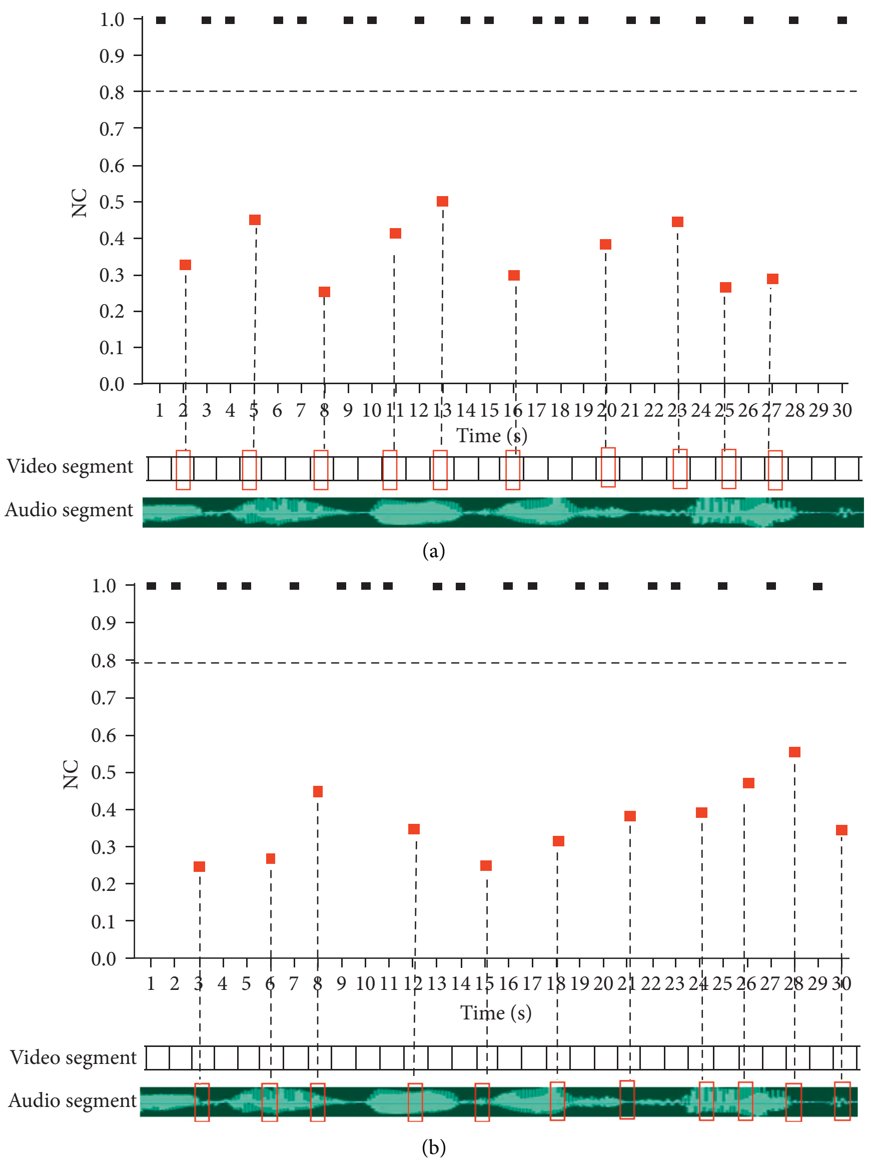


Figure 8: Experimental results of matching detection and location of audio and video. (a) Experimental results of matching detection after video tampering. (b) Experimental results of matching detection after audio tampering.

compression in the range of quality factor 10–90 with increments of 10 intervals, which is shown in Figure 10. According to experiment results shown in the figure, when the quality factor improves, the NC values which were extracted from key frames steadily rise and the distribution becomes more concentrated, and the NC values also increase with the improvement of quality factor. Within the experimental range, the NC values are greater than 0.96, indicating that the algorithm in this paper has good robustness in resisting JPEG compression.

(3) *Filter Attack*. In the research of image and video, image filtering is one of the most common operations. In this section, it is a Gaussian filtering attack that is applied on the video. As shown in Figure 11, when facing the Gaussian filtering attack, with the increase of the filter window size and the surrounding scale, the NC value of the watermark decreases, but it is still greater than 0.94, which shows that the robustness of the Gaussian filtering is better.

