IEICE TRANS. INF. & SYST., VOL.E100–D, NO.1 JANUARY 2017

INVITED PAPER Special Section on Enriched Multimedia —New Technology Trends in Creation, Utilization and Protection of Multimedia Information—

Information Hiding and Its Criteria for Evaluation

Keiichi IWAMURA[†a)], **Masaki KAWAMURA**[††b)], **Minoru KURIBAYASHI**[†††c)], ***Senior Members***,
Motoi IWATA[††††d)], **Hyunho KANG**[†e)], ***Members***, **Seiichi GOHSHI**[†††††f)],
and **Akira NISHIMURA**[††††††g)], ***Senior Members***

SUMMARY Within information hiding technology, digital watermarking is one of the most important technologies for copyright protection of digital content. Many digital watermarking schemes have been proposed in academia. However, these schemes are not used, because they are not practical; one reason for this is that the evaluation criteria are loosely defined. To make the evaluation more concrete and improve the practicality of digital watermarking, watermarking schemes must use common evaluation criteria. To realize such criteria, we organized the Information Hiding and its Criteria for Evaluation (IHC) Committee to create useful, globally accepted evaluation criteria for information hiding technology. The IHC Committee improves their evaluation criteria every year, and holds a competition for digital watermarking based on state-of-the-art evaluation criteria. In this paper, we describe the activities of the IHC Committee and its evaluation criteria for digital watermarking of still images, videos, and audio.

key words: *information hiding, evaluation criteria, digital watermarking, still image, video, audio, IHC committee*

1. Introduction

The 2015 white paper on information and communications in Japan [1] shows that the Japanese content market was valued at roughly 11.3 trillion JPY, and that the market for digital content (downloaded or streamed via the Internet to computers or mobile phones) grew to roughly 2.34 trillion JPY, accounting for 20.8% of the entire content market. The market for digital content is thus growing, while other content markets are not. For example, the Recording Industry Association of Japan (RIAJ) reported that the amount of CD production is continuously dropping since 1998 (587.9 billion JPY) to 2015 (180.1 billion JPY), except for the expansion induced by the AKB48 phenomena in 2012. Hardware or software for capturing is able to make copies of the unprotected media i.e. CD, and also is able to make copies of the protected disc media, DVD and BD. One possible reason for this is that current content protection technology is insufficient.

There are two typical types of content protection technology. The first is cryptography, which enables viewing of and listening to content only by the legally authorized user. The second is information hiding, which enables us to verify the copyright of content by embedding copyright information in the content. Digital watermarking is a representative technology. To protect the copyright of digital content, it is desirable that these two technologies should be combined.

In cryptographic technology, security is carefully evaluated in terms of robustness against typical attacks such as cipher-text attacks, known plain-text attacks, chosen plain-text attacks, and chosen cipher-text attacks. Security of cryptography can be theoretically or mathematically proven in many cases. In addition, current cryptographic technologies use open algorithms, allowing evaluation by third parties.

In contrast, information hiding technology is not evaluated using common evaluation criteria, and security can not be proven because (in many cases) the basis of security is the non-disclosure of the algorithm. Therefore, information hiding technology is not discussed in terms of security in open community, and the evaluation criteria for security are loosely defined. In addition, research into evaluation criteria is scarce. As a result, academic digital watermarking sees little practical use.

In this situation, we have tried to establish useful and globally applicable evaluation criteria for information hiding technology. Trials have been started by the Information Hiding and its Criteria for Evaluation (IHC) Committee in workshops presented by the Institute of Electronics, Information and Communication Engineers (IEICE), establishing new evaluation criteria every year, inviting presentations of digital watermarking methods exceeding the evaluation criteria, and holding competitions for watermarking. In this paper, we describe the activities of the IHC Committee and the evaluation criteria for digital watermarking for still images, videos, and audio as decided by the IHC Committee.

The IHC Committee recently hosted two international

Manuscript received March 28, 2016.
Manuscript revised August 2, 2016.
Manuscript publicized October 7, 2016.
[†]The authors are with Tokyo University of Science, Tokyo, 125–8585 Japan.
[††]The author is with Yamaguchi University, Yamaguchi-shi, 753–8512 Japan.
[†††]The author is with Okayama University, Okayama-shi, 700–8530 Japan.
[††††]The author is with Osaka Prefecture University, Sakai-shi, 599–8531 Japan.
[†††††]The author is with Kogakuin University, Tokyo, 163–8677 Japan.
[††††††]The author is with Tokyo University of Information Sciences, Chiba-shi, 265–8501 Japan.
a) E-mail: iwamura@ee.kagu.tus.ac.jp
b) E-mail: m.kawamura@m.ieice.org
c) E-mail: kminoru@okayama-u.ac.jp
d) E-mail: iwata@cs.osakafu-u.ac.jp
e) E-mail: kang@ee.kagu.tus.ac.jp
f) E-mail: gohshi@cc.kogakuin.ac.jp
g) E-mail: akira@rsch.tuis.ac.jp
DOI: 10.1587/transinf.2016MUI0001

