Proposal of Blind Extractable Additive Video Watermarking Method

Nao Harada, Rinka Kawano Masaki Kawamura †
Graduate School of Sciences and Technology for Innovation, Yamaguchi University
† E-mail: m.kawamura@m.ieice.org

***Abstract*—To improve the quality of stego video, it is necessary to reduce the total amount of information embedded in the video as a watermark. This paper proposes two techniques. The first technique is distributed watermarking. A watermark is embedded in some of the blocks. The same watermark is embedded in multiple frames. By time averaging the extracted watermarks, error can be reduced. The second technique is detection signal embedding. In order to detect stego frames, a detection signal is introduced. This allows the watermark to be embedded in any frame. In addition, resistance to frame removal attacks can be provided. We embedded a watermark in videos and evaluated its performance using normalized correlations of the extracted watermarks. We also evaluated the quality of the stego video. The results showed that compared with the DCT-DWT method, the performance of the proposed method was better under different types of attacks, including compression attacks and frame removal attacks. In addition, the image quality was high with a PSNR of over 41 dB.**

I. Introduction

Digital watermarking is a technique for the covert embedding of information into digital content such as still images, videos, and music. Depending on their purpose, watermarking methods can be divided into robust watermarking [1] and fragile watermarking [2]. Robust watermarking is used to protect copyrights and is resistant to compression and content modification. Fragile watermarking is for detecting content tampering and is therefore not resistant to compression or attacks. Watermarking methods include non-blind watermarking [3] and blind watermarking [4]. The former requires the original content to extract the watermark. The latter does not. Since non-blind methods are not practical in most cases where it is difficult to obtain the original image or where there is no permission to access the original image, blind methods are preferred.

In this paper, we focus on robust video watermarking. Similar to watermarking still images, there are two embedding domains for video frames: the spatial domain [5] and the frequency domain [6], [7]. In general, it is known that embedding in the frequency domain is more resistant to attacks than embedding in the spatial domain. Sang *et al.* [7] proposed a robust DCT-DWT method that embeds the watermark in a domain combining the discrete cosine (DCT) and discrete wavelet (DWT). However, their method was not practical because it was not blind. In addition, video embedding techniques also include additive embedding [6], [7] and quantization index modulation (QIM) [8]. Video is computationally intensive because it contains a lot of information. Therefore, it is desirable to be able to embed with a simple process such as additive embedding. However, existing methods using additive embedding are mostly non-blind. The method of Yokota and Kawamura [6] was able to extract additively embedded watermarks using blind source separation (BSS). During the extraction, stego frames can be detected from signals separated by BSS using the correlation between signals. However, the stego frames and the non-embedded frames were adjacent to each other. As a result, the video image flickered during playback. Therefore, it was easy to detect that the watermark existed. The method was also vulnerable to the frame removal attack.

We propose two techniques that improve the quality of stego video. The first one is to reduce the number of watermarks embedded in a single frame. The proposed method embeds one watermark per frame, while Yokota and Kawamura's method [6] embeds six or more of the same watermark in a frame. In this technique, the frame is divided into blocks. A watermark is embedded in some of the blocks to suppress flickering. Because the extracted watermark contains errors, the same watermark is embedded in multiple frames. By time averaging, a watermark with fewer errors can be achieved. The other technique is to introduce a detection signal to find stego frames. By embedding a detection signal in stego frames, the frames can be detected even under frame removal attacks. In addition, the frames can be selected at arbitrary rather than fixed intervals (e.g., one every three frames). These two techniques can reduce the total amount of information embedded as a watermark in a video. They also eliminate the need for complex BSS processing. The watermark can be extracted by simply taking the difference between two adjacent frames of the video.

To demonstrate the effectiveness of the proposed method, the accuracy of the extracted watermarks and the quality of the stego frames are evaluated by computer simulations. We will also compare the proposed method with the DCT-DWT method proposed by Sang *et al.* [7] in terms of filtering and frame removal attack results. In addition, the video is compressed using MPEG4 and its compression resistance is evaluated.

The organization of this paper is as follows: Section II describes the embedding process of the proposed method, Section III presents the experimental results, and Section IV concludes.

