BSS-BASED EXTRACTION FOR ADDITIVE VIDEO WATERMARKING

Akane Yokota and Masaki Kawamura
Yamaguchi University, Yamaguchi, Japan
E-mail: kawamura@sci.yamaguchi-u.ac.jp Tel: +81-83-933-5701

***Abstract*—We propose a method for extracting watermarks additively embedded in video frames by using blind source separation (BSS). A video consists of a sequence of multiple still images, and similarity between successive frames is high. By using the similarity between frames and the BSS technique, we show that the proposed method can perform blind decoding. After separating the original signal and watermark from the stego-frames by BSS, it is necessary to identify the frame containing the watermark from the separated signals. We propose a method to identify watermarked frames on the basis of correlation. We embedded a binary image, or logo, into frames as a watermark. Since the embedding strength β affects the quality of the video and bit error rate (BER) of the estimated image, the peak signal-to-noise ratio (PSNR) for stego-frames and BER for the image were evaluated. In the case where the strength was low, the PSNR was over 47 dB, but the BER was 0.3 at maximum. For a high strength, the PSNR was 30 to 42 dB and the BER was less than 0.1. As a result, an embedding strength of $\beta = 0.02$ was recommended, which achieved $\mathrm{PSNR} > 42$ dB and $\mathrm{BER} < 0.1$.**

I. Introduction

In recent years, many people enjoy watching movies and videos. In addition, video posting sites such as YouTube and social networking services enable people to post their videos easily. However, piracy has also occurred, such as illegally acquiring videos by using a screen recording function and posting copied videos. Watermarking methods [1]–[9] can be used to control the unauthorized use of videos. The watermarking method [1] is a technique for embedding secret information into digital contents such as still images, videos, audios, etc. It can be used to protect copyrights and detect image tampering. The method needs to be able to retrieve the embedded watermark even from contents that have been subjected to attacks such as compression and scaling. At the same time, it is necessary to be able to embed a sufficient number of watermarks while maintaining the quality of the contents.

In this work, we consider a digital watermarking method for videos. A video consists of a series of still images, called frames. For each block in a frame, motion vectors are calculated from the differences between frames. There are several methods of video watermarking, including those using motion vectors [2], [3] and video streams [4]. Moreover, the watermarking method for still images can be directly applied to embed the watermark in the frames [5]–[7].

The watermarking methods that modify the motion vectors [2], [3] may cause unnaturalness since they change the number of object movements in frames. Furthermore, the algorithms of embedding depend on a coding standard, e.g., MPEG-4 and H.264. For streaming, the embedding process needs to be fast. Since decompression and compression take a long time to process, the embedding process is performed in the compressed domain [4].

Watermarks are often embedded in the discrete Wavelet transform (DWT) domain. To achieve blind watermarking, the watermark is quantized and embedded in the low-low (LL) subband of the DWT coefficients [5]. Quantization is effective against compression, but it causes a degradation in quality. Additive embedding methods are weak against compression, but the embedding strength can be easily controlled. In [6], each frame is divided into 8×8 blocks, and the embedding position is determined by predicting the motion vector of the blocks. The watermark is additively embedded into the frame after the second-level DWT is applied. In [7], the DWT is performed for each frame. The LL and high-high (HH) subbands are divided into blocks and principal component analysis (PCA) is performed on each block. The watermark is embedded additively in the coefficients of the PCA. These methods require the original frames when extracting the watermark. In general, blind watermarking that requires no original frames is preferred.

In the case that the watermark is embedded additively, the original signal is usually required for decoding. We propose a blind watermarking method using the blind source separation (BSS) technique. BSS is a method of separating mixed signals into their original sources without using any information about the signal sources [10], [11]. To separate the mixed signals, more observation signals are needed than the number of signal sources. Although image watermarking using BSS has been proposed [8], there is only one stego-image while there are two source signals, the original image and watermark. Therefore, in addition to the stego-image, an auxiliary image is necessary. In video watermarking, considering the video frame and watermark as two signal sources, the stego-frame can be regarded as a mixed signal. By using BSS, it may be possible to separate the two signal sources from the stego-frame. Two stego-frames are needed to separate them. However, for videos, several frames can be used.

Since the watermark is embedded additively, the embedding method is not an issue. In this paper, we will show that BSS can be used to blindly extract watermarks from stego-frames. The issue is to determine which of the signals separated by

BSS contains the watermark signal. We propose a method to identify the watermark signal by using correlation.