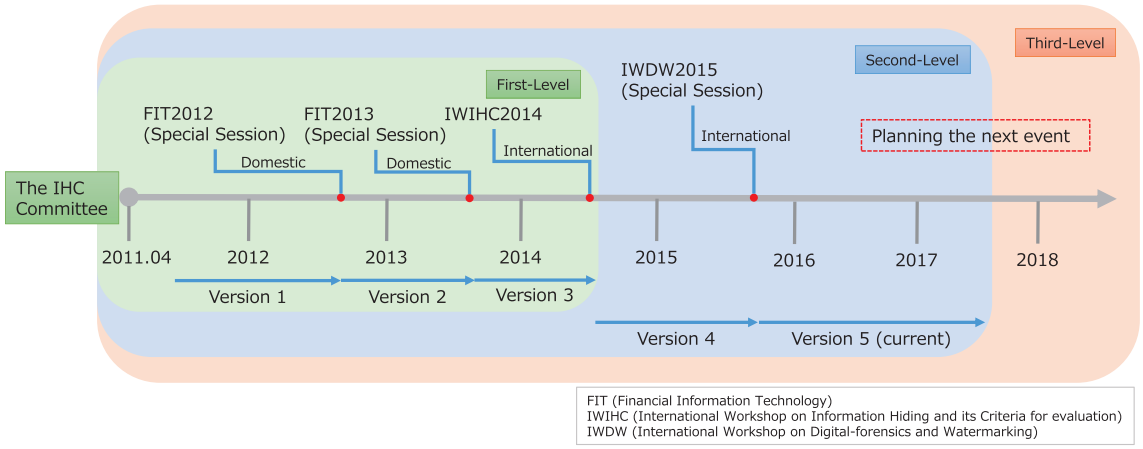


Fig. 1 The levels and history of the IHC evaluation criteria.

events in Japan. The First International Workshop on Information Hiding and Its Criteria for Evaluation (IWIHC2014), which included a watermarking competition, was held in Japan in conjunction with a major security conference, called ASIACCS2014 (ACM Symposium on Information, Computer and Communications Security) [2]. The second event was the 14th International Workshop on Digital-Forensics and Watermarking, IWDW 2015, held in Tokyo, Japan, in October 2015 [3].

The remainder of this paper is organized as follows. In Sect. 2, the IHC Committee and its activities are described. The evaluation criteria and techniques of digital watermarking for still images are presented in Sect. 3. Videos are discussed in Sect. 4. Audio is discussed in Sect. 5. In the final section, we present our conclusions and future directions for the evaluation criteria.

2. The IHC Committee

In this section, we will describe the activity of the IHC Committee including the history of the IHC evaluation criteria and contents for evaluation. The IHC Committee has been established in 2012. The history of the IHC Committee is explained using Fig. 1. We hope this schematic diagram is self-explanatory (see our Web site for more details [10]).

2.1 The Activity of the IHC Committee

Each year, the IHC Committee carries out the following process.

1. Compilation and public presentation of new evaluation criteria
2. Call for information hiding schemes exceeding the newest evaluation criteria
3. Competition among the submitted schemes based on the evaluation criteria
4. Call for opinions on the evaluation criteria and the attacks for the submitted schemes
5. Discussion of the results of the competition, and compilation of opinions on the evaluation criteria and attacks

Based on the result obtained in step 5, the new evaluation criteria for step 1 (the following year) are decided. By repeating this process, the evaluation criteria, which are discussed from many viewpoints and incorporate the consensus of many researchers, are developed. However, step 4 was not performed until recently (for reasons mentioned later).

In the following sections, we explain each step of the process in detail.

2.1.1 Step 1

Researchers and committee members of IHC update the evaluation criteria based on the results of step 5 (from the previous year). The robustness that evaluation criteria require is roughly classified into three levels. The first level is the robustness required for transmission of content, namely, robustness in compression and clipping of content. The second level is the robustness required for utilizing content in addition to the first level, namely, robustness in the conversion between analog and digital (such as printing and scanning of a still picture). This robustness includes robustness against rotation and scaling. The third level is robustness against malicious attack in addition to the second level (such as collusion attacks comparing the watermarked content). The evaluation criteria of the IHC Committee began at the first level, and have reached the second level. This approach attempts to synergistically improve evaluation criteria and digital watermarking.

2.1.2 Step 2

The IHC Committee calls for information hiding schemes exceeding the currently available evaluation criteria. The exhibited evaluation criteria are discussed in detail later in this section.

2.1.3 Step 3

The IHC Committee holds a competition for the subscribed schemes based on the evaluation criteria. The IHC Committee has held a domestic competition twice at Forum on Information Technology in 2012 and 2013, and an international competition twice at the ACM Symposium on Information, Computer, and Communications Security in 2014 and at the International Workshop on Digital-forensics and Watermarking in 2015. Therefore, there are four versions of the evaluation criteria. However, some part of this paper is given to the evaluation criteria for the next version (ver. 5).

2.1.4 Step 4

Cryptographic technology can consider security in open community, because the algorithm is made public as is already mentioned. Therefore, information hiding technology also needs open community in which to consider security. This process is a suitable open place to discuss the security, because the algorithms of submitted schemes are exhibited. However, at present, the evaluation criteria and digital watermarking for this activity are at the second level. Therefore, this process will begin once this activity arrives at the third level.

2.1.5 Step 5

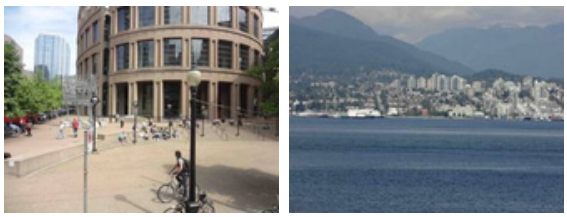
The IHC Committee discusses the results of the competition. When our activity arrives at the third level, the discussion will include opinions on the evaluation criteria and attacks.

2.2 Contents for Evaluation

In the study of digital watermarking for still images, in many cases, LENA is used as an image for evaluation. When researchers use a common image, it is believed that the reproducibility of the experiment and comparison with previous research are easy. However, LENA is an image that was captured using a scanner at 512×512 pixels roughly 30 years ago. Current sampling accuracy, modulation transfer functions (MTFs), and amplitude responses (ARs) are significantly better. Today, when a high resolution image spreads, LENA is not suitable as a standard image for studying today's digital watermarking. For this reason, the IHC Committee has prepared new standard images shown in Fig. 2. These images are 4608×3456 pixels, and contain different features.

