IHC Evaluation Criteria and Competition

http://www.ieice.org/iss/emm/ihc/en

1 Scope

While digital technologies have completely changed our lifestyles by giving us a plethora of convenient digital tools, they have also created problems. Digital copyright infringement is one of them. Twenty years ago, this was not a big problem. Today, however, the great variety of digital tools has made it very easy to copy digital content. This has led to a rapidly growing amount of illegal digital content being distributed all over the world. As a result, digital copyright protection has become an important issue. Although much research has been done on digital watermarking, the state of the art has not yet reached the level needed. The Information Hiding Criteria (IHC) Committee is working to improve this situation by promoting the development of digital watermarking technologies. In particular, it aims to help develop standard evaluation criteria and to sponsor watermark competitions based on those criteria.

2 Watermark Criteria for Images (ver.5)

Since image content is delivered after coding, tolerance against coding is considered to be the top priority. The evaluation criteria will be revised in accordance with advances in watermarking technology, the needs of the content industry, and the practicality of the competition.

This competition requires as a minimum both coding tolerance and clipping tolerance. The tolerance for scaling, rotation, and their combination is required as the additional attack. Entrants should explain in their entries all of the tolerances of their watermarking scheme.

2.1 Image Quality Assessment

The six images provided by the Information Hiding Criteria (IHC) Committee for quality assessment are shown in Fig.1. They can be downloaded at http://www.ieice.org/iss/emm/ihc/en/image/image.php. They are 4608×3456 (about 16M) pixel color images. They should be watermarked and then compressed using the YUV422 format. The size of the compressed file should be less than 1/25 that of the original file. The original unwatermarked images should also be compressed using the same parameters. Both sets of images should then be decompressed, and the peak signal-to-noise ratio (PSNR) and the mean structural similarity (MSSIM)[1] should be calculated for each pair. The PSNR of each pair, which is calculated with luminance (luma) signal, should be higher than 30 dB. Luma is derived according to the ITU-R BT.709 standard, which is shown as follows.

$$Y = 0.2126R + 0.7152G + 0.0722B \tag{1}$$

[1] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Processing, vol. 13, no. 4, pp. 600-612, 2004.

(a) Image 1 (Flower garden)

(b) Image 2 (Street view)

(c) Image 3 (Library)

(d) Image 4 (Port view)

(e) Image 5 (Bus)

(f) Image 6 (Flower pot)

Figure 1: IHC standard images.

The compression process to reduce the file size to less than 1/25 the original size consists of two steps, which are explained in Sect.2.2. Although JPEG and JPEG2000 are candidate compressing tools, other compression tools can be used as long as they meet the requirements explained in Sect.2.2. There are also many candidate tools for the scaling and rotation. It is recommended for entrants to use the ImageMagick[2] tool. If another tool is used, entrants should include relevant information about the tool along with their entry. The IHC Committee will conduct subjective assessments if necessary to evaluate the watermark technologies.

2.2 Tolerance Assessment

- Watermark information should be embedded into the whole image, and the compression-decompression cycle should be performed twice. The file size should be less than 1/15 of the original size after the first compression, and the decompressed images should be compressed on the second compression. After the second compression, the file size should be less than 1/25 of the original size. The compression ratio is determined not by the RGB files but by the YUV422 files. The quality factor (QF) of the second compression should be stored.

- As the additional attacks, scaling $s \in [80, 120]$ [%], rotation $\theta = [0, 10]$ [°], and their combination (s, θ) should be checked for the evaluation.

- The rotation center is the image center, where θ represents the degrees of clockwise rotation. Empty triangles left over from rotating the image are filled with the background color (i.e., white or black).

- After the additional attacks, four HDTV-size (1920×1080) images should be clipped from the attacked image. The center coordinates of these clipped images are $(x \pm 700, y \pm 500)$,

[2] `http://www.imagemagick.org/`

where (x, y) is the center of the attacked image. The watermark embedded in each clipped image should be detectable.