This paper is organized as follows. In Section II, we explain the principle of BSS. In Section III, we describe the proposed method. In Section IV, we show results obtained by computer simulations, and we conclude our study in Section V.

II. Blind Source Separation

BSS is a method of estimating the original independent signal sources from observed signals generated by a mixture of independent sources. The BSS methods include the PCA and the independent component analysis (ICA) [10]. PCA is a method of estimating the original sources under the assumption that each component of the sources is uncorrelated, whereas ICA is a method of estimating the original sources under the assumption that each component of the original sources is independent. For signal separation using BSS, the number of observations must be greater than or equal to the number of sources. Since original videos and watermarks can be regarded as independent sources, the proposed method applies ICA, especially FastICA [10].

A. *Mathematical framework of ICA*

We assume that M statistically independent signal sources $\boldsymbol{s}(t) = (s_1(t), s_2(t), \cdots, s_M(t))^\top$ are mixed, and then N mixed signals $\boldsymbol{x}(t) = (x_1(t), x_2(t), \cdots, x_N(t))^\top$ are observed, that is,

$$\boldsymbol{x}(t) = A\boldsymbol{s}(t), \tag{1}$$

where A is a mixing matrix and $^\top$ denotes the transpose of a vector. The goal of BSS is to estimate the unknown signal source $\boldsymbol{s}(t)$ and the unknown mixing matrix A from only mixed signals $\boldsymbol{x}(t)$.

Assume that the separated signals $\boldsymbol{u}(t) = (u_1(t), u_2(t), \cdots, u_M(t))^\top$, which are the estimation of the sources $\boldsymbol{s}(t)$, are given by

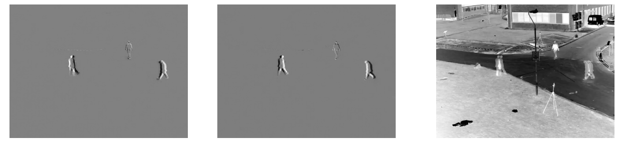
$$\boldsymbol{u}(t) = W\boldsymbol{x}(t), \tag{2}$$

where $W = [\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_N]$ is the separation matrix with $\boldsymbol{w}_i = (w_{i1}, w_{i2}, \cdots, w_{iN})^\top$. FastICA [10] and the natural gradient method [11] have been proposed for the estimation of the separation matrix W. The number of observations must be greater than or equal to the number of sources.

B. *BSS for still image*

A stego-image can be regarded as a mixture of the original image and the watermark. To separate the stego-image into the original image and watermark, at least two observation signals, i.e., two stego-images, are required. Since it is not possible to separate them with only one stego-image, methods using auxiliary images have been proposed. Nguyen et al. [8] proposed a method using a public key image as an auxiliary image. The key image is generated from the original and random images, and third parties can obtain it. Since two images are obtained, BSS can be applied for separation. However, it is not a blind watermarking method since it uses an image created from the original image. Thus, the use of BSS for image watermarking is inherently difficult.

(a) 1st frame (b) 2nd frame (c) 3rd frame



(d) Differential Frame 1 (e) Differential Frame 2 (f) Dominant frame

Fig. 1. Original frames (a)–(c) and separated frames by ICA (d)–(f)

C. *ICA for video frames*

In video watermarking, the number of observed signals is also an issue. Malik et al. [9] proposed a detection method for spread-spectrum (SS)-based watermarking. They found that the host signal and watermark obey non-Gaussian distributions in the case of SS-based watermarking. Therefore, a mixed signal could be separated by using a noisy ICA model.

Here, we consider watermarking without the SS technique. When successive frames are extracted from a video, the extracted frames have a high similarity. Therefore, the assumption that the frames are independent signal sources does not hold. In other words, the number of effective sources may be reduced. The three successive original frames and results of applying ICA are shown in Fig. 1. As a result, it was found that one dominant frame and two differential frames of the original video were extracted. We could determine that the differential frames were the difference between the second and first original frames and that between the third and second original frames. Due to the differential frames having low information than the original ones, we believe that even if a watermark is embedded in one of the three frames, it can be separated by ICA.

III. Proposed Method

In the proposed method, stego-frames are generated by embedding watermarks in the discrete cosine transform (DCT) domain. The watermark is additively embedded in one of the three frames. When extracting the watermark, ICA is applied to the three stego-frames to separate the watermark from the original frames.