In recent years, video formats with a large number of pixels are also frequently used, such as 2K (1920×1080) or 4K (3840×2160). By contrast, the videos used for examination currently available for free have low resolution, and standard video with a high resolution is expensive in many cases. Therefore, the IHC Committee prepared standard pictures that are raw images at 2K and 4K. Images of the free

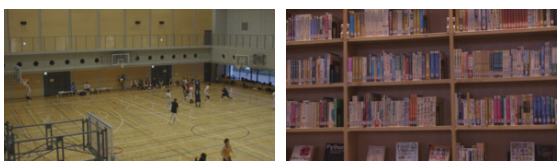
(a) iheval-01 (flower garden) (b) iheval-02 (street view)



(c) iheval-03 (library) (d) iheval-04 (port view))

(e) iheval-05 (bus) (f) iheval-06 (flower pot)

Fig. 2 IHC Standard Images.



(a) Basketball (b) Library

(c) Walk1 (d) Walk2

(e) Lego

Fig. 3 IHC Standard Video.

videos are shown in Fig. 3.

For audio, CD tracks that are well known in this field are used.

The circulation model of content is as follows. First, after copyright information is embedded in the content using digital watermarking and the first compression is per-

formed, the content is sent to consumers using a regular distribution channel and is restored. Typically, circulation stops at this juncture. A user can (illegally) attack this content, performs a second compression, and circulates this version on a network. When this modified content arrives, the embedded copyright information is first detected from the decompressed content. Therefore, compression and decompression are performed twice, and attacks occur between the first decompression and the second compression.

In the detection process, it is assumed that the original content is not available. This assumption is made because it is assumed that an authorized third-party organization (which is not the owner of a digital content) performs detection of digital watermarking. In this case, the owner is assumed to want to avoid giving original content to this organization.

3. Evaluation Criteria for Still Images

In this section, we explain the requirements for the evaluation of still images, and describe the procedure in terms of the evaluation criteria. Based on the criteria, two major approaches [4], [5] are presented at the workshops organized by the IHC Committee. We briefly introduce these methods and discuss their advantages and drawbacks.

3.1 Transition of Edition

There are three metrics used in the evaluation of watermarking methods: capacity, degradation, and robustness. For a given capacity, the evaluation criteria require competition entrants to achieve robustness against specified attacks under constraints on image quality.

In the first version, the amount of watermark information is 64 bits. In the second version, this number was increased to 200 bits. The watermark information should be generated using eight ordered maximal-length sequences (M-sequences). From 10 initial vectors specified by the IHC committee, 10 kinds of watermark information should be used in the evaluation of watermarking methods using the six images shown in Fig. 2. Thus, 10 types of watermarked images are generated for each original image. Any error correcting code can be used to encode the watermark information, as long as the 200 bits are recovered after decoding.

A watermarked image is produced by embedding the watermark information in an original image; then, the watermarked image is compressed using the YUV422 format. As stated in Sect. 2, we assume two compressions during the process of illegal distribution. Although JPEG compression is the most popular image encoding algorithm, other compression tools can be used, as long as the compression ratio satisfies the following criteria. The file size must be less than 1/15 (1/25) after 1st (2nd) encoding, relative to the original size of the YUV422 format.

The assessment of image quality, the PSNR, and the mean structural similarity (MSSIM [6]) should be calculated for each pair of luminance signals (i.e., from the compressed original image and the compressed watermarked image). The luminance signal (Y) must be calculated from RGB color signals R, G, and B using the ITU-R BT.709 standard, defined as follows.

$$Y = 0.2126R + 0.7152G + 0.0722B \quad (1)$$

The attacks used for evaluation are updated with the increase of the version of the criteria. In ver. 1 of the criteria, only the double compression is listed as a common attack, because tolerance against compression is the highest priority. Then, a clipping attack was added to the criteria in ver. 2. After the first compression, 10 HDTV-sized (1920×1080 pixels) images are clipped (the clipping positions are specified). Then, the second compression is performed on the clipped images. The first two competitions were targeted only at domestic researchers in Japan. At the third competition, the IHC Committee extended their activities to international workshops using the same criteria as ver. 2. In ver. 4, scaling and rotation attacks were added to the criteria; the scaling parameter s and rotation angle θ were specified by the IHC Committee. Scaling and rotation attacks should be performed on the watermarked image after the first compression. Its inverse operation is performed on extracting watermark bits in the evaluation process. Then, the clipping attack should be performed in the evaluation. In ver. 5, the parameters s and θ were not given and the clipping attack was directly performed on the scaled or/and rotated image without performing the inverse operation.

The procedure for evaluation in ver. 5 is summarized as follows.

[Embedder]

E1. Embed the 200-bit watermark information.
E2. Compress the watermarked image (first compression).

The image is called a stego-image.

[Attacker]

A1. Perform scaling or/and rotation operation(s).
A2. Clip the HDTV-sized images.
A3. Compress the image (second compression).

The image is called an attacked image.

[Detecter]

D1. Detect the 200-bit watermark information from each clipped image.
D2. Calculate the average bit error rate (BER).

At detection, watermark information must be extracted from the attacked images without using reference information (which includes the original image and additional information). One fixed secret key is used for all detections.

There are two categories of approach: one is to achieve the highest tolerance and the other is to achieve the highest image quality. The former approach targets entries with the smallest file size after the second compression without errors (where the PSNR of the stego-image should be more

than 30 dB), whereas the latter targets the file with the highest average PSNR (where the average BER should be less than 1% or, at worst, less than or equal to 2%).

3.2 Two Potential Approaches

Two potential approaches are introduced in this section. These approaches met the IHC evaluation criteria for still images in ver. 4, but have differences in block segmentation.

Several techniques are found to be useful to devise a method that meets the IHC evaluation criteria. Moreover, watermarking methods have been almost converged with two potential approaches through the previous competitions. These approaches satisfied the IHC evaluation criteria for still images in ver. 4. However, the design concepts of the two approaches are much different. The main difference is expressed in block segmentation. One approach uses a small block that is the same size as the JPEG algorithm, the other uses a much larger block to widely spread the watermark signal. In this paper, we call the method in which a small number of watermark bits are embedded into each DCT block the small-block approach and the method in which a large number of watermark bits are embedded into large blocks the large-block approach.