II. Proposed Method

The size of a video frame is $L_w \times L_h$, and the total number of frames is M. m frames that are not adjacent to each other are randomly selected as stego frames. In this section,

we describe the procedure for generating a stego frame by additively embedding a watermark in the μ-th frame. Fig. 1 shows an overview of the proposed method. Using the similarity between consecutive frames, the watermark is extracted by computing the difference between the μ-th and $(\mu-1)$-th frames.

A. Embedding watermark

The main points of the proposed embedding process are as follows.

- The watermark is embedded in P blocks in each frame.
- Eight different patterns ($\ell = 1, 2, 3, \cdots, 8$) for selecting blocks are generated as spreading codes.
- The generated spreading codes are embedded in the frame as a detection signal.
- Both the watermark and the detection signal are embedded using additive embedding.

Let Y^μ be the Y-component of the μ-th frame. Let $\tilde{Y}^\mu$ be the Y-component normalized so that the maximum luminance value Y^μ is 1. In other words, the luminance value at the coordinate (x, y) is normalized by

$$\tilde{Y}^\mu(x, y) = Y^\mu(x, y)/255. \tag{1}$$

The frame $\tilde{Y}^\mu$ is divided into blocks of $h \times h$ pixels. P blocks are randomly selected as embedded blocks. In this paper, the number of combinations of selected blocks is eight different patterns. In other words, by specifying the label of the pattern, the selected blocks can be determined. The label indicating the pattern is also embedded in the frame as a detection signal. The seed value of the random number is kept as the key for extraction.

The watermark $\boldsymbol{w}$ is a signal created by converting a two-dimensional binary logo image of e pixels into a one-dimensional one. That is,

$$\boldsymbol{w} = (w_1, w_2, \cdots, w_e), \tag{2}$$

where $w_i \in \{+1, -1\}$, $i = 1, 2, \cdots, e$. A discrete cosine transform (DCT) is performed on each embedding block. To form the DCT coefficient vector $\boldsymbol{d}^\mu$, the L coefficients are selected from the intermediate frequency components in a zigzag scan order. Note that the number of DCT coefficients, L, and the number of embedded blocks, P, are determined such that $L \times P = e$. That is, the DCT coefficient vector $\boldsymbol{d}^\mu$ at the μ-th frame is given by

$$\boldsymbol{d}^\mu = (d_1^\mu, d_2^\mu, \cdots, d_e^\mu). \tag{3}$$

Let $\tilde{\boldsymbol{d}}^\mu$ be the DCT coefficient vector after the additive embedding process, as given by

$$\tilde{\boldsymbol{d}}^\mu = \left(\tilde{d}_1^\mu, \tilde{d}_2^\mu, \cdots, \tilde{d}_e^\mu\right), \tag{4}$$

where

$$\tilde{d}_i^\mu = d_i^\mu + \alpha w_i,\ i = 1, 2, \cdots, e, \tag{5}$$

where $\alpha > 0$ is the embedding intensity of the watermark. After embedding, an inverse discrete cosine transform (IDCT) is performed on each block. A normalized stego frame $\tilde{Y}_s^\mu$ is generated by combining all blocks, including non-embedded blocks. This frame is then restored to the original 256 levels of luminance to create the stego frame Y_s^μ. The value of Y_s^μ may overflow due to the embedding of the watermark. Therefore, we introduce a boundary process given by

$$\chi = \lfloor 255 \times \tilde{Y}_s^\mu(x, y) \rfloor, \tag{6}$$

$$Y_s^\mu(x, y) = \begin{cases} 255, & 255 < \chi \\ \chi, & 0 \le \chi \le 255 \\ 0, & \chi < 0 \end{cases}, \tag{7}$$

where $\lfloor \cdot \rfloor$ is a floor function.

B. Embedding detection signal

The index $\ell = 1, 2, 3, \cdots, 8$ identifies the blocks selected during watermark embedding. In other words, eight detection signals $\boldsymbol{\xi}^\ell$ are generated as spreading codes. The detection signals are stored as secret keys. For error correction during detection, the length of the detection signal $\boldsymbol{\xi}^\ell$ is k bits as given by

$$\boldsymbol{\xi}^\ell = \left(\xi_1^\ell, \xi_2^\ell, \cdots, \xi_k^\ell\right), \tag{8}$$

where $\xi_j^\ell \in \{+1, -1\}$. The detection signals are generated to satisfy

$$E\left[\boldsymbol{\xi}^\ell \cdot \boldsymbol{\xi}^\lambda\right] = k\ (\ell = \lambda), \tag{9}$$

$$E\left[\boldsymbol{\xi}^\ell \cdot \boldsymbol{\xi}^\lambda\right] \simeq 0\ (\ell \neq \lambda). \tag{10}$$