A. *Embedding Process*

I-frames are extracted from all video frames and divided every three frames. Since the same processes are performed, we will focus on three consecutive frames, F_1, F_2, F_3. The size of the frame is $L_w \times L_h$ pixel. The stego-frame S_1 is generated by embedding a watermark in the DCT coefficients of the first original frame F_1. The flow of the embedding

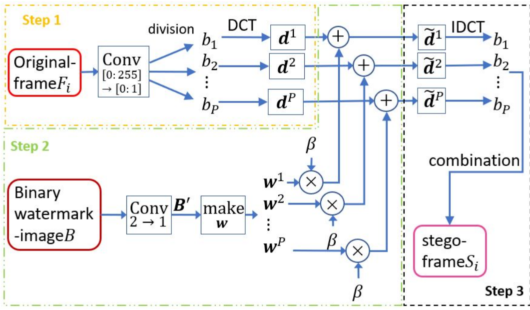


Fig. 2. Embedding process

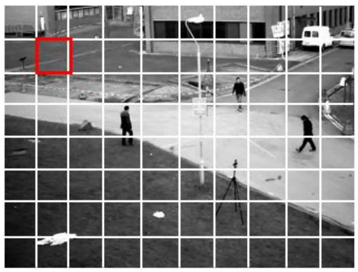


Fig. 3. Blocks in a frame

process is shown in Fig. 2. In Step 1, to equalize the intensity of a watermark image with that of the original frame, the luminance of the first original frame, F_1, is normalized to a real number of [0:1]. That is,

$$\tilde{F}_1\left(m,n\right) = F_1\left(m,n\right)/255, \tag{3}$$

where $m = 1,2,\cdots,L_w,\ \ n = 1,2,\cdots,L_h$. The original frame $\tilde{F}_i$ is divided into P blocks $b_1, b_2, \cdots, b_P$ of $u \times u$ pixels as shown in Fig. 3, and DCT is performed on each divided block, where the number of blocks is $P = \left(L_w \times L_h\right)/\left(u \times u\right)$.

From each block, e coefficients are selected from the intermediate frequency components of the DCT coefficients in accordance with the zigzag scan order. From the first original frame F_1, $e \times P$ DCT coefficients are extracted. Let the DCT coefficient vector of the μ-th block in the first frame be $\boldsymbol{d}^\mu,\ \mu = 1,2,\cdots,P$, which has e elements and is given by

$$\boldsymbol{d}^\mu \quad = \quad \left(d^\mu(1), d^\mu(2), \cdots, d^\mu(e)\right). \tag{4}$$

In Step 2, the binary image or logo $B(k,l)$ with $Le_w \times Le_h$ pixels is converted to a one-dimensional logo signal $\boldsymbol{B}' = (B'(1), B'(2), \cdots, B'(Le_w \times Le_h))$, where $B(k,l) \in \{-1,+1\},\ k = 1,2,\cdots,Le_w,\ l = 1,2,\cdots,Le_h$. The logo signal $\boldsymbol{B}'$ is repeatedly concatenated M times so that its size is the same as $e \times P$ pixels. We call the concatenated signal a watermark signal $\boldsymbol{w}$, which is given by

$$\boldsymbol{w} = \underbrace{\left(\boldsymbol{B}', \boldsymbol{B}', \cdots, \boldsymbol{B}'\right)}_{M\text{ times}}. \tag{5}$$

The watermark signal can be again divided into P blocks with e elements to embed it into each block.

$$\boldsymbol{w} \quad = \quad \left(\boldsymbol{w}^1, \boldsymbol{w}^2, \cdots, \boldsymbol{w}^P\right), \tag{6}$$

where

$$\boldsymbol{w}^\mu \quad = \quad \left(w^\mu(1), w^\mu(2), \cdots, w^\mu(e)\right). \tag{7}$$

The size of $\boldsymbol{w}^\mu$ is the same as that of $\boldsymbol{d}^\mu$. Each block of the watermark signal, $\boldsymbol{w}^\mu$, is additively embedded into the DCT coefficients $\boldsymbol{d}^\mu$ of the μ-th block of the first frame. The DCT coefficient $\tilde{\boldsymbol{d}}^\mu$ of the stego-frame is given by

$$\tilde{d}^\mu(x) \quad = \quad d^\mu(x) + \beta w^\mu(x),\ x = 1,2,\cdots,e, \tag{8}$$

where $\beta \geq 0$ is an embedding strength.

In Step 3, inverse discrete cosine transform (IDCT) is performed on each divided block. The blocks are combined to generate the stego-frame S_1 of $L_w \times L_h$ pixels.