3.2.1 Small-Block Approach

In the small-block approach, each watermark bit is embedded into each DCT block. Therefore, this approach is compatible with JPEG compression. Moreover, the computational costs for synchronization can be reduced thanks to the small size. The core techniques are the error correcting codes and the weighted majority voting in order to reduce errors.

To meet the criteria, robust watermarking methods must be robust against several types of attack. First, let us consider the clipping attack. When a stego-image experiences a clipping attack, the embedding position of the watermarks is unknown. Therefore, synchronization codes or markers are embedded in the watermarks to indicate the embedding position. The interval of the synchronization code is determined by the possible size of the clipped image. No one has previous knowledge of the embedding position; therefore, synchronization will be performed using a brute force search. In the small-block approach, for the synchronization, 64 possible positions are searched because of the size of the pixel block, i.e., 8×8 pixels [4], [8].

Watermarks can be extracted from any clipped region of the stego-image. Therefore, the same watermarks are embedded at different locations throughout the stego-image. We call the area that is surrounded by the synchronization code a segment. Several watermarks are embedded into each segment. If many watermarks are extracted from a clipped image, errors can be reduced by employing weighted majority voting. However, the reliability of the extracted watermark bits may be different in different areas. Therefore, to measure the reliability, check bits are also embedded. This check is a public bit sequence. The reliability can be estimated using the coincidence ratio of the check bits.

Watermark information may be incorrectly extracted because of attacks; therefore, it should be encoded using error correcting codes. For example, low-density parity-check (LDPC) and concatenated codes are employed in [4], [8], and convolutional code is employed in [9]. In general, the image quality of a stego-image could be improved by employing fewer watermark bits. Therefore, the code length of the encoded watermark should be smaller.

Next, let us consider scaling and rotation attacks. If parameters such as s and θ can be estimated, synchronization can be accomplished using the inverse transformation. Otherwise, a brute force search for s and θ may be used to attempt synchronization. However, this type of search requires a large amount of computational resources. Therefore, feature extraction methods are promising approaches in ver. 5.

Based on the above explanation, we briefly summarize the latest method [4] for a small-block approach. The 200-bit watermark information is encoded using a 1012-bit codeword by the concatenated code with Bose-Chaudhuri-Hocquenghem (BCH) and LDPC codes. The length of the check bits is 25 bits. The codeword is embedded in five different non-overlapped regions in each segment. When decoding, upon finding the synchronization code, synchronization can be accomplished. Using the reliability obtained from the check bits, the extracted watermarks are estimated using weighted majority voting. Then, the estimated watermark information is decoded using the error correcting code. Because these robust techniques are used, a near-zero error rate can be achieved.

3.2.2 Large-Block Approach

The large block approach spreads watermark signal much wider range of frequency components. The robustness against noise is much higher than the small block approach while the computational costs are larger. In order to reduce the computational costs, the fast frequency transformation algorithm is employed at the embedding/detecting watermark. The core technique in the large block approach is the combination of spread spectrum and orthogonal frequency division multiplexing (OFDM) techniques.

Generally, the spread spectrum method and its variants must remove the interference among sequences at the embedding stage to improve the robustness. If orthogonal sequences are applied, no interference occurs among sequences. It is possible to use the method of OFDM.

Quantization index modulation (QIM) [7] is a method that can extract a watermark without the original image. The IHC evaluation criteria requires a high compression ratio and few errors. In JPEG compression, high frequency components in the discrete cosine transform (DCT) domain are strongly quantized. Therefore, if the watermarks are embedded in lower frequency components by QIM, the stego-image has robustness against JPEG compression.

Suppose that the host data is an L-dimensional vector $\boldsymbol{x}$ selected randomly from the low frequency components of an image. Initially, a sequence $\boldsymbol{d}$ is calculated using

$$\boldsymbol{d} = \mathrm{DCT}(\boldsymbol{\rho} \otimes \boldsymbol{x}), \tag{2}$$

where $\boldsymbol{\rho}$ is a secret PN sequence, $\otimes$ indicates element-wise multiplication. For a given watermark bit, one element of $\boldsymbol{d}$ is modified using the QIM method. Because the vector $\boldsymbol{d}$ has L elements, at most L bits can be embedded. Therefore, the synchronization vector $\boldsymbol{s}$ and the watermark information $\boldsymbol{w}$ are embedded into the vector $\boldsymbol{d}$ and the watermarked vector $\boldsymbol{d'}$ is obtained. Using the following operation, the watermarked data is calculated.

$$\boldsymbol{x'} = \boldsymbol{\rho} \otimes \mathrm{IDCT}(\boldsymbol{d'}). \tag{3}$$

Based upon the above DCT-OFDM method using the QIM embedding method, watermark information and synchronization information are repeatedly embedded into a host image in the following way. A host image is partitioned into blocks of 256×256 pixels. The watermark and synchronization information are partitioned into r pieces, and are embedded into every r blocks. If at least r blocks are presented, the detection of watermark information is possible. At each block, two-dimensional DCT is performed and 512 DCT coefficients are selected as the host data $\boldsymbol{x}$ of length $L = 512$ from low and middle frequency based on a secret key. In [5], the 200-bit watermark information is directly embedded into r blocks. For the enhancement of robustness, the watermark information is encoded using convolutional code at an encoding rate of roughly 1/2; it is decoded by the Viterbi algorithm presented in [9].