One of the eight detection symbols is embedded in the V-component $\tilde{V}^\mu$ of the μ-th frame. The V-component frame V^μ is normalized so that it has a maximum size of 1 using (1). The frame $\tilde{V}^\mu$ is divided into blocks of $h \times h$ pixels. On each block, the DCT is performed. The k blocks are selected in raster scan order. A coefficient is selected from the mid-frequency component of the block. This sequence of DCT coefficients is represented by the vector $\boldsymbol{D}^\mu$ given by

$$\boldsymbol{D}^\mu = (D_1^\mu, D_2^\mu, \cdots, D_k^\mu). \tag{11}$$

One of the selected detection signals $\boldsymbol{\xi}^\ell$ is additively embedded into the DCT coefficients, one bit at a time. Let $\widetilde{\boldsymbol{D}}^\mu$ be the embedded DCT coefficients, as given by

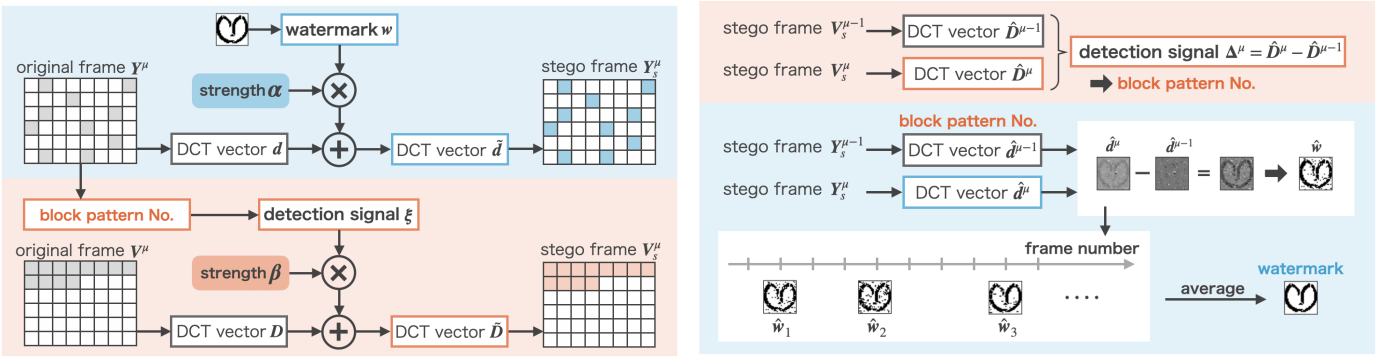
$$\widetilde{\boldsymbol{D}}^\mu = \left(\widetilde{D}_1^\mu, \widetilde{D}_2^\mu, \cdots, \widetilde{D}_k^\mu\right), \tag{12}$$

where

$$\widetilde{D}_j^\mu = D_j^\mu + \beta \xi_j^\ell,\ j = 1, 2, \cdots, k, \tag{13}$$

where $\beta > 0$ is the embedded intensity of the detection signal.

After the embedding process, the IDCT process is performed on each block. A frame $\tilde{V}_s^\mu$ is generated by concatenating all blocks, including non-embedded blocks. Using (6) and (7), the values of the frame $\tilde{V}_s^\mu$ are recovered to their original luminance levels. The stego frame V_s^μ is generated. A color frame is generated by combining the Y, U, and V components with embedded watermarks or detection signals. All frames, including non-embedded frames, are lined up to generate stego video.



(a) Embedding process (b) Extraction process

Fig. 1. Overview of proposed method

C. Extracting detection signal

The main points of the proposed extraction process are as follows.

- As in the process of embedding the detection signal, the V-components of the frames are divided into blocks, the DCT is performed, and the μ-th DCT coefficient vector $\widehat{\boldsymbol{D}}^\mu$ is generated.
- From the time difference of vectors $\widehat{\boldsymbol{D}}^\mu$ and $\widehat{\boldsymbol{D}}^{\mu-1}$, the frame embedding the detection signal is detected.
- The detection signal is extracted from the difference in the DCT coefficients.
- The correlations between the extracted detection signal and the eight original detection signals are calculated, and the label with the largest correlation is found to be the pattern used in that frame.
- From the Y-components of the frames, the watermark is extracted from the blocks indicated by the label patterns.