B. Proposed extraction method

Since the watermark is embedded in every three frames, the watermark can be extracted using any three consecutive frames of the stego-video. BSS is applied to extract the watermark blindly. The separated signals are independent signals, but it is not clear which one contains the watermark.

1) preprocess: Figure 4 shows the processing flow, which includes the preprocess and ICA. In Step 1, preprocess is performed before ICA. Three frames of the stego-video, $S_i,\ i = 1,2,3$, are divided into P blocks $b_1, b_2, \cdots, b_P$ of $u \times u$ pixels, and DCT is performed on each divided block. From each block, e intermediate frequencies of the DCT coefficients are extracted in a zigzag scan order. Low frequencies may affect the image quality and high frequencies are vulnerable to compression, so intermediate frequencies are used. $e \times P$ DCT coefficients are extracted from each of the three stego-frames $S_i\ i = 1,2,3$. Let the DCT coefficient vector of the μ-th block in the i-th frame be $\boldsymbol{D}_i^\mu,\ i = 1,2,3,\ \mu = 1,2,\cdots,P$, and the DCT coefficient vector of all blocks in the i-th frame be

$$\boldsymbol{D}_i \quad = \quad \left(\boldsymbol{D}_i^1, \boldsymbol{D}_i^2, \cdots, \boldsymbol{D}_i^P\right),\ i = 1,2,3. \tag{9}$$

2) separation process: In Step 2, ICA is applied to the vector $\boldsymbol{D}_i\ i = 1,2,3$. Independent separated signals $\boldsymbol{y}_i,\ i = 1,2,3$ are obtained by applying ICA. As discussed in Section II, one of the signals is the dominant frame and the others are the difference frames, which may contain the watermark.

3) identification process: Figure 5 shows the processing flow, which includes the identification and reconstruction processes of a watermark image. Note that although images are shown as frames in Figure 5, they are actually signals after DCT. In Step 3, the dominant frame $\boldsymbol{f}$ is identified by calculating the cross-correlations of all combinations between the separated signals $\boldsymbol{y}_i,\ \ i = 1,2,3$ and the stego-signals $\boldsymbol{D}_j,\ j = 1,2,3$. If a separation signal $\boldsymbol{y}_i$ was the dominant frame $\boldsymbol{f}$, the cross-correlation $r\left(\boldsymbol{y}_i, \boldsymbol{D}_j\right)$ between $\boldsymbol{y}_i$ and the stego-signals $\boldsymbol{D}_j$ would be close to 1.0. Therefore, the sum of the cross-correlations $r\left(\boldsymbol{y}_i, \boldsymbol{D}_j\right)$ is used as an estimator

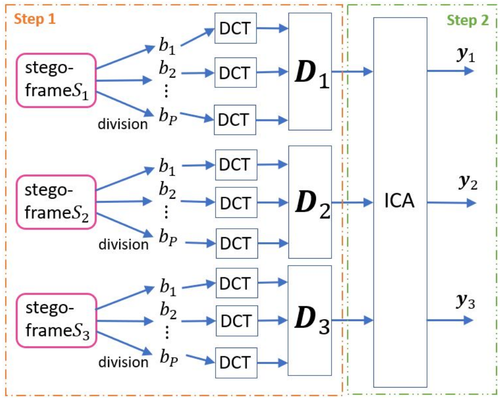
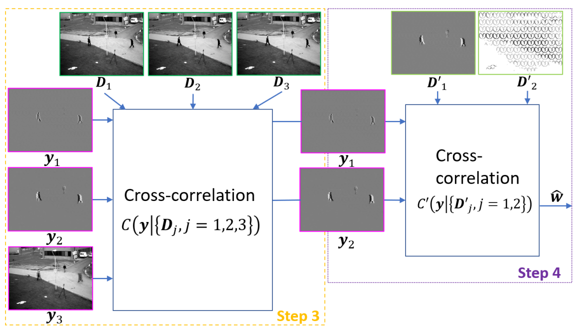
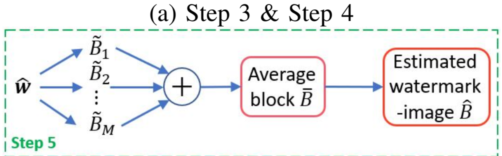