(4) *Shear Attack*. The algorithm in this paper conducts an attack experiment of cutting 1/20, 1/16, and 1/8 of the video on the upper left, lower left, upper right, and lower right, and the results are illustrated in Figure 12. The results demonstrate that because the algorithm extracts the features of the key frames when generating the watermark, even if the clipping attack will lead to the loss of a large number of features of the key frame image, the mean value of NC in the experiment is still above the matching detection threshold, which ensures the accuracy of the matching detection.

Table 1: Experimental results of video robustness.

Attack type	Attack parameter	PSNR	NC	BER (%)
Gaussian noise	0.01	20.7781	0.9947	0.1558
	0.05	14.3098	0.9853	0.4748
Salt and pepper noise	0.01	21.3424	0.9948	0.1646
	0.05	16.7343	0.9892	0.4720
Shear attack	Upper left 1/16	13.2199	0.9821	0.4759
	Lower right 1/20	14.8016	0.9930	0.1546
Scaling	1/2	33.5737	0.9985	0.1498
	2	45.3691	0.9994	0.1454
Frame attack	Frame average	28.8532	0.9932	0.15
	Frame reorganization	34.7721	0.9972	0.1499
Recompression	Mpeg4	32.2128	0.9853	0.4819
	H.264	33.2266	0.9875	0.4812
Rotate attack	15°	18.3589	0.98	0.4823
	90°	19.1075	0.9807	0.4817
	180°	19.9432	0.9914	0.1745
JPEG compression	90	45.8949	0.9984	0.1440
	50	35.4672	0.9917	0.1599
	20	30.9995	0.9841	0.4798
Combined attack	JPEG60 + rotate 30°	8.803	0.9523	1.7499
	JPEG30 + rotate 45°	8.2203	0.9340	2.2831
	JPEG10 + rotate 90°	6.2247	0.9120	3.4837
	Gaussian filtering 3*3 + upper left shear 1/20	14.1580	0.9835	0.4827
	Gaussian filtering 5*5+ upper left shear 1/20	14.0569	0.9815	0.4850
	Gaussian filtering 7*7 + upper right shear 1/20	14.4891	0.9841	0.4814

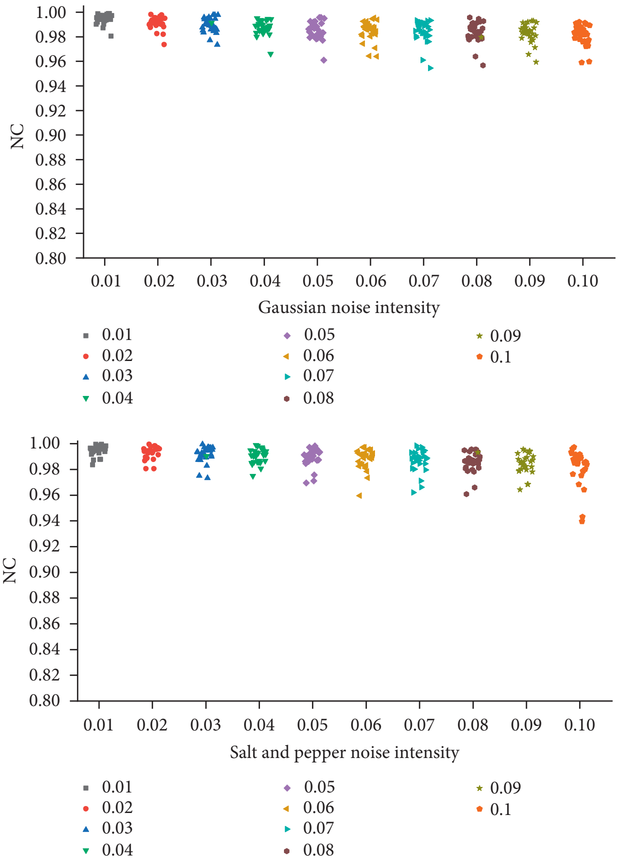


Figure 9: NC value under noise attack.