First, the detection signal is extracted from the video. As in the process of embedding the detection signal, the DCT coefficient vector $\widehat{\boldsymbol{D}}^\mu$ of the μ-th frame V_s^μ is generated, as expressed in

$$\widehat{\boldsymbol{D}}^\mu = \left(\widehat{D}_1^\mu, \widehat{D}_2^\mu, \cdots, \widehat{D}_k^\mu\right). \tag{14}$$

To detect a stego frame from the whole video, the detection signal is first detected. To determine the presence of the detection signal, the difference vector between the μ-th and the $(\mu-1)$-th DCT coefficient vectors is defined as

$$\boldsymbol{\Delta}^\mu = \widehat{\boldsymbol{D}}^\mu - \widehat{\boldsymbol{D}}^{\mu-1}. \tag{15}$$

Then, the intensity of the difference vector is defined as

$$\|\boldsymbol{\Delta}^\mu\|_1 = \sum_{j=1}^{k} \left|\widehat{D}_j^\mu - \widehat{D}_j^{\mu-1}\right|. \tag{16}$$

If the video is not attacked by anything, then $\widehat{\boldsymbol{D}}^\mu = \boldsymbol{D}^\mu$. Assuming that the difference between the μ-th vector and the $(\mu-1)$-th vector is sufficiently small, i.e., $\left|D_j^\mu - D_j^{\mu-1}\right| \simeq 0$, then from (13), the intensity of the difference would be

$$\|\boldsymbol{\Delta}^\mu\|_1 = \sum_{j=1}^{k} |\beta\xi_j| \simeq \beta k. \tag{17}$$

It is determined that the detection signal is embedded in the μ-th frame if the intensity $\|\boldsymbol{\Delta}^\mu\|_1$ is greater than a threshold γ.

Now that the frame with the embedded detection signal is known, the next step is the extraction of the detection signal. By finding the correlation between the difference vector and the eight original detection signals, we can find the detection signal with the highest correlation. This index represents the selected blocks used for embedding. In other words, the detected index $\hat{\ell}$ is obtained by

$$\hat{\ell} = \arg\max_{\ell} E\left[\boldsymbol{\xi}^\ell \cdot \boldsymbol{\Delta}^\mu\right]. \tag{18}$$

Hence, the detection signal is $\boldsymbol{\xi}^{\hat{\ell}}$.

D. Extracting watermark

The detection signal has been extracted. Accordingly, the watermarked frame can be identified as Y_s^μ. The watermark is extracted from the block specified by the detected index $\hat{\ell}$. The Y-components Y_s^μ and $Y_s^{\mu-1}$ of two consecutive stego frames are divided into blocks, respectively. Then, the DCT is performed on each block. The coefficient vector of the DCT is given by

$$\hat{\boldsymbol{d}}^\mu = \left(\hat{d}_1^\mu, \hat{d}_2^\mu, \cdots, \hat{d}_e^\mu\right). \tag{19}$$

We can assume that the coefficient vector $d^{\mu-1} \simeq d^\mu$ holds because the two successive stego frames are similar. From (5), the estimated watermark can be calculated from the difference in the coefficient vectors as

$$\hat{w}_i^\mu = \operatorname{sgn}\left(\hat{d}_i^\mu - \hat{d}_i^{\mu-1}\right). \tag{20}$$

E. Averaged watermark

Although a watermark can be extracted from a single frame, it contains many errors. Since the same watermark is embedded in multiple frames, multiple watermarks extracted from a video can be averaged to generate an averaged watermark. Let Ω

TABLE I
EVALUATION VIDEOS (CIF FORMAT)

Video	No. of frames, M	Frame size [px]
football_a	360	352×288
tempete	260	352×288

Fig. 2. Example of logo image used as watermark (32×32 pixels)

be the set of frame numbers that have been detected as stego frames. Hence, the averaged watermark is given by

$$\bar{w}_i = \mathrm{sgn}\left(\sum_{\mu\in\Omega} \hat{w}_i^{\mu}\right),\ i = 1, 2, \cdots, e. \tag{21}$$