Fig. 4. Separation process



(a) Step 3 & Step 4



(b) Step 5

Fig. 5. Identification process

$C(\boldsymbol{y}_i | \{\boldsymbol{D}_j, j = 1,2,3\})$ to estimate the dominant frame $\boldsymbol{f}$. That is, the estimated dominant frame $\hat{\boldsymbol{f}}$ is given by

$$\hat{\boldsymbol{f}} = \arg\max_{\boldsymbol{y} \in Y} C(\boldsymbol{y} | \{\boldsymbol{D}_j, j = 1,2,3\}), \tag{10}$$

where Y is a set of the separation signals, which is given by $Y = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3\}$, and

$$C(\boldsymbol{y} | \{\boldsymbol{D}_j, j = 1,2,3\}) = \sum_{j=1}^{3} |r(\boldsymbol{y}, \boldsymbol{D}_j)|. \tag{11}$$

Note that since the separated signal $\boldsymbol{y}_i$ may be a sign-reversed version of the original frame, the operation of the absolute value is introduced.

In Step 4, the dominant frame with the watermark is identified. The signals $\boldsymbol{y}_i$ excluding the dominant frame are difference frames as described in Section II. The watermark is included in these difference frames. Therefore, it is not possible to find the watermarked frame by using the correlation (11) with the original frame. To find the watermarked frame, the cross-correlations with the difference frames are used instead of the original frame itself. Let the difference frame be $\boldsymbol{D}'_j$, which is given by

$$\boldsymbol{D}'_j = \boldsymbol{D}_{j+1} - \boldsymbol{D}_j, \; j = 1,2. \tag{12}$$

The cross-correlation between the separated signal $\boldsymbol{y}_i$ and the difference frame $\boldsymbol{D}'_j$ is represented as $r(\boldsymbol{y}_i, \boldsymbol{D}'_j)$. Since the watermark signal tends to appear in the difference frames, the frame with the largest correlation is chosen as a watermark frame. Therefore, the estimated watermarked frame $\hat{\boldsymbol{w}}$ is given by

$$\hat{\boldsymbol{w}} = \arg\max_{\boldsymbol{y} \in Y \backslash \boldsymbol{f}} C'(\boldsymbol{y} | \{\boldsymbol{D}'_j, j = 1,2\}), \tag{13}$$

where $Y \backslash \boldsymbol{f}$ denotes the difference set of the set Y excluding the dominant frame $\boldsymbol{f}$, and

$$C'(\boldsymbol{y} | \{\boldsymbol{D}'_j, j = 1,2\}) = \sum_{j=1}^{2} |r(\boldsymbol{y}, \boldsymbol{D}'_j)|. \tag{14}$$

In Step 5, the logo can be extracted. Since the size of the logo is $Le_w \times Le_h$ pixels, the watermark signal $\hat{\boldsymbol{w}}$, which is the one-dimensional vector of $e \times P = Le_w \times Le_h \times M$ elements, is divided into M blocks with $Le_w \times Le_h$ pixels, and m-th block is denoted as $\tilde{B}_m(k,l)$ in a two-dimensional form. Since there are M blocks, the average block is given by

$$\bar{B}(k,l) = \frac{1}{M} \sum_{m=1}^{M} \tilde{B}_m(k,l). \tag{15}$$

The value of the average block $\bar{B}(k,l)$ is real number. The estimated logo $\hat{B}(k,l)$ is given as a binarized block of the average block, that is,

$$\hat{B}(k,l) = \begin{cases} 1, & \bar{B}(k,l) \geq \theta \\ 0, & \bar{B}(k,l) < \theta \end{cases}, \tag{16}$$

where θ is the threshold calculated by the discriminant analysis method [12].

IV. Computer Simulation

We additively embed a logo as a watermark in one of every three video frames and show that watermarks can be extracted using BSS technique. The image quality of the stego-frames and the bit error rate (BER) of the logo are evaluated.

A. Experimental conditions

The logo is a binary image of 64×64 pixels ($Le_w = 64, Le_h = 64$), as shown in Fig. 8 (a). It is converted to a watermark as described in Section III-A.

The watermark is embedded into one of every three video frames. The total frame number is 100. The first three frames are shown in Fig. 1. The size of the frame is $L_w = 768 \times L_h = 576$ pixels. Each frame is divided into blocks of 8×8 pixels ($u = 8$) and transformed by DCT. For each block, a watermark of $e = 4$ bits is embedded in the DCT domain. The e bits are selected as the 12th to 15th DCT coefficients in a zigzag scan order.