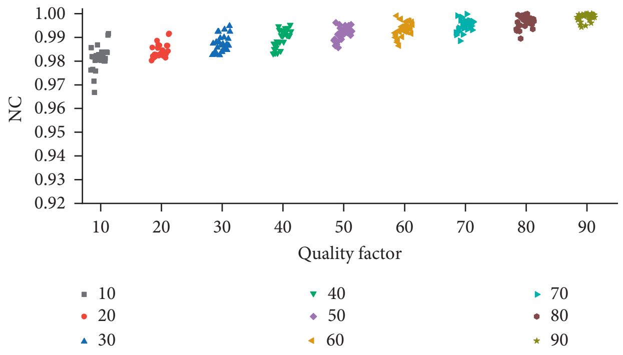


Figure 10: NC value under JPEG compression attack.

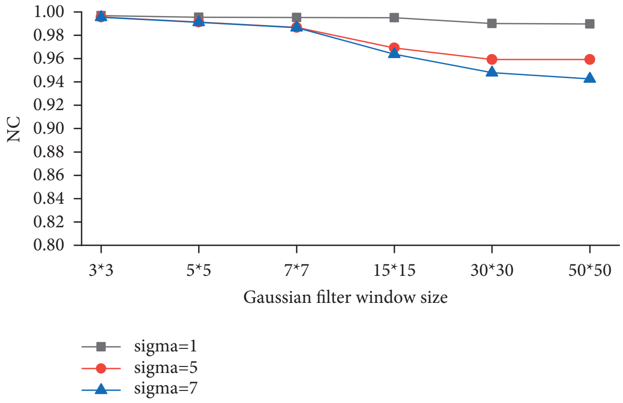


Figure 11: NC value under Gaussian filter attack.

(5) *Rotate Attack*. The algorithm in this paper carries out rotation attack from 0° to 180° on the video. It can be seen from Figure 13 that, with the increase of rotation angle, the NC value is gradually decreasing, but all of them are above 0.96, indicating that the algorithm can resist rotation attack well.

(6) *Scaling Attack*. The algorithm in this paper uses different scaling multiples to attack the video key frame images, respectively. As can be seen from Figure 14, the NC values are above 0.96, indicating that the algorithm in this paper has good robustness to scaling attacks.

(7) *Combined Attack*. In actual audio and video transmission, video often suffers from more than one attack, and there may be multiple attacks acting at the same time. Robustness under combined attack is also an important aspect of algorithm performance. The algorithm in this paper selects three combined attack methods of rotation and JPEG compression attack, shearing and Gaussian filtering attack, and H.264+ other attacks to conduct experiments. The results are shown in Figures 15–17, respectively. In general, for the three combined attacks, the NC value of the watermark is above 0.9, and there is still a large margin space from the threshold of 0.8, indicating that the algorithm can well resist the combined attack.

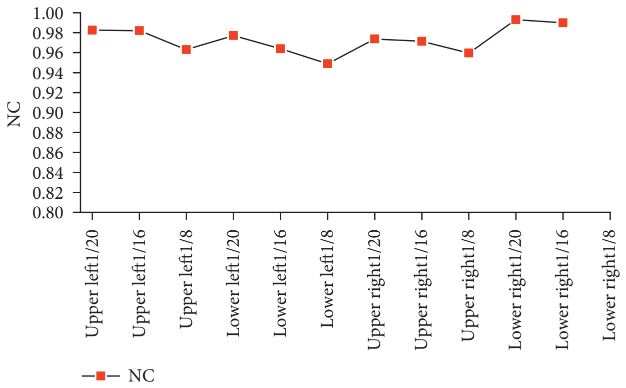


Figure 12: NC mean under shear attack.

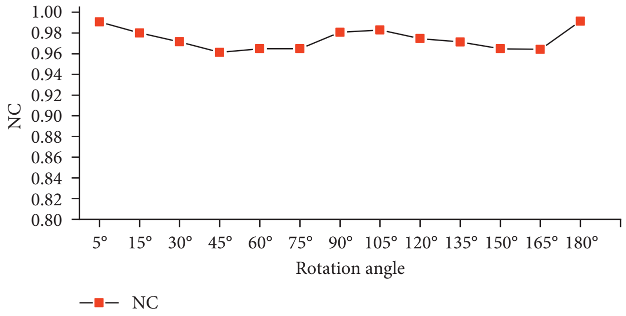


Figure 13: NC mean under rotation attack.

For Figure 15, the NC value of the watermark extracted by the algorithm under small-scale cropping and Gaussian filtering attacks can reach more than 0.96, which has strong robustness. Compared with the two attacks, the NC value of the algorithm is lower under the large-scale cropping attack, and the sensitivity to the cropping attack is slightly higher than that of the Gaussian filter.