- The clipped images are compressed with JPEG using the same QF of the above second compression.
- During the review process, the IHC Committee may request the detection rate for different areas.

2.3 Embedding and Detection of Watermark Information

- No reference information including the original image can be used in the detection.
- The same watermark information should be embedded in all six images.
- Ten types of watermarked images should be generated for each original image using ten different bit sequences (as explained below). The average error rate and image quality (PSNR and MSSIM) should be calculated for these ten images.
- No additional information can be used in the detection.
- One fixed secret key should be used for all detections.

2.4 Watermark Information

The amount of watermark information to be embedded is 200 bits. The watermark information should be generated by using eight ordered maximal length sequences (M-sequences). Each polynomial should be generated in the form $x^8 + x^4 + x^3 + x^2 + 1$. The initial values should be given as follows:

$$a_7x^7 + a_6x^6 + a_5x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 \rightarrow (a_7, a_6, a_5, a_4, a_3, a_2, a_1, a_0)$$

1.	(1,0,1,0,1,0,1,0)	2.	(1,0,1,0,1,0,1,1)	3.	(1,0,1,1,1,0,1,0)
4.	(1,1,1,0,1,0,1,0)	5.	(1,0,1,0,1,0,0,0)	6.	(1,0,1,0,0,0,1,0)
7.	(1,0,0,0,1,0,1,0)	8.	(0,0,1,0,1,0,1,0)	9.	(1,1,1,1,1,0,1,0)
10.	(1,0,1,0,1,1,1,0)				

The watermarks should be sufficiently tolerant to be detectable in no less than 200 bits in each clipped image. Any error-correcting code can be used to encode the watermark information as long as the 200 bits are recovered from the codeword after decoding.

2.5 Content Flow

The watermarking technology being entered should be used to embed the watermark information given in Sect.2.4 into all six images. After each image file has been compressed twice, the file size should be smaller than 1/25 that of the original image file. The QF of the second compression should be stored. The file size percentages are based on the original file size, being

the size of a YUV422 file. For the first compressed files, after each image file has been decompressed, the PSNR and MSSIM between the original and the watermarked images should be calculated. The PSNR should be higher than 30 dB for the luminance signal given by Eq.(1). The 200 bits of watermark information should be detectable. The evaluation flow is summarized as follows.

[Watermarker's Operation]

- Watermark information (200 bits) is embedded into the original image.
- The watermarked image is compressed with JPEG whose file size should be smaller than 1/15 the size of the original YUV422 image file. The compressed file is called a stego image. The PSNR value between the compressed original image and the stego image is calculated.

  Note that the watermark information (200 bits) can be encoded by an error-correcting code whose code length is more than 200 bits.

[Attacker's Operation]

- The stego image is compressed once again with JPEG whose file size should be smaller than 1/25 the size of the original YUV422 image file. The QF of the JPEG compression is stored.
- Scaling, rotation, and their combination should be performed on the stego image. The attack parameters such as s and θ are NOT available at the extraction of the watermark information.
- After the above geometrical attack, the image is clipped by an HDTV-size area (1920×1080) at the four specified coordinates.
- The clipped images are compressed with JPEG using the QF. The obtained image is called an attacked image.

[Detection Operation]

For a given attacked image of HDTV size (1920×1080), the watermark information (200 bits) is extracted without referring to the original image, attack parameters, and any information related to the original image and attacks. The assistant of the decoder of the error-correcting code is available.

[Remark]

- The watermark information must be detected from at least three clipped images among four clipped images. For the three clipped images, the bit error rate of the detected watermark information must be no more than 1% on average. At the worst case, the bit error rate for the three clipped images must be less than 2%. In case of highest tolerance category, the bit error rate must be 0 for every attack parameter.

Table 1: Average compression ratio, PSNR value, and MSSIM value.