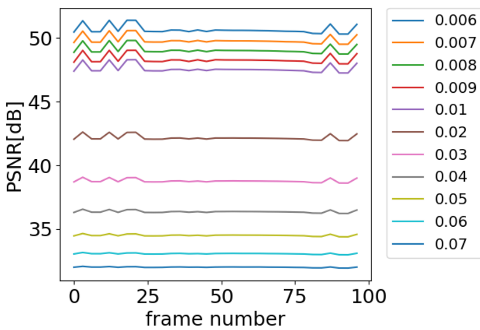


Fig. 6. PSNRs for different embedding strengths.

The image quality of a stego-frame S_i is evaluated by the peak signal-to-noise ratio (PSNR). The estimated logo $\hat{B}(k,l)$ is evaluated by BER. The BER of the estimated logo $\hat{B}(k,l)$ is defined by

$$\mathrm{BER} = \frac{1}{Le_w \times Le_h} \sum_{k=1}^{Le_w} \sum_{l=1}^{Le_h} B(k,l) \oplus \hat{B}(k,l), \quad (17)$$

where $\oplus$ denotes the exclusive-OR.

B. Evaluation of image quality of stego-frames

The image quality of the stego-frame S_i was evaluated by PSNR. The PSNR depends on the embedding strength β in (8). Therefore, the PSNRs for different strengths β were calculated. We examined two types of strength; $\beta = 0.006, 0.007, 0.008, 0.009, 0.010$ as the low one, and $\beta = 0.020, 0.030, 0.040, 0.050, 0.060, 0.070$ as the high one. Figure 6 shows the PSNR at each stego-frame for these embedding strengths. The abscissa represents the frame number and the ordinate represents PSNR $[\mathrm{dB}]$. The PSNRs were calculated every three frames, which were the stego-frames. All PSNRs were over 30 dB. In particular, for the low strength case, it was more than 47 dB. In contrast, even with the highest embedding strength $\beta = 0.070$, the stego-frames did not seem to be significantly degraded. From Fig. 6, we found that the PSNR for each strength β was almost flat and did not change much depending on the position of the stego-frame. The generated stego-frames for $\beta = 0.010, 0.070$ are shown in Fig. 7. Due to the embedding in the DCT domain, there is less degradation in these frames.

C. Evaluation of watermark estimation

Watermarks were embedded in one of the three frames. Even when three frames are extracted from an arbitrary frame number, we can show that watermarks can be extracted by ICA. First, to determine the dominant frame, we calculated the sum of correlations $C(\boldsymbol{y}_i | \{\boldsymbol{D}_j, j = 1,2,3\})$. Table I shows the sum of the correlations, $C(\boldsymbol{y}_i | \{\boldsymbol{D}_j, j = 1,2,3\})$, $i = 1,2,3$ for the frame number $i = 1,2,3$ as an example. From Table I, we found that only one of three takes a large value. This means that it is possible to distinguish between the signal close to the original frame and other signals. Therefore, we could estimate the dominant frame $\hat{\boldsymbol{f}}$ by (10).

(a) $\beta = 0.010, 47.4$ dB (b) $\beta = 0.070, 32.0$ dB

Fig. 7. Example of stego-frames

TABLE I
SUM OF CORRELATION, $C(\boldsymbol{y}_i | \{\boldsymbol{D}_j, j = 1,2,3\})$.

$C(\boldsymbol{y}_1 \| \{\boldsymbol{D}_j\})$	$C(\boldsymbol{y}_2 \| \{\boldsymbol{D}_j\})$	$C(\boldsymbol{y}_3 \| \{\boldsymbol{D}_j\})$
0.83	0.49	2.56

Next, to detect the frame containing the watermark, we calculated the sum of cross-correlations $C'\left(\boldsymbol{y}_i | \{\boldsymbol{D}'_j, j = 1,2\}\right)$ of (14). Table II shows the sum of the correlations, $C'\left(\boldsymbol{y}_i | \{\boldsymbol{D}'_j, j = 1,2\}\right), i = 1,2$ as an example. We found that only one of two takes a large value. This means that it is possible to detect the difference frame containing the watermark. Therefore, we could estimate the estimated watermarked frame by (13). In summary, we found that the proposed method could separate and identify the signals.