III. EXPERIMENT

A. Experimental conditions

As in the DCT-DWT method [7], the two evaluation videos [9] shown in Table I were used for comparison with the DCT-DWT method [7]. We also evaluated four other videos. All of them showed similar results. Only the two results reported in [7] were shown for comparison. The original videos were in the CIF format with 352×288 pixels. These videos were compressed to the MPEG4 format in lossless mode using FFmpeg [10]. The total number of frames, M, varied from video to video, but the number of frames used for embedding was set to $m = 60$ frames. Fig. 2 shows an example of a logo image used as a watermark. Its size was $e = 32 \times 32$ pixels. Each frame was divided into blocks with a size of 8×8 pixels ($h = 8$). The DCT was performed on the $P = 128$ blocks, and $L = 8$ coefficients were selected from each block. The watermark was embedded into the DCT coefficient vector.

The performance of the proposed method was evaluated in terms of the normalized correlation (NC) of the estimated logo image. The NC between two images I^1 and I^2 is defined by

$$\mathrm{NC} = \frac{\sum_{i=1}^{e}\left(I_i^1 - \bar{I}^1\right)\left(I_i^2 - \bar{I}^2\right)}{\sqrt{\sum_{i=1}^{e}\left(I_i^1 - \bar{I}^1\right)^2}\sqrt{\sum_{i=1}^{e}\left(I_i^2 - \bar{I}^2\right)^2}}, \tag{22}$$

where $\bar{I}^1$ and $\bar{I}^2$ are the average pixel values of images I^1 and I^2, respectively. In addition, the stego video has to be generated in such a way that the watermark can be extracted and the degradation of the video quality is minimized. The image quality of a stego frame was evaluated in terms of the peak signal-to-noise ratio (PSNR) given by

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{255^2}{\mathrm{MSE}}\right)[\mathrm{dB}], \tag{23}$$

where MSE is the mean squared error between two color frames $F^1(x, y)$ and $F^2(x, y)$, given by

$$\mathrm{MSE} = \frac{1}{3L_hL_w}\sum_{x=1}^{L_h}\sum_{y=1}^{L_w}\sum_{c\in\Omega_c}\left\{F_c^1(x, y) - F_c^2(x, y)\right\}^2, \tag{24}$$

where Ω_c is the set of color channels $\{Y, U, V\}$. Since the PSNR of non-embedded frames diverges, the quality of a stego video is expressed as the average PSNR of stego frames excluding non-embedded frames.

B. Evaluating robustness against attacks

The robustness of the stego videos when using the proposed method was evaluated. The bit length of the detection signal was $k = 16$ bits. The embedding strength and threshold for the detection signal were set to $\beta = 0.2$ and $\gamma = 2.0$, respectively. The embedding strength for the watermark was set to $\alpha = 0.005, 0.010, 0.015, \cdots, 0.050$. The generated stego videos were saved in a lossless format. Then, attacks, i.e., Poisson filter, salt and pepper noise (noise density: 2.0%), Gaussian noise (SD $\sigma = 0.05$), sharpening, cutting frames (5 frames), cutting frames (10 frames) were applied to the stego videos. For comparison with the DCT-DWT method [7], the same attacks were used for evaluation. The cutting frames attack implies the frame removal attack. Five or ten frames were randomly removed.