For Figure 16, the experimental results show that the algorithm has good antiattack ability against the combined attack of rotation + JPEG compression. Most of the extracted watermark NC values are about 0.94, and it can be seen that the sensitivity of the algorithm to rotation attack is higher than that of JPEG compression attack.

For Figure 17, under the combined attack of format conversion and other attacks, the NC value of the extracted watermark is relatively high, which can be used for matching detection. Further analysis will find that the sensitivity of different video frames to the attack is different, and the ability to resist the combined attack has a certain relationship with the image content.

5.2.2. Audio Robustness Test. Audio with watermark may encounter attacks in the process of transmission. Some attacks may be unintentional, such as noise. Although they may not affect the visual perception, they also affect the reliability of watermark; some attacks may be intentional, such as cutting, filtering, and compression. The algorithm needs to have sufficient attack ability to resist various attacks and ensure the reliability and security of the watermark. In

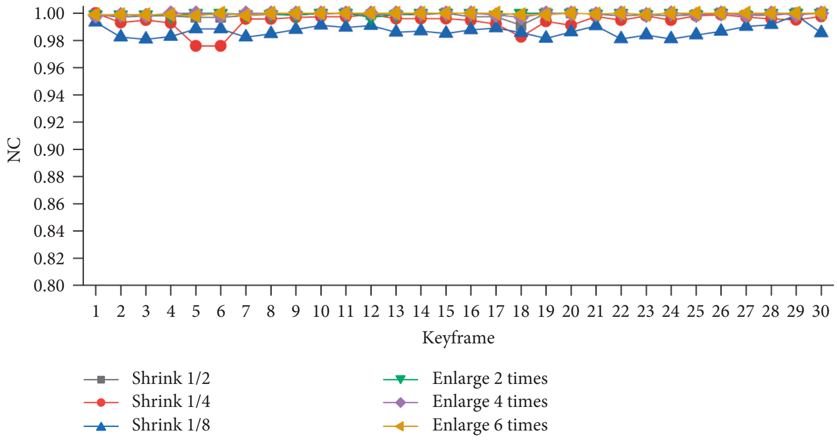


Figure 14: NC value under scaling attack.

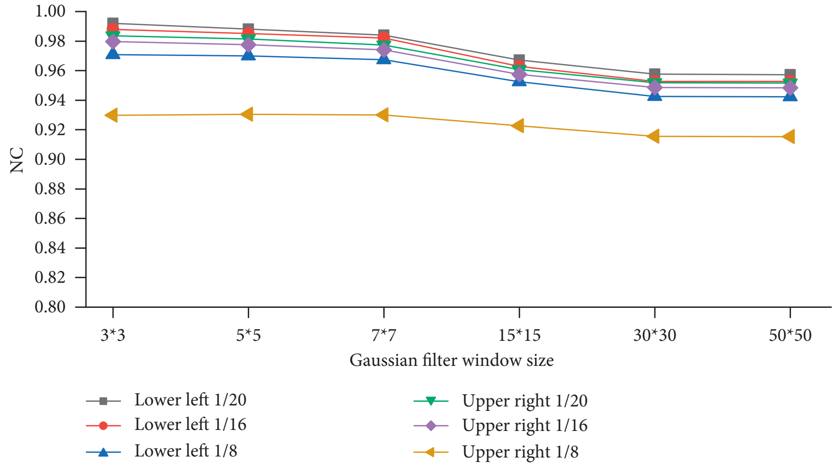


Figure 15: NC mean under combined attack of shear and Gaussian filter.