In the case of a clipped image, we must find synchronization points. From the top-left coordinate, a clipped image is partitioned into blocks of 256×256 pixels. Then, we attempt to extract r pieces of synchronization information from r blocks. Because a cyclically shifted sequence may be extracted, we test r cyclically shifted patterns. If the Hamming distance between the original sequence and the extracted sequence is within a certain window, we determine that the point is synchronized with the number of cyclic shifts. Otherwise, the coordinate is shifted, and the same operation is performed until synchronization is recovered. This operation is performed at most 256×256 times. In the case of failure, we determine that no watermark is contained in the given image.

3.3 Considerations

Generally, the longer the spread-spectrum sequences become, the higher the robustness against noise. The block size is 256×256 in the DCT-OFDM method. The watermark information is modulated into a waveform of length $L = 512$ and is embedded (widely spread) throughout the frequency domain in the block. Because of the long length, the method achieves high robustness against attacks. On the other hand, the large block size increases the computational costs. Even though DCT operations can be performed using the fast algorithm, synchronization recovery is very time-consuming because the number of candidate coordinates is related to the block size.

In both approaches, the robustness against attacks is relatively high. It seems important to investigate the synchronization method, including the embedding of the synchronization signal and the recovery operation. The amount of energy assigned to the synchronization signal should be controlled so as not to sacrifice the image quality. The accuracy and the computational time should be considered in the recovery operation. In [9], part of the recovery operation is hierarchically performed to minimize the computational costs. We can say that it is desirable to investigate the method in which computational costs are not linearly increased with the number of candidates of parameters for geometrical attacks.

4. Evaluation Criteria for Video

4.1 Summaries of the IHC Criteria for Video

There have been many studies of digital watermarking; however, the state-of-the-art remains imperfect. The IHC Committee is working to improve this situation by promoting the development of digital watermarking techniques. In particular, it aims to help develop standard evaluation criteria and to sponsor watermarking competitions based on these criteria [10].

In this section, we summarize the IHC evaluation criteria for video. More detailed information is given in the document describing the IHC criteria [10].

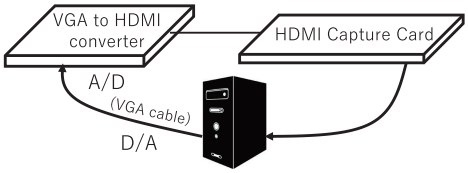
4.1.1 Image Quality Assessment

Watermarked video clips should be compressed using MPEG-4 part 10 (H.264) or the MPEG-2 codec. The size of the compressed bit stream should be less than 1/100 that of the original video clip. The original unwatermarked video clips should be compressed using the same parameters. Both sets of clips should then be decompressed, and the PSNR should be calculated for each pair of luminance signals from the RGB channel using Eq. (1) (PSNR should be greater than 30 dB).

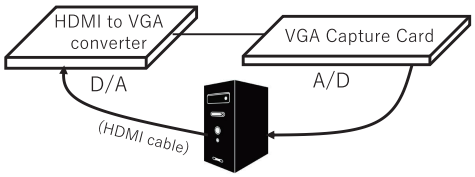
The bit rate of the original video clip should be 1.2 Gbps, and the average size of the coded video stream should be less than 12 Mbps.

4.1.2 Tolerance Assessment

Although digital content is protected using a digital rights management system, analogue content can easily be copied. This phenomenon is called the analogue hole. The watermark must remain after the content has been changed into an analogue signal using a D/A converter.



(a) Example 1



(b) Example 2

Fig. 4 D/A and A/D conversion test.

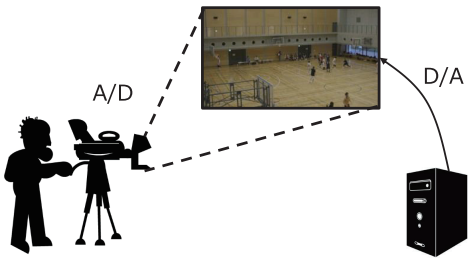


Fig. 5 Camcorder jamming test.

(1) D/A and A/D Conversion

After the watermarked video clips are compressed as described above, they should be decompressed, converted from digital to analog (D/A), and then converted from analog to digital (A/D). All of the embedded watermark information should be detectable in the digitized video. For converting digital video sources to analog ones, we can employ a D/A converter like HDMI–VGA as well as the analog output of a common video card (see Fig. 4).

(2) Camcorder jamming

The watermarked video without lossy compression should be subjected to the "camcorder jamming" test. This test performs D/A and A/D conversion using a screen and a camcorder. There are no limitations on the equipment that the entrants can use for testing (see Fig. 5).

4.1.3 Amount of Data to be Embedded

The amount of data embedded into each 15-s clip should be 16 bits.

4.2 Potential approach

Here, we introduce the method developed by M. Iwata, one of the authors, as an example of one of the potential approaches.

4.2.1 Embedding Procedure

First, a watermark of length N bits is coded using convolutional code to obtain the code array, where the code rate and the constraint length of the convolutional code are $1/2$ and 7, respectively. Then, the length of the code array is $2N$ bits. Second, each frame of an original video is divided into $X \times Y$ blocks. The blocks (excluding corner blocks) are used to embed the code array. Then, the relationship among X, Y, and N is $N = (XY - 4)/2$. Finally, the code array is embedded into the differences between the corresponding blocks in the $2f$-th and $(2f + 1)$-th frames by controlling their signs, where f is the index of a frame. The sign of the difference of a block is modified so that the sign is the same as the signs of corner blocks when the corresponding bit of the code array is 1 and vice versa. Here, the signs of all corner blocks are positive when f is even and negative when f is odd.

4.2.2 Extracting Procedure

The inputs of the extracting procedure are M frames obtained using the estimation procedure for the watermarked region, as described in Sect. 4.2.3.