	Compression ratio	PSNR [dB]	MSSIM
image1			
image2			
image3			
image4			
image5			
image6			
average			

- The scaling parameter $s[\%]$ is assumed to be an integer varying in the range $80 \leq s \leq 120$. (In the future, it will be a rational number.)
- The rotation parameter $\theta[^\circ]$ is assumed to be an integer varying in the range $0 \leq \theta \leq 10$. (In the future, it will be a rational number.)
- After the scaling or rotation, four HDTV-size images whose center coordinates are $(x \pm 700, y \pm 500)$ are clipped from the attacked image.

It is recommended to use the ImageMagick tool to perform the above operations including the JPEG compression, scaling, and rotation.

The PSNR and MSSIM calculation and the detection of watermark information should be done by the entrant, and the results should be included in the entry. The entry should also include the details of the embedding and detection algorithms.

2.6 Information Required for Submission

- Embedding and detection algorithms
- Compression ratio, PSNR value for the luminance signal given by Eq.(1), and MSSIM value[3] for the six images after the 1st compression (see Table 1)

  The contest category enumerated in Sect.2.7 must be specified for the data
- Average error rates for four HDTV-size areas after the additional attack and second decompression (see Table 2)
- Scaling $s = \{80, 90, 110, 120\}$, Rotation $\theta = \{3, 5, 7, 10^\circ\}$, and their Combination $(s, \theta) = \{(80, 9), (90, 7), (110, 5), (120, 3)\}$ should be checked for the evaluation.

In Table 2, "No attack" means that no additional attack is performed, namely, only JPEG compression and clipping are performed. In the case of "Scaling," "Rotation," and "Combination," there are four candidates for parameters s and θ. Therefore, the average error rates for total thirteen attacks must be calculated from ten initial values as specified in Sect.2.4 and best three clipped images of the provided images. That is, each error rate in each field of Table 2 is averaged over 10×3 data for each attack.

[3]The value of MSSIM should be calculated by using the default parameter.

Table 2: Average error rate for four HDTV-size areas with additional attacks.

	No attack	Scaling s				Scaling θ				Combi (s, θ)			
		80	90	110	120	3°	5°	7°	10°	80, 9	90, 7	110, 5	120, 3
image1													
image2													
image3													
image4													
image5													
image6													
average													

2.7 Contest Categories

It is required to show respective data (tables) for the following two categories.

- Highest Tolerance
  This category targets entries with the smallest file size after the second compression for the six images under the conditions of the IHC standards, ver. 5 (PSNR should be more than 30 dB). In this category, 200 bits of watermark information must be extracted with no error from at least three clipped images among four clipped images for each attacked image.

- Highest Image Quality
  This category targets entries with the highest average PSNR after the first compression. The watermark information (200 bits) must be detectable, allowing a small bit error rate from at least three clipped images among four clipped images for each attacked image. The average error rate should be lower than 1% ($<$ 2 bits for 200 bits of watermark information), and at the worst error rate, it should be equal to or less than 2% ($\leq$ 4 bits for 200 bits of watermark information) during the detection. A subjective assessment will be made if necessary.

3 Watermark Criteria for Videos (ver. 5)

3.1 Image Quality Assessment

The watermarked video clips should be compressed using the MPEG-4 part 10 (H.264) or 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 the luminance signal from the RGB channel according to the following equation (PSNR should be more than 30 dB).

$$Y = 0.2126R + 0.7152G + 0.0722B \tag{2}$$

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.

3.2 Tolerance Assessment

3.2.1 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. The analog output of the video equipment can be used as the digital video input for the analog video conversion (see Fig.2).