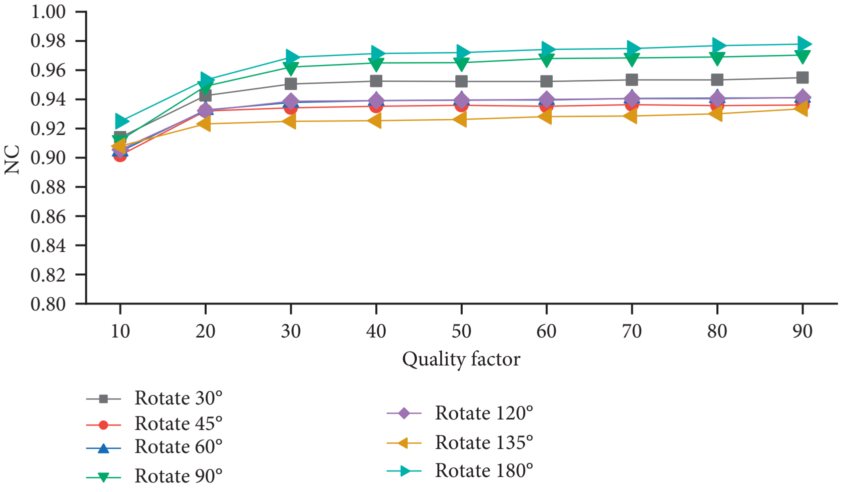


Figure 16: NC mean under combined attack of rotation and JPEG compression.

this paper, noise, weight, resampling, MP3 compression, and other attacks on audio are carried out, and the audio robustness experiments are carried out. The results are shown in Table 2. It can be seen from the data in the table that, under several attacks on the experiment, the watermark NC value obtained by the algorithm in this paper is more than 0.91, most of which are more than 0.99, and the robustness of the algorithm is good.

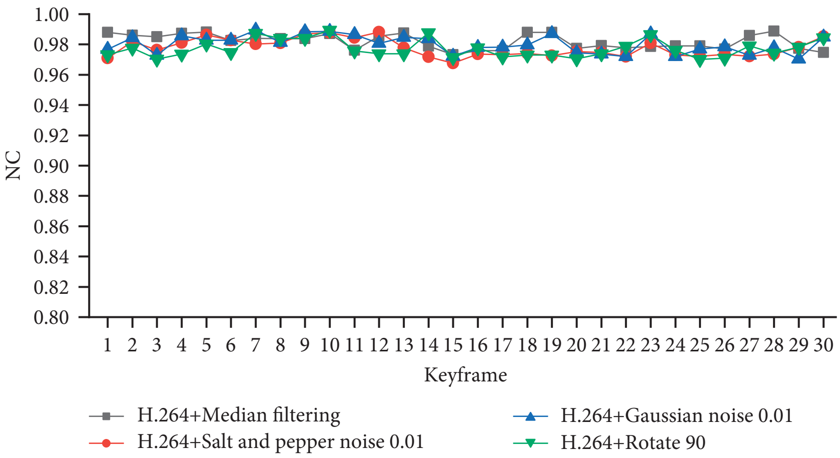


Figure 17: NC mean under H.264 and other attacks.

Table 2: Experimental results of audio robustness.

Attack type	Attack parameter	NC mean	PSNR mean	BER mean (%)
White Gaussian noise	20 dB	0.9191	28.8435	3.4829
	25 dB	0.9954	32.4838	0.1482
	30 dB	1	32.5631	0
Requantization	Up quantization	0.9995	31.6278	0.1473
	Down quantization	0.9962	32.0576	0.1492
Resampling	Upsampling	0.9939	31.5582	0.1495
	Downsampling	0.9879	33.5632	0.4796
Low-pass filtering	4 kHz	0.9954	32.9113	0.1489
	8 kHz	0.9938	32.9013	0.1496
MP3 compression	128 kbps	0.9917	31.4401	0.1507
	64 kbps	0.9938	31.9790	0.1495

Table 3: The experimental results of the proposed algorithm are compared with those of literature algorithms (NC value).

Experiment audio	Attack type	Literature [26] algorithm	Literature [27] algorithm	Proposed algorithm
Classical	Requantization	0.9910	0.9963	0.9978
	Resampling	0.9874	0.9899	0.9914
	Gaussian noise (20 dB)	0.9158	0.9467	0.9235
	Gaussian noise (30 dB)	0.9756	0.9999	1
	Low-pass filtering	0.9845	0.9913	0.9946
	MP3 compression	0.9389	0.9864	0.9969
Pop	Gaussian noise (20 dB)	0.9295	0.9665	0.9451
	Gaussian noise (30 dB)	0.9819	0.9993	1
	Low-pass filtering	0.9876	0.9911	0.9985
	MP3 compression	0.9539	0.9899	0.9984
	Requantization	0.9945	0.9985	0.9992
	Resampling	0.9956	0.9918	0.9996

5.3. *Comparison Experiment with Similar Algorithms.*

This paper makes relevant experiments on similar algorithms in literature [26, 27] and compares and investigates them with the algorithms in this paper. Literature [26] selects the audio segment according to the local time domain characteristics of the audio signal and uses DWT and SVD algorithms to construct a zero watermark for the selected audio segment. Reference [27] is a zero-watermarking method based on DWT-DCT-SVD. Compared with the two algorithms, the algorithm in this paper is different in feature extraction and decomposition methods.