D. BER of estimated logo

Since we could determine the watermark frame, we reconstructed the logo and evaluated it by BER. Figure 8 shows the estimated logos reconstructed by the proposed method. (a)–(c) show the original image, the estimated image for the embedding strength $\beta = 0.070$, and the image for $\beta = 0.008$, respectively. The estimated logo may be flipped in black and white. This is because ICA cannot estimate the amplitude strength. Therefore, the inverted image is considered to be identical to the original. The BERs for the embedding strengths $\beta = 0.070$ and 0.008 were 0.002 and 0.042, respectively.

TABLE II
SUM OF CORRELATION, $C'\left(\boldsymbol{y}_i | \left\{\boldsymbol{D}'_j, j = 1,2\right\}\right)$.

$C'\left(\boldsymbol{y}_1 \| \left\{\boldsymbol{D}'_j\right\}\right)$	$C'\left(\boldsymbol{y}_2 \| \left\{\boldsymbol{D}'_j\right\}\right)$
1.12	0.02

(a) Original (b) Estimated image ($\beta = 0.070$) (c) Estimated image ($\beta = 0.008$)

Fig. 8. Original and estimated images [2×], 64×64 pixels

The BERs for embedding strengths are shown in Fig. 9. The abscissa and ordinate represent the frame number and BER, respectively. The BERs were calculated using three consecutive frames. For low embedding strengths as shown in (a), the BERs were almost $\mathrm{BER} < 0.3$. It was difficult to say that images were reconstructed with sufficient accuracy. For high embedding strengths as shown in (b), the BERs were almost $\mathrm{BER} < 0.1$. We found that to extract the logo with a low BER, it is better to use a high embedding strength.

A high embedding strength results in poor image quality, and a low embedding strength results in a large BER. Therefore, multiple extracted logos estimated with a low embedding strength are combined to create a synthesized logo. This approach has the potential to reduce the BER. Since the extracted logo consists of zeros and ones, the synthesized logo can be generated by the majority vote of the extracted logos. However, as previously mentioned, a reversed logo may be estimated as shown in Fig. 8 (c). For a logo with a white background, if it contains more than half the number of black pixels, it is assumed to be an inverted logo, and then black and white pixels are swapped.

The BERs of the synthesized logo were evaluated as shown in Fig. 10. The abscissa represents number of logos to generate the synthesized logo. The ordinate represents BER in a logarithmic scale. (a) and (b) show the BERs for low and high embedding strengths, respectively. With exception to the case of $\beta = 0.006$, the BERs became smaller as the number of logos increased. We found that the BER could be reduced sufficiently if at least 40 logos were synthesized. As a result, even with a low embedding strength, the BER could be reduced to zero.

V. Conclusion

We proposed a method that can extract a watermark in the frames of a video using ICA, which is one of the BSS methods. For videos, since there are many similar frames, the watermark could be separated from the original frame without increasing the number of stego-frames. The issue was to identify in which of the separated signals the watermark was embedded. We calculated the correlations between the separated signals and stego-frames, and then succeeded in identifying the watermarked frame.

The embedding strength affects the image quality and the BER of the estimated logo. Hence, we evaluated them using the PSNR and BER. For low embedding strengths, the PSNR was over 47 dB and the image quality was very high. However, although the BER was not so good for a single frame, it could be reduced to zero by using multiple frames. For high embedding strengths, the PSNR was from 30 to 42 dB, and all BERs were under 0.1. Therefore, the embedding strength β, where the PSNR and BER are both better, should to be selected. We recommend the embedding strength $\beta = 0.02$ for a single frame, which achieves $\mathrm{PSNR} > 42$ dB and $\mathrm{BER} < 0.1$.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 20K11973. The computation was carried out using the super computer facilities at the Research Institute for Information Technology, Kyushu University.

References

[1] F.A. Petitcolas, R.J. Anderson, and M.G. Kuhn, "Information hiding-a survey," Proceedings of the IEEE, vol. 87, no. 7, pp. 1062–1078, 1999

[2] J. Zhang, J. Li, and L. Zhang, "Video watermark technique in motion vector," Proceedings XIV Brazilian Symposium on Computer Graphics and Image Processing, pp. 179–182, 2001

[3] K. Niu, X. Yang, and Y. Zhang, "A novel video reversible data hiding algorithm using motion vector for H.264/AVC," Tsinghua Science and Technology, vol. 22, no. 5, pp. 489–498, 2017

[4] H. Ling, Z. Lu, and F. Zou, "New real-time watermarking algorithm for compressed video in VLC domain," 2004 International Conference on Image Processing, vol. 4, pp. 2171–2174, 2004