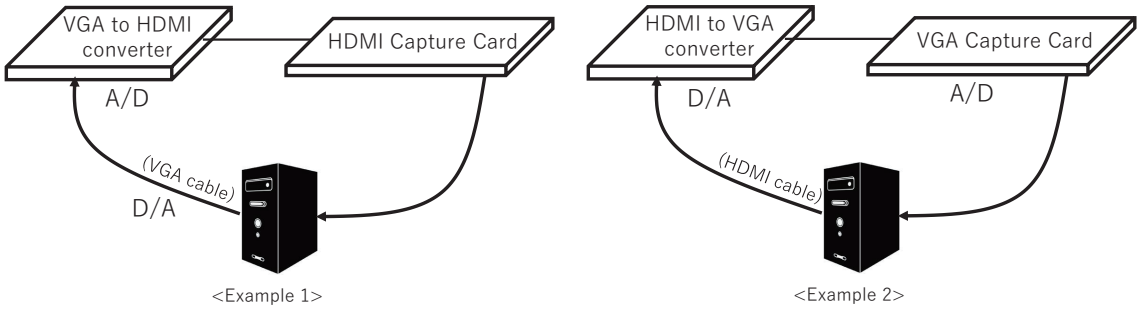

Figure 2: D/A and A/D conversion test

3.2.2 Camcorder jamming

The watermarked video without lossy compression should try to use the "camcorder jamming" test. This test performs D/A and A/D conversion using the screen and the camcorder. There are no limitations to what equipment the entrants will need for a testing (see Fig.3).

3.3 Amount of Data (Information) to be Embedded

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

3.4 Embedding and Detection of Watermark Information

No additional information can be used in the detection.

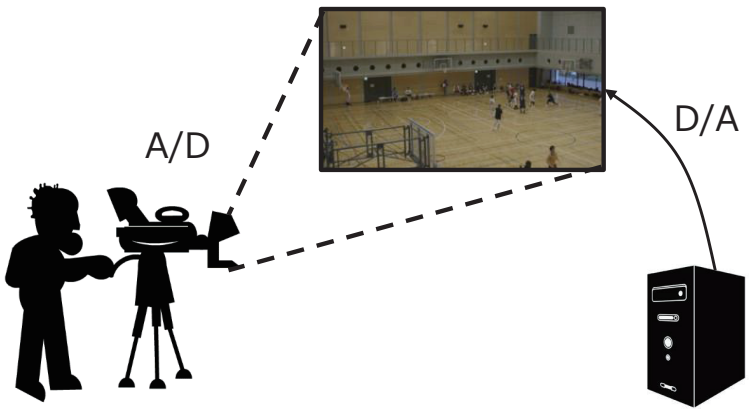

Figure 3: Camcorder jamming test

3.5 Video Clips

The applicants for the competition must use our five video clips, whose thumbnail images are shown in Fig.4. The files can be downloaded from the following URL at no charge.

`https://ds0n.cc.yamaguchi-u.ac.jp/~m.kawamu/IHC/dataset/video/`

You can see two folders, one named "`2K_RAW_16bit`" and another named "`2K_RAW_8bit`". For your convenience, a 16-bit raw image file is quantized to 8-bit-depth uncompressed AVI files. You can freely use the folder named "`2K_RAW_8bit`".

The 2K raw video clips were taken with a Canon Cinema EOS C500 system with support from Canon Inc. The IHC Committee would like to thank this company for its valuable contributions.

3.6 Content Flow

The information mentioned above should be embedded into the five HDTV video sequences specified above, and the sequences should be coded using MPEG-2 or MPEG-4 at less than 1/100 of the original HDTV bit rate (1.2 Gbps). Since the bit rate of the coded sequence is less than 12 Mbps, the average bit rate of the coded video stream should be less than 12 Mbps. The coded bit stream should be decompressed, and the decompressed 1.2-Gbps HDTV sequences should be converted into an analog video signal using a D/A converter. The analog video signal should then be converted into a digital bit stream using an A/D converter. These D/A and A/D processes are necessary since the digital HDTV content is protected by a digital rights management system. However, the content can be easily copied if the content is converted into analog format. Robust watermarking technologies must have tolerance against D/A and A/D processes. Detection of the embedded information should be tested after the A/D conversion. The volume of embedded information should be 16 bits per 15 s. The average bit error rate for the embedded information and the average PSNR for each video sequence should be calculated and included in the entry.