In the experiment, two different styles of audio signals, classical and pop, are selected as the original audio carrier. They are mono audio signals with the sampling frequency of 44.1 kHz and quantization accuracy of 16 bits; [26, 27] adopt 32 × 32 fixed binary watermark image, and the algorithm in these papers adopts 32 × 32 binary video watermark image generated from video. The attack experimental results of this algorithm and two comparative literature algorithms are shown in Table 3. It clearly shows that the proposed algorithm has excellent robustness against Gaussian noise, weighting, resampling, and low-pass filtering attacks. These attacks are better than the comparison algorithm, and the advantage is more prominent under MP3 compression attack.

6. Conclusion

In this paper, a zero-watermarking algorithm for audio and video matching based on NSCT transform is proposed, which can detect whether the audio and video are mismatched due to tampering attacks, locate the mismatched audio and video segments, and play a role in protecting and identifying digital media information security. The algorithm uses NSCT, DCT, SVD, and Schur decomposition to extract video features and audio features and generate zero watermarking after synthesis. From the experimental results, the algorithm not only has strong robustness to common single-item attacks but also has high antiattack ability to combined attacks. Information security is a subject of constant development and change.

With the advancement of technology, the forms and types of attacks are also changing and improving, further improving antiattack capabilities, anti-new attack capabilities, robustness, and positioning speed and accuracy. It needs continuous research and continuous improvement.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

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

References

[1] J. Abraham and V. Paul, “An imperceptible spatial domain color image watermarking scheme,” *Journal of King Saud University-Computer and Information Sciences*, vol. 31, no. 1, pp. 125–133, 2019.

[2] D. Fan, Y. Li, and S. Gao, “A novel zero watermark optimization algorithm based on Gabor transform and discrete cosine transform,” *Concurrency and Computation: Practice and Experience*, vol. 2020, no. 2, 13 pages, 2020.

[3] M. Moosazadeh and G. Ekbatanifard, “An improved robust image watermarking method using DCT and YCoCg-R color space,” *Optik*, vol. 140, pp. 975–988, 2017.

[4] C. Kumar, A. K. Singh, P. Kumar, and S. Siddharth, “SPHIT-based multiple image watermarking in NSCT domain,” *Concurrency and Computation: Practice and Experience*, vol. 32, no. 1, pp. e4912.1–e4912.9, 2020.

[5] C. V. Narasimhulu, “A robust hybrid video watermarking algorithm using NSCT and SVD,” in *Proceedings of the IEEE International Conference on Power, Control, Signals and Instrumentation Engineering*, pp. 1495–1501, New York, NY, USA, September 2017.

[6] N. N. Zermi, A. Khaldi, R. Kafi, F. Kahlessenane, and S. Euschi, “A DWT-SVD based robust digital watermarking for medical image security,” *Forensic Science International*, vol. 320, no. 3.20, Article ID 110691, 2021.

[7] Yu Yang, M. Lei, X. Liu, Z. Qu, and C. Wang, “Novel zero-watermarking scheme based on DWT-DCT,” *China Communications*, vol. 13, no. 7, pp. 122–126, 2016.

[8] S. B. Latha, D. V. Reddy, and A. Damodaram, “Video watermarking using neural networks,” *International Journal of Information and Computer Security*, vol. 14, no. 1, pp. 40–59, 2021.

[9] C. Priya and C. Ramya, “Robust and secure video watermarking based on cellular automata and singular value decomposition for copyright protection,” *Circuits, Systems, and Signal Processing*, vol. 40, no. 5, pp. 2464–2493, 2020.

[10] A. Bhardwaj, V. S. Verma, and R. K. Jha, “Robust video watermarking using significant frame selection based on coefficient difference of lifting wavelet transform,” *Multimedia Tools and Applications*, vol. 77, no. 15, pp. 19659–19678, 2018.

[11] R. Ahuja and S. S. Bedi, “Video watermarking scheme based on IDR frames using MPEG-2 structure,” *International Journal of Information and Computer Security*, vol. 11, no. 6, pp. 585–603, 2019.