[5] H. Inoue, A. Miyazaki, T. Araki, and T. Katsura "A digital watermarking method using the wavelet transform for video data," IEICE trans. Fundamentals, vol. E-83-A, no. 1, pp. 90–96, 2000

[6] J. Hussein and A. Mohammed "Robust video watermarking using multiband wavelet transform," International Journal of Computer Science Issues, vol. 6, no. 1, pp. 44–49, 2009

[7] S.A.K Mostafa, A.S. Tolba, F.M. Abdelkader, and H.M. Elhindy, "Video watermarking scheme based on principal component analysis and wavelet transform," International Journal of Computer Science and Network Security, vol. 9, no. 8, pp. 45–52, 2009

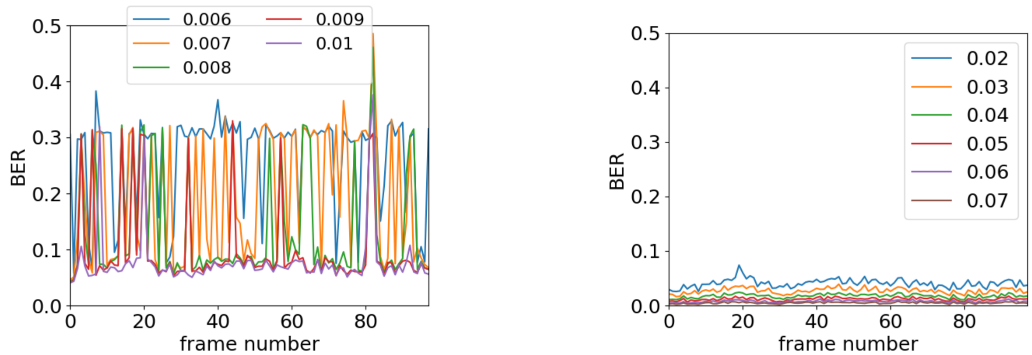
[8] T.V. Nguyen, T. Vlet, and J.C. Patra, "A simple ICA-based digital image watermarking scheme," Digital Signal Processing, vol. 18, no. 5, pp. 762–776, 2008

[9] H. Malik, A. Khokhar, and R. Ansari, "Improved watermark detection for spread-spectrum based watermarking using independent component analysis," DRM '05: Proceedings of the 5th ACM workshop on Digital rights management, pp. 102–111, 2005

[10] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," IEEE transactions on Neural Networks, vol. 10, no. 3, pp. 626–634, 1999

[11] S. Amari, "Natural gradient works efficiently in learning," Neural Computation, vol. 10, no. 2, pp. 251–276, 1998

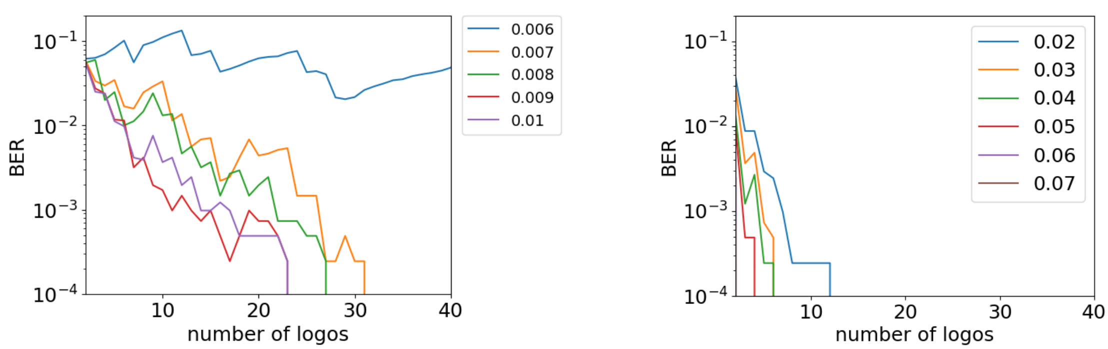
[12] P.K. Sahoo, S. Soltani, A.K.C. Wong, and Y.C. Chen, "A survey of thresholding techniques," Computer Vision, Graphics, and Image Processing, vol. 41, no. 2, pp. 233–260, 1988



(a) BERs for weak case $\beta \leq 0.010$

(b) BERs for strong case $\beta \geq 0.020$

Fig. 9. BERs for different embedding strengths β.



(a) BERs for weak case $\beta \leq 0.010$

(b) BERs for strong case $\beta \geq 0.020$

Fig. 10. BER of synthesized logo