3.7 Information Required for Submission

- Embedding and detection algorithms

(a) Basketball

(b) Library

(c) Walk1

(d) Walk2

(e) Lego

Figure 4: Thumbnail images of our video clips.

- PSNR data after the compression process and average error rates for five video sequences. (It is not mandatory to include PSNR data after the D/A and A/D conversion and camcorder jamming. If possible, please show us the results for reference.)
- Additional data for demonstrating the advantage of the proposed method, if any

3.8 Contest Categories

- Highest Tolerance
  This category targets entries with the highest compression ratio. No error should occur during the detection.
- Highest Image Quality
  This category targets entries with the highest average PSNR. No error should occur during the detection.

4 Watermark Criteria for Audio (ver. 5)

4.1 Host Signals

Sixteen-bit linear quantization, a sampling frequency of 44.1 kHz, and stereo format should be used. Previous criteria have caused difficulties in embedding payload data into long silent periods, i.e., amplitude of zero. Therefore, the initial and final segments of zero amplitude should be removed from the audio clips from SQAM[4] (CD Tracks 27, 32, 35, 40, 65, 66, 69, and 70). Their new initial and final samples are shown in Table 3. They should be clipped from the initial sample to the final samples and should be used repetitively for a duration of 60 s.

Table 3: New initial and final samples to clip from each SQAM track.

SQAM no.	Initial	Final	Repetition is required
27	25,390	726,147	yes
32	24,910	3,145,364	no
35	24,904	2,446,802	yes
40	25,302	2,193,536	yes
65	23,937	4,803,885	no
66	22,832	642,775	yes
69	22,888	1,269,672	yes
70	23,365	733,343	yes

4.2 Payload

Ninety-bit payloads per 15 s of the host signal should be embedded, meaning that 360 bits per 60 s should be embedded. Error correction schemes can be used to embed an actual payload of 90 bits per 15 s. Random binary data to be used as the payload are available on the IHC web page[5].

4.3 Criteria for Objective Quality Degradation

PQevalAudio v2r0[6], which is an implementation of PEAQ (perceptual evaluation of audio quality) and is recommended by ITU-R BS.1387-1, should be used to measure the objective difference grade (ODG) of the eight stego signals. All of the following measurements require converting the sampling frequency from 44.1 to 48 kHz.

- Calculate the ODG between the original PCM host signal (the reference signal) and the stego signal in which the payload is embedded. The ODG should be more than -2.5.
- Calculate the ODG between the original PCM host signal (the reference signal) and the stego signal in which the payload is embedded and then compress the MP3 at 128 kbps joint stereo signal and decompress it as a degraded signal. The arithmetic mean of eight ODGs should be more than -2.0.
- If only a left- or right-channel signal is available for embedding, calculate the monaural ODG using the first method above and use it for the embedded channel signal.

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

[5] http://www.ieice.org/iss/emm/ihc/en/

[6] http://www-mmsp.ece.mcgill.ca/documents/Downloads/AFsp/

4.4 Signal Processing Attacks

The following signal processing or perceptual coding attacks should be applied to the stego signals, after which the payload should be extracted. These attacks have been confirmed to be realistic in terms of sound quality degradation of decompressed signals or of signals after inverse processing[7].

The mandatory attacks are MP3 coding and a series of attacks that mimic D/A and A/D conversions. Four of the eight optional attacks are required at the least. Changing the parameters and/or their values and/or the embedding algorithm for each stego audio is prohibited.

Mandatory

- MP3 128 kbps joint stereo (LAME ver. 3.99.3[8])
- A series of attacks that mimic D/A and A/D conversions; see Sect.4.5 for details

Optional

- Gaussian noise addition (overall average SNR 36 dB)
- Bandpass filtering from 100 Hz to 6 kHz, −12 dB/oct (filter coefficients are available on the IHC web page[2])
- Frequency scale modification (time invariant) ±4% (PICOLA[9])
- Linear speed change ±10% (ResampAudio v5r1[3])
- A single echo addition, 100 ms, −6 dB
- MP3 128 kbps (joint stereo) tandem coding
- MPEG4 HE-AAC 96 kbps (NeroAAC[10])
- A series of attacks that mimic aerial transmission; see Sect.4.6 for details

4.5 A Series of Attacks That Mimic D/A And A/D Conversions

The following signal processing steps mimic attacks using D/A and A/D conversions.