[12] Li Chen, Yi Yang, K. Liu, L. Tian, and H. Lu, “A semi-fragile video watermarking algorithm based on H.264/AVC,” *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 8848553, 11 pages, 2020.

[13] F. Madine, M. A. Akhaee, and N. Zarmehi, “A multiplicative video watermarking robust to H.264/AVC compression standard,” *Signal Processing: Image Communication*, vol. 68, pp. 229–240, 2018.

[14] S. Anguraj, S. P. Shantharajah, and E. J. Jeba, “A steganographic method based on optimized audio embedding technique for secure data communication in the Internet of Things,” *Computational Intelligence*, vol. 36, no. 2, pp. 557–573, 2019.

[15] P. Hu, D. Z. Peng, Z. Yi, and Y. Xiang, “Robust time-spread echo watermarking using characteristics of host signals,” *Electronics Letters*, vol. 52, no. 1, pp. 5-6, 2016.

[16] S. M. Pourhashemi, M. Mosleh, and Y. Erfani, “A novel audio watermarking scheme using ensemble-based watermark detector and discrete wavelet transform,” *Neural Computing & Applications*, vol. 33, no. 11, pp. 6161–6181, 2020.

[17] M. Abdelwahab Khaled, M. Abd El-atty Saied, Wi El-Shafa, S. El-Rabaie, and F. Abd El-Samie, “Efficient SVD-based audio watermarking technique in FRT domain,” *Multimedia Tools and Applications: International Journal*, vol. 79, no. 9-10, pp. 5617–5648, 2020.

[18] A. R. Elshazly, M. E. Nasr, M. M. Fouad, and F. E Abdel-Samie, “Intelligent high payload audio watermarking algorithm using colour image in DWT-SVD domain,” *Journal of Physics: Conference Series*, vol. 2128, no. 1, Article ID 012019, 2021.

[19] J. Dittmann, A. Steinmetz, and R. Steinmetz, “Content-based digital signature for motion pictures authentication and content-fragile watermarking,” in *Proceedings of the IEEE International Conference of the Multimedia Systems Multimedia Computing and Systems*, pp. 574–579, Florence, Italy, June 1999.

[20] X. Wang and Y. Pan, “Audio and video cross watermarking algorithm based on visual saliency model,” *Electronic Measuremeent Technology*, vol. 40, no. 8, pp. 112–115, 2017.

[21] Z. Esmaeilbeig and S. Ghaemmaghami, “Compressed video watermarking for authentication and reconstruction of the audio part,” in *Proceedings of the 2018 15th International ISC (Iranian Society of Cryptology) Conference on Information Security and Cryptology (ISCISC)*, pp. 1–6, IEEE, Tehran, Iran, August 2018.

[22] W. Sun, H. Zhao, X. Zhang et al., “Zero-watermarking algorithm for audio and video matching verification,” *AIMS Mathematics*, vol. 7, no. 5, pp. 8390–8407, 2022.

[23] M. Bao, G. Lu-yang, L. Xiao-dong, and J. Tian, “A study on optimum classification character based on the distributive entropy of euclidian distance,” *Journal of Optoelectronics - Laser*, vol. 2, no. 3, pp. 469–473, 2007.

[24] K. Huda and M. Mahmoud, “An imperceptible, robust, and high payload capacity audio watermarking scheme based on the DCT transformation and Schur decomposition,” *Analog Integrated Circuits and Signal Processing*, vol. 99, no. 3, pp. 571–583, 2019.

[25] W. Liu, S. Sun, and H. Qu, “Fast zero-watermarking algorithm based on Schur decomposition,” *Journal of Frontiers of Computer Science and Technology*, vol. 13, no. 3, pp. 494–504, 2019.

[26] X. Feng, G. Feng Naiguang, and Y. Wang, “Watermarking algorithm of audio signal based on discrete wavelet transform and singular value decomposition,” *Journal of Huaqiao University*, vol. 37, no. 6, pp. 770–773, 2016.

[27] S. Liu, Q. Du, H. Long, Y. Shao, and Y. Peng, “A robust audio watermarking algorithm based on DWT-DCT-SVD,” *Journal of Optoelectronics - Laser*, vol. 32, no. 9, pp. 1015–1022, 2021.