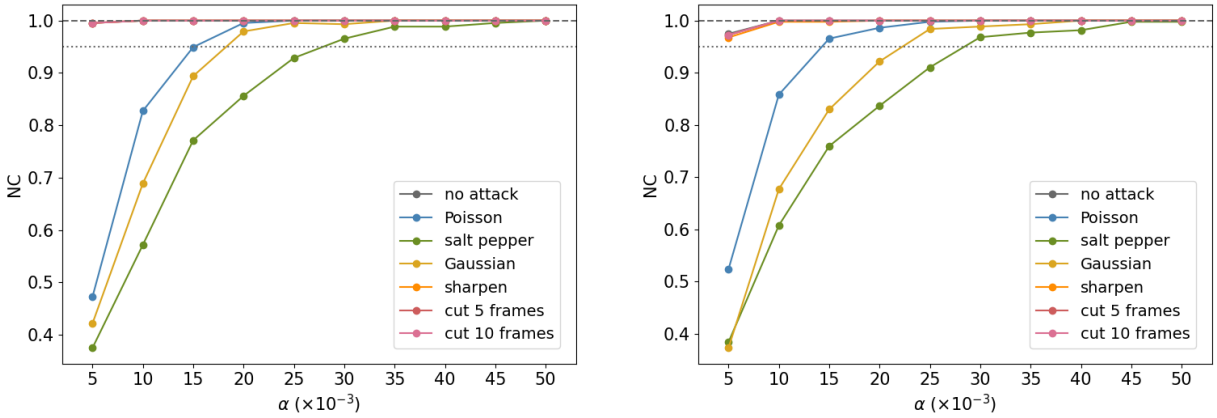
An averaged watermark $\bar{\boldsymbol{w}}$ was extracted from the attacked videos. Its NC value was evaluated. Fig. 3 shows the NC value for the embedding strength α. (a) and (b) represent the NC for football_a and tempete, respectively. The attacks are represented by each of the colors shown in the legend. NC values of 0.95 and 1.0 are represented by the dotted and dashed lines, respectively. For both videos, when $\alpha = 0.030$ or higher, the watermark was extracted with an NC value of 0.95 or higher. With respect to the image quality of the frame, the minimum embedding strength α with an NC greater than or equal to 0.95 was defined as the optimal value. Accordingly, the optimal embedding strength was $\alpha = 0.030$.

The NC values of the proposed method and the DCT-DWT method [7] were compared for two evaluation videos, football_a and tempete. The proposed method was evaluated on these stego videos with the optimal embedding strength $\alpha = 0.030$. Table II shows the NC values of the averaged watermark $\bar{\boldsymbol{w}}$ of both methods. The NC values of the proposed method were equal to or better than those of the DCT-DWT method.

Next, the average PSNR of the stego frames for the proposed method and the DCT-DWT method are shown in Table III. We found that the image quality of the proposed method was better than that of the DCT-DWT method. Considering the fact that the DCT-DWT method is a non-blind method [7] and looking at the results of the NC values and the PSNR, the proposed method is superior to the DCT-DWT method.

C. Evaluating compression resistance

The compression resistance of the proposed method was evaluated. The bit length of the detection signal was set to



(a) football_a (b) tempete

Fig. 3. NC of average watermark under various attacks

TABLE II
NC VALUES OF AVERAGED WATERMARK

	football_a		tempete	
	Proposed Method	DCT-DWT Method	Proposed Method	DCT-DWT Method
non-attack	1.0000	0.9982	1.0000	0.9985
Poisson filter	1.0000	0.9780	1.0000	0.9843
salt-pepper filter (0.02)	0.9654	0.9655	0.9681	0.9644
Gaussian filter (0.05)	0.9931	0.9333	0.9885	0.9258
sharpening	1.0000	0.9985	1.0000	0.9978
cutting frames (5)	1.0000	0.9670	1.0000	0.9663
cutting frames (10)	1.0000	0.8951	1.0000	0.8881

TABLE III
COMPARISON OF AVERAGE PSNR OF STEGO FRAMES AT STRENGTH $\alpha = 0.030$.

Video	Method	PSNR [dB]
foorball_a	Proposed Method	46.5816
	DCT-DWT Method	39.3984
tempete	Proposed Method	46.8538
	DCT-DWT Method	39.6583

$k = 44$ bits because it is difficult to find the detection signal due to compression. The embedding strength β and the threshold γ of the detection signal were set to $\beta = 0.3$ and $\gamma = 2.5$, respectively. The embedding strength for the watermark was set to $\alpha = 0.01, 0.02, \cdots, 0.10$. To compress the videos at a constant quality, a constant rate factor of $\mathrm{crf} = 18$ was specified during compression.

As shown in Fig. 4, we evaluated the NC value of the averaged watermark $\bar{\boldsymbol{w}}$. The horizontal axis is the embedding strength of the watermark, α. The vertical axis is the NC after compression with $\mathrm{crf} = 18$. The watermark was extracted almost correctly when $\alpha \geq 0.05$. Next, the quality of the stego videos was evaluated. The optimal embedding strength for the compression attack was $\alpha = 0.05$. Since the PSNR is more than 41 dB, we can see that the image quality of the proposed method is good.