First, the $M - 1$ difference frames are obtained from all areas of two successive frames in the M frames. Second, the half of $M - 1$ difference frames with higher absolute pixel values are selected for extracting a watermark because half of them are used for embedding (with the same f in the embedding procedure) and the remaining are not (with different values of f). Third, the $(M - 1)/2$ difference frames are divided into XY blocks in the same manner as used in the embedding procedure. The code array is extracted from the blocks (excluding the corner blocks) in the raster scan order based on the condition for extraction. The condition for extraction is that the extracted bit is 1 (0) if the sign of the sum of the pixels in the block is the same as that of (is different from that of) the pixels in all of the corner blocks. Finally, the extracted code array is fixed based on the majority rule for each extracted bit. Then, the watermark is obtained by decoding the extracted code array using Viterbi decoding, where the cost of the shortest path is calculated so as to evaluate the reliability. The extraction procedure starts again if the cost of the shortest path would be larger than R, where R is a pre-defined threshold.

4.2.3 Estimation Procedure for Watermarked Region

The described method can estimate the watermarked region from the recaptured frames in the following manner, where the inputs of this procedure are the recaptured M frames.

First, the $M - 1$ difference frames are obtained from every pair of successive frames in the M frames. Then, the $M - 1$ difference frames are divided into blocks for an estimation of size $B_x \times B_y$ pixels. Second, the trace of embedding for each block for estimation is checked based on the

Fig. 6 Experimental setup.

Fig. 7 Implementation.

magnitude correlation among the means of pixel values in the blocks at the same coordinate in successive frames. Finally, the four corners of a watermarked region are estimated so that the watermarked region is a projectively transformed rectangle. Then, the region surrounded by the four corners is corrected using a projective transformation such that the four corner points of the region fit the corresponding corner points of the size of the recaptured frames.

4.2.4 Implementation

Figure 6 and Fig. 7 show the experimental setup and the implemented application, respectively. In Fig. 7, the four circles in the application indicate the estimated corners of the watermarked region. The number in the central area of the application is the extracted watermark (which is correct). As indicated by Fig. 6 and Fig. 7, the implementation is practical.

5. Evaluation Criteria for Audio

In this section, we explain the requirements for the evaluation and the procedure for audio watermarking. Three basic techniques of embedding and extraction were presented at workshops organized by the IHC Committee. We briefly introduce these methods and discuss their advantages and drawbacks.

5.1 Editions of Evaluation Criteria

Sixteen-bit linear quantization, a sampling frequency of 44.1 kHz, stereo format, and the SQAM (sound quality assessment materials) database[†] (CD Tracks 27, 32, 35, 40, 65, 66, 69, and 70) are used repetitively for a duration of 60 s each. The SQAM program signals are carefully chosen so as to reveal listener impairments that have been observed in testing of both analog and digital audio systems. These signals are also used to evaluate audio quality for memory audio [11].

Ninety-bit random payloads for each 15 seconds of the host signal should be embedded, meaning that 360 bits per 60 seconds should be embedded. PQevalAudio v2r0, which is an implementation of PEAQ (perceptual evaluation of audio quality), should be used to measure the ODG (objective difference grade) of the eight stego-signals. The ODG between the original PCM host signal (the reference signal) and the stego-signal should be calculated. The ODG should be greater than -2.5. Additionally, the stego-signal is then compressed as an MP3 128-kbps joint stereo signal and decompressed as the degraded signal. The arithmetic mean of eight ODGs should be greater than -2.0.

The following signal processing or perceptual coding attacks should be applied to the stego-signals, after which, the payload should be extracted. The attacks are SDMI (Secure Digital Music Initiative) phase II screening, and STEP2000 and STEP2001 conducted by JASRAC (Japanese Society for Rights of Authors, Composers and Publishers), which were developed to test commercial audio watermarking technologies. These attacks have been confirmed to be realistic in terms of sound quality degradation of either decompressed signals or signals after inverse processing [12]. The mandatory attacks and optional attacks are pre-defined. The criteria before ver. 4 required three of the optional attacks. Four or more of the optional attacks have been required since ver. 5. These attacks are realized using general signal processing tools and codecs that are freely available on the Internet.

Mandatory tests

- MP3, 128 kbps, joint stereo
- A series of attacks that mimic DA and AD conversions (since ver. 3)

Optional tests

- Gaussian noise addition (overall average SNR 36 dB)
- Bandpass filtering 100 Hz to 6 kHz, -12 dB/oct.
- Frequency scale modification (time invariant) $\pm 4\%$
- Linear speed change $\pm 10\%$
- A single echo addition, 100 ms, -6 dB
- MP3 128 kbps (joint stereo) tandem coding
- MPEG4 HE-AAC 96 kbps (since ver. 2)
- A series of attacks that mimic aerial transmission (since ver. 5)

[†]http://tech.ebu.ch/publications/sqamcd/

Forty-five seconds of the modified stego-audio from which the initial sample is randomly chosen in the initial 15 seconds for each simulation should be used for extracting the payload; this selection is intended to simulate a clipping attack on the stego-audio. The BER (bit error rate) is defined as the average number of mismatched bits over 100 trials between the embedded and extracted payloads relative to the 180 bits that are embedded into 15 to 45 seconds of stego-audio.

The host audio clips used from ver.1 to ver.4 contain long silent periods, i.e., periods with an amplitude of zero, at initial and final segments, which caused difficulties in embedding payload data. Such segments were removed from the host audio clips on editing ver.4. At the beginning of the investigation of the criteria, error correction schemes were not introduced, because we wanted to know the nature of the proposed watermarking techniques. Taking account of practical use of watermarking, error correction schemes have been permitted from ver. 4 onwards.

5.2 Basic Embedding and Extraction Techniques

The number of audio watermarking technologies that is certificated as satisfying the criteria is limited. Basic techniques of embedding and detection that satisfy the criteria are QIM in phase angles between left and right channels [13], AM (amplitude modulation) applied to subband signals [14], and thresholding of magnitude differences between coefficients obtained from two different wavelet filters [15].