- Additive Gaussian noise at −80 dB (relative to a maximum amplitude of 16-bit quantization as 0 dB)
- Amplification of −2 dB to the above signal, followed by 16-bit quantization
- Linear speed change conversion (pitch and time-scale conversion) of −0.1 %

[7]Nishimura, A., Unoki, M., Ogiwara, A., Kondo, K.: Objective evaluation of sound quality for attacks on robust audio watermarking. In: International Congress on Acoustics 2013, POMA. vol. 19 (2013)

[8]`http://sourceforge.net/projects/lame/files/lame/3.99/`

[9]`http://www.ieice.org/iss/emm/ihc/audio/picola_tdhs2006Nov30.tar.gz`

[10]`http://www.nero.com/enu/company/about-nero/nero-aac-codec.php`

4.6 A Series of Attacks That Mimic Aerial Transmission

The following signal processing steps mimic attacks that emit a stego signal from a loudspeaker and record it by a microphone, i.e., aerial transmission and D/A and A/D conversion. This attack uses the left channel of the host signal, so that the size of the payload is half of the recommended amount.

- Convolving the room impulse response (reverberation time of 0.4 s, direct-to-reverberant energy ratio of 3.7 dB) [11] to the stego signal
- Additive extended Hoth noise (extends the highest frequency from 8 kHz to 20 kHz) [12] at −36 dB to the convolved stego signal
- Linear speed change conversion (pitch and time-scale conversion) of −0.1 %

4.7 Bit Error Rate and Criteria

The host signals should not be used in the processing for payload detection. The bit pattern of the payload should be unknown in the detection process. The detection process should require only a stego signal, i.e., it should be a "blind detection." Key data and embedding parameters that do not depend on the host signal can be used for the detection. Robustness testing should be conducted by extracting a robust payload from the modified stego signal. Forty-five seconds of the modified stego audio from which the initial sample is randomly chosen in the initial 15 s for each simulation should be used for extracting the payload, which is intended to simulate a clipping attack on the stego audio. This random clipping attack is repeated 100 times for each detection condition.

The bit error rate (BER) is defined as the averaged number of mismatched bits over 100 trials between the embedded and the extracted payloads relative to the 180 bits that are embedded into 15 to 45 s of the stego audio. BERs are recommended to calculate the embedding intensity of the averaged ODG obtained from the MP3-coded stego signals exhibited just above −2.0. If a challenger tries the optional attack of aerial transmission, the amount of payload to be extracted is 90 bits, which is embedded into the left channel of the host signal.

BERs should be calculated and reported for every combination between the host signals and the attacks. The acceptable maximum BER for all measurements is less than 10%.

4.8 Attack of Sampling Frequency Conversion

The sampling frequency conversion attack requires expansion and contraction of the duration and initial time for detection at the same rates as for conversion. For example, +10% conversion requires random selection from the initial 16.5-s samples (a 10% increase from the original) of the stego audio. The detection period is 49.5 s.

4.9 Information Required for Submission

- Embedding and detection algorithms
- ODGs between the original PCM host signals and the stego signals

[11] `http://www.ieice.org/iss/emm/ihc/audio/imp_spk_room_mic.txt`
[12] `http://www.ieice.org/iss/emm/ihc/audio/extendedHoth.html`

- ODGs between the original PCM host signals and the MP3-coded stego signals
- BERs obtained from combinations of all stego signals for two mandatory and at least four selected optional attacks

5 Contact Information

- IHC web site: `http://www.ieice.org/iss/emm/ihc/en`
- IHC Committee Secretary: `emm-ihc@mail.ieice.org`