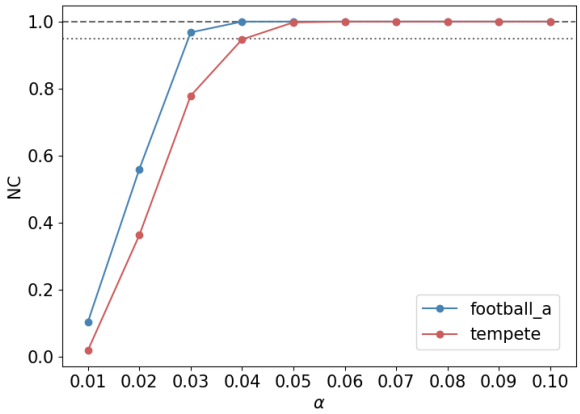


Fig. 4. NC value of averaged watermark after compression

TABLE IV
AVERAGE PSNR OF STEGO FRAMES AT STRENGTH $\alpha = 0.05$.

Video	PSNR[dB]
football_a	41.3555
tempete	42.0173

IV. CONCLUSION

In video watermarking, the watermark should be embedded at high speed because of the large amount of data. Thus, an additional embedding is desirable. However, there has been no

blind decoding method other than the method using BSS [6]. A blind extraction method using BSS was proposed by Yokota and Kawamura [6]. Their method was vulnerable to removal attacks because it required embedding watermarks at fixed intervals.

In this paper, we proposed a method that can perform blind decoding while using additive embedding. Two techniques were introduced: distributed watermark embedding and detection signal embedding. Stego frames are divided into blocks, and watermarks are embedded in a distributed manner to improve the quality of a stego video. The same watermark is embedded in several frames. An averaged watermark is created by averaging the watermarks extracted from these frames. Stego frames are selected by using eight different patterns. Each of the selection patterns is represented by an index. The index is encoded by a spread spectrum code and embedded as a detection signal. Stego frames can be easily detected with the detection signal. Resistance to frame removal attacks is also provided. As a result of implementing these techniques, it was found that the average PSNR of stego frames could be higher than 41 dB, even when using a large embedding strength ($\alpha = 0.05$) for compression tolerance. Furthermore, by using the average of the extracted watermarks, a watermark with small errors could be extracted.

We compared the proposed method with the DCT-DWT method [7]. Comparing the two methods in terms of NC values against non-geometric attacks, the NC values of the proposed method were larger than those of the DCT-DWT method. We also evaluated the compression resistance of the proposed method. When the embedding was $\alpha = 0.05$ or more, an NC value over 0.95 could be achieved.

V. Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP20K11973 and JP24K15106.

References

[1] A. Sverdlov, S. Dexter, A. M. Eskicioglu, "Robust DCT-SVD domain image watermarking for copyright protection: Embedding data in all frequencies," 2005 13th European Signal Processing Conference, Antalya, Turkey, pp. 1-4, 2005

[2] X. Zhang, S. Wang, "Statistical Fragile Watermarking Capable of Locating Individual Tampered Pixels," IEEE Signal Processing Letters, vol. 14, no. 10, pp. 727-730, 2007

[3] C. Pradhan, S. Rath, A. K. Bisoi, "Non Blind Digital Watermarking Technique Using DWT and Cross Chaos," Procedia Technology, vol. 6, pp. 897-904, 2012

[4] I. Hong, I. Kim, S. S. Han, "A blind watermarking technique using wavelet transform," 2001 IEEE International Symposium on Industrial Electronics Proceedings, Korea, vol. 3, pp. 1946-1950, 2001

[5] P. T. Yu, H. H. Tsai, J. S. Lin, "Digital watermarking based on neural networks for color images," Signal Processing, Vol. 81, Issue 3, pp. 663-671, 2001

[6] A. Yokota, M. Kawamura, "BSS-Based Extraction For Additive Video Watermarking," APSIPA ASC 2021, pp. 1640-1646, 2021

[7] J. Sang, Q. Liu, C. L. Song," Robust video watermarking using a hybrid DCT-DWT approach," Journal of Electronic Science Technology vol. 18, issue 2, 2020

[8] N. I. Yassin, N. M. Salem, M. I. El Adawy, "QIM blind video watermarking scheme based on Wavelet transform and principal component analysis," Alexandria Engineering Journal, vol. 53, Issue 4, pp. 833-842, 2014

[9] Xiph.org Video Test Media, https://media.xiph.org/video/derf/, (Viewed 2024/5/20)

[10] FFmpeg, https://ffmpeg.org/, (Viewed 2023/5/15)