5.2.1 Embedding Into Stereo Phase Difference

Embedding the payload in the stereo phase difference [13] is achieved by dividing 2π phase difference into 16 quantization steps in which '0' and '1' are alternately assigned. A 4096-sample Hamming window, shifted in 2048-sample steps, is applied to calculate the phase and amplitude spectra of the host signal. Therefore, the maximum amount of embedding is 2023 bits per 46.4 ms. The maximum angle of phase shift is $1/32 \times 2\pi$ and the maximum stereo phase difference is $1/16 \times 2\pi$.

This technique slightly modifies the phase spectra, which is imperceptible by the monoaural human auditory system. However, the binaural human auditory system is very sensitive to interaural time differences. A $1/16 \times 2\pi$ phase difference between left and right channels corresponds to an interaural time difference of 62.5 μs at 500 Hz. This value exceeds the perceptual threshold of interaural time difference at 500 Hz, which corresponds to roughly 10 to 20 μs. Objective quality degradation might be underestimated, because the PEAQ algorithm does not consider interaural phase distortion. The above discussion considers the worst case scenario of perceptual degradation of embedding in stereo phase differences. Subjective evaluation of the stego-audio will be studied in future.

This technique is robust against MP3 and MPEG4AAC conversions, bandpass filtering, Gaussian noise addition, and single echo addition because these attacks retain phase information at the low-frequency and intense energy region. However, this technique is vulnerable to geometric attacks such as frequency-scale modification, linear speed changes, and DA/AD conversion including slight sampling frequency mismatch because these attacks shift the frequency-axis of the stego-audio. Also, applying independent all-pass filters to the left and right channels is a critical attack for this technique, without severe quality degradation.

5.2.2 Embedding by Amplitude Modulation

Embedding the payload in sinusoidal AM at relatively low modulation frequencies that are applied with opposite phase to neighboring subband signals can be used as the carrier of embedded information and for synchronization of data frames [14]. The embedded information is encoded in the form of relative phase differences between the AM signals applied to several groups of subband signals.

This technique is robust against MP3 and MPEG4AAC conversions, DA/AD conversions, bandpass filtering, and single echo addition, but is vulnerable to geometric attacks, because the bandwidths of the subbands are constant over the center frequencies. To resist geometric attacks, the detection of frame synchronization should be improved by compensating for the frequency scale and/or the time scale in the time-frequency domain. The maximum amplitude of AM collected from the subband group for synchronization is observed at the frequency-scale compensation that corresponds to the amount of frequency-scale modification applied [14].

This method is also vulnerable to Gaussian noise addition to track no. 35, which has small stego-energy segments. Payload extraction from such low SNR segments of the stego-audio should be improved. Adaptive decision of watermarking intensity that takes the signal power level into account may be explored in future studies.

5.2.3 Embedding Into Difference of Different Wavelet Filters

Embedding the payload in the high frequency region of the host signal is achieved by thresholding the magnitude difference between coefficients obtained from two different wavelet filters [15]. The magnitude difference between two different wavelet filters contains amplitude and phase spectra of the high frequency region of the host signal. Therefore, this method is robust against attacks that retain the high-frequency spectral feature, such as MP3 conversion, single echo addition, and Gaussian noise addition attacks. However, this method is vulnerable to bandpass filtering, geometric attacks, and MPEG4AAC conversion (which replicates the high frequency spectra from the low frequency spectra).

5.3 Effective Techniques

Several important techniques are combined with the basic technique, as required to satisfy the criteria: frame synchronization, which achieves temporal synchronization between embedding frames of the host signal and detection frames of the stego-signal, and multiple and distributed embedding into the time-frequency domain of audio contents.

The former method is required to cope with the clipping attack that randomly varies the initial sample of the stego-signal. Embedding synchronization code in conjunction with payload data into the host signal is an effective technique for achieving frame synchronization. However, the computational cost is relatively high because the decoding process is repeated by incrementally shifting a temporal window to search for the synchronization code. The addition of a synchronization signal composed of M-sequences added to the host signal is another effective technique. The M-sequence signal can be rendered less perceptible by replacing the amplitude of its spectrogram with that of the host signal before addition. At the decoder, cross correlation between the stego-signal and the spectrally-weighted M-sequence signal is obtained to realize whitened cross correlation, which creates a sharp peak to indicate the synchronization point.

The latter method is required to achieve reliable detection from stego-audio that has a sparse energy distribution in the time-frequency domain. The same information bits are repeatedly embedded into the frequency domain [13], [14] and the time domain [13]–[15]. Detection of the payload bits is based on the amplitude-weighted average [13], the average of AM waveforms [14], or the majority decision [15].

6. Conclusion

In this paper, we introduce the activities of the IHC Committee and their evaluation criteria for digital watermarking of still images, videos, and audio. Currently, the evaluation criteria is at the second level, which achieves robustness for transmission and utilization of content. We aim to achieve the third (and final) level, namely, robustness against malicious attacks.

References

[1] Ministry of Internal Affairs and Communications, Japan. White paper 2015: Information and Communications in Japan. http://www.data.go.jp/data/dataset/soumu_20160201_0005 (accessed on 20 July 2016).
[2] Proceedings of the 1st International Workshop on Information Hiding and its Criteria for Evaluation (IWIHC2014), ACM, New York, NY, USA, 2014.
[3] Post-conference proceedings of the 14th International Workshop on Digital-Forensics and Watermarking (IWDW2015), LNCS 9569, Springer-Verlag, 2016.
[4] N. Hirata and M. Kawamura, "Watermarking method using concatenated code for scaling and rotation attacks," Proc. IWDW2015, LNCS, vol.9569, pp.259–270, Springer-Verlag, 2016.
[5] M. Hakka, M. Kuribayashi, and M. Morii, "DCT-OFDM based watermarking scheme robust against clipping attack," Proc. IWIHC2014, pp.18–24, 2014.
[6] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., vol.13, no.4, pp.600–612, 2004.
[7] B. Chen and G.W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," IEEE Trans. Inform. Theory, vol.47, no.4, pp.1423–1443, 2001.
[8] N. Hirata and M. Kawamura, "Digital watermarking method using LDPC code for clipped image," Proc. IWIHC2014, pp.25–30, 2014.
[9] H. Ogawa, M. Kuribayashi, M. Iwata, and K. Kise, "DCT-OFDM based watermarking scheme robust against clipping, rotation, and scaling attacks," Proc. IWDW2015, LNCS, vol.9569, pp.271–284, Springer-Verlag, 2016.
[10] IHC. Information Hiding and its Criteria for Evaluation. Available online: http://www.ieice.org/iss/emm/ihc/en/index.php (accessed on 20 July 2016).
[11] AV & IT Equipment Standardization Committee, "JEITA CPR-2601 the designation of audio quality for memory audio," 2010. http://www.jeita.or.jp/japanese/standard/book/CPR-2601/ (accessed on 20 July 2016).
[12] A. Nishimura, M. Unoki, K. Kondo, and A. Ogihara, "Objective evaluation of sound quality for attacks on robust audio watermarking," Proceedings of Meetings on Acoustics (POMA), International Congress on Acoustics 2013, Montreal, vol.19, no.1, 2013.
[13] N. Ono, "Robust audio information hiding based on stereo phase difference in time-frequency domain," Proc. of IIHMSP2014, pp.260–263, 2014.
[14] A. Nishimura, "Detection of frequency-scale modification using robust audio watermarking based on amplitude modulation," Proc. IWDW2015, LNCS, vol.9569, pp.299–311, Springer-Verlag, 2016.
[15] T. Ito, H. Kang, K. Iwamura, K. Kaneda, and I. Echizen, "Audio watermarking using different wavelet filters," Proc. IWDW2015, LNCS, vol.9569, pp.312–320, Springer-Verlag, 2016.

Keiichi Iwamura received B.S. and M.S. degrees in Information Engineering from Kyushu University in 1980 and 1982, respectively. During 1982–2006, he was with Canon Inc. He received a Ph.D. from Tokyo University. He is now a Professor at the Tokyo University of Science. His subjects are coding theory, information security, and digital watermarking. He is a fellow of the Information Processing Society of Japan and chairperson of Information Hiding and its Criteria for Evaluation.

Masaki Kawamura received B.E., M.E., and Ph.D. degrees from the University of Tsukuba in 1994, 1996, and 1999, respectively. He joined Yamaguchi University as a research associate in 1999. Currently, he is an associate professor at the same institute. His research interests include associative memory models and information hiding. He is a senior member of IEICE and a member of JNNS, JPS, and IEEE.

Minoru Kuribayashi received B.E., M.E., and D.E degrees from Kobe University, Kobe, Japan, in 1999, 2001, and 2004, respectively. From 2002 to 2007, he was a Research Associate in the Department of Electrical and Electronic Engineering, Kobe University. In 2007, he was appointed as an Assistant Professor at the Division of Electrical and Electronic Engineering, Kobe University. Since 2015, he has been an Associate Professor in the Graduate School of Natural Science and Technology, Okayama University. His research interests include digital watermarking, information security, cryptography, and coding theory. He received the Young Professionals Award from IEEE, Kansai Section, in 2014.

Motoi Iwata received M.E. and D.E. degrees in computer and systems sciences from Osaka Prefecture University, in 1999 and 2005, respectively. From 1999 to 2016, he was an Assistant Professor in the Graduate School of Engineering, Osaka Prefecture University. Since 2016, he has been an Associate Professor in the Graduate School of Engineering, Osaka Prefecture University. His current research focuses on digital watermarking, data hiding, image retrieval, comics image analysis and reading-life log. He is a member of the Imaging Society of Japan, the Institute of Image Information and Television Engineers, and IEEE.

Hyunho Kang is currently an Assistant Professor in the Department of Electrical Engineering at Tokyo University of Science, Japan; he has held this position since April 2013. He received his Ph.D. from the University of Electro-Communications, Tokyo, in 2008. From 2008 to August 2010, he was a Researcher/Assistant Professor at Chuo University, Tokyo, where he was part of a team that developed Biometric Security technologies. From September 2010 to March 2013, he was an AIST Postdoctoral Researcher at the National Institute of Advanced Industrial Science and Technology (AIST), Japan, where his research work focused mainly on the evaluation of physical unclonable functions. His main interests are digital watermarking, biometric security, and physical unclonable functions. He is a member of IEICE, IPSJ, and IEEE.

Seiichi Gohshi is a professor of Kugakuin University. He received his BS degree, MS degree and PhD degree from Waseda University in 1979, 1981 and 1997. He joined Japan Broadcasting Corporation (NHK) in 1981. He started his research at NHK Science Technical Research Laboratories (STRL) in 1984. He helped to develop the HDTV broadcasting system, transmission systems, signal processing systems. He was the project leader of the Super Hi-Vision (8K) transmission system and successfully conducted the first Super Hi-Vision transmission test at IBC2008. He also developed a watermark system that was used in movie theaters. He joined Sharp Corporation as a division deputy general manager in 2008 and developed high resolution systems. He is currently a professor of Kogakuin University. His research interests are video and image signal processing especially for super resolution and forensic technologies.

Akira Nishimura received B. Eng. and M. Eng. degrees in acoustics from Kyushu Institute of Design in 1990 and 1992, respectively. He received a Ph. D. degree in audio information hiding from Kyushu University in 2011. Since 1996, he has been a faculty member of Tokyo University of Information Sciences. He is a professor in the Department of Informatics. His current research interests are auditory modeling, audio information hiding, musical acoustics, and the psychology of music. He is a member of the Acoustical Society of Japan, the Audio Engineering Society, IEEE, and the Japanese Society of Music and Cognition. He won the Sato Prize from the Acoustical Society of Japan in 2012.