IHC Evaluation Criteria and Competition

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

1 Watermark Competition

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.3)

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 cropping tolerance. 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 IHC Committee for quality assessment are shown Fig.1. They can be downloaded at `http://www.ieice.org/iss/emm/ihc/en/image/image.php`. They are color images with more than 10M pixels each. 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 with ITU-R BT.709, which is shown as follows.

$$Y = 0.2126R + 0.7152G + 0.00722B \tag{1}$$

The compressing 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 JPEG 2000 are candidate

[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.

compressing tools, other compressing tools can be used as long as they meet the requirements explained in Sect.2.2. 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

- Information should be embedded into the whole image, and the compressing-decompressing cycle should be performed twice. The file size should less than 1/15 the original size after the first compression, and the decompressed images should be compressed on the second compression. After the second compression, the files size should be less than 1/25 the original size. The compression ratio is determined not by the RGB files but by the YUV files.
- The files should be decompressed after the second compression. Ten HDTV-size (1920×1080) images should be cropped from each decompressed 4608×3456 image. The vertices of these cropped images are listed in Table 1. The watermark embedded in each cropped image should be detectable.
- During the review process, the IHC committee may request the detection rate for different areas.

2.3 Amount of Data(Information) to be Detected

The watermarks should be sufficiently tolerant to be detectable in no less than 200 bits in each cropped image.

Table 1: Cropping positions

Position	(x_1, y_1)	(x_2, y_2)	(x_3, y_3)	(x_4, y_4)
1	(16,16)	(1935,16)	(1935,1095)	(16,1095)
2	(1500,16)	(3419,16)	(3419,1095)	(1500,1095)
3	(2617,16)	(4536,16)	(4536,1095)	(2617,1095)
4	(16,770)	(1935,770)	(1935,1849)	(16,1849)
5	(1500,770)	(3419,770)	(3419,1849)	(1500,1849)
6	(2617,770)	(4536,770)	(4536,1849)	(2617,1849)
7	(1344,768)	(3263,768)	(3263,1847)	(1344,1847)
8	(16,1520)	(1935,1520)	(1935,2599)	(16,2599)
9	(1500,1520)	(3419,1520)	(3419,2599)	(1500,2599)
10	(2617,1520)	(4536,1520)	(4536,2599)	(2617,2599)

2.4 Embedding and Detecting of 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.

2.5 Watermark Information

The amount of watermark information to be embedded is 200 bits. The 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)		

Any error correcting code can be used to encode the watermark information as long as the 200 bits are recovered from the codeword after decompression.

Table 2: Average compression ratio, PSNR value, and MSSIM value.

	Compression ratio		PSNR [dB]		MSSIM	
	1st coding	2nd coding	1st coding	2nd coding	1st coding	2nd coding
Image 1						
Image 2						
Image 3						
Image 4						
Image 5						
Image 6						
Average						

2.6 Content Flow

The watermarking technology being entered should be used to embed the watermark information given in Sect.2.5 into all six images. After each image file has been compressed twice, the file should be smaller than 1/25 the size of the original image file. The file size percentages are based on the original file being the size of a YUV422 file. After each image file has been decompressed twice, the PSNR and MSSIM between the original and watermarked images should be calculated. The PSNR should be higher than 30 dB for the luminance signal given by Eq.(1). After the second decompression, the embedded information should be detected from HDTV-size (1920×1080) cropped images. The embedded 200 bits should be detectable.

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

2.7 Information Required for Submission

- Embedding and detection algorithms
- Compression ratio, PSNR value for the luminance signal given by Eq.(1), and MSSIM value[2] for six images (see Table 2)
- Average error rates for ten HDTV-size areas after second decompression (see Table 3)
- Additional data, if any, on robustness against other attacks

2.8 Contest Categories

- Highest Tolerance
  This category targets entries with the highest compression ratio for the six images under the conditions of the IHC standards, ver.3. No error should occur during detection.
- Highest Image Quality
  This category targets entries with the highest average PSNR after the second compression. No error should occur during detection. A subjective assessment will be made if necessary.

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

Table 3: Average error rate for ten HDTV-size areas after second decompression (%).

	Position									
	1	2	3	4	5	6	7	8	9	10
Image 1										
Image 2										
Image 3										
Image 4										
Image 5										
Image 6										

3 Watermark Criteria for Videos (ver.3)

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 given by Eq.(1). 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

After the watermarked video clips are compressed as described above, they should be decompressed, converted from digital to analog, and then converted from analog to digital. All of the embedded information should be detectable in the digitalized video. The analog output of video equipment can be used as the digital video input for the analog video conversion.

3.3 Amount of Data (Information) to be Embedded

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

3.4 Embedding and Detecting of Additional Information

No additional information can be used in the detection.

3.5 Video Clips

The video clips should come from the ITE/ARIB Hi-Vision Test Sequence 1st Edition: numbers 2, 8, 20, 23, and 46. The thumbnail images of these video clips are shown in Fig.2.

About video clips for the watermark criteria for videos: You can buy the ITE/ARIB Hi-Vision Test Sequence 1st Edition through the inquiry form of NHK Engineering System Inc.. You can find the inquiry form as the direct link tagged "Contact Us" on the NHK top page[3]. Please input your purchasing video sequence numbers in the message form in the inquiry form.

[3] `https://www.nes.or.jp/en/index.html`

(a) ITE/ARIB 2

(b) ITE/ARIB 8

(c) ITE/ARIB 20

(d) ITE/ARIB 23

(e) ITE/ARIB 46

Figure 2: Thumbnail images of ITE/ARIB Hi-Vision Test Sequence 1st Edition.

For example, "I would like to buy the sequence #2, #8, #20, #23 and #46 in the ITE/ARIB Hi-Vision Test Sequence 1st Edition."

3.6 Content Flow

The information mentioned above should be embedded into the five HDTV video sequences specified in the ITE standard movie, and the sequences should be coded using MPEG-2 at less than 1/100 the original HDTV bit rate (1.2 Gbps). Since the bit rate of the coded sequence is less than 100 Mbps, the average of the coded video stream should be less than 100 Mbps. The coded bit stream should be decompressed, and the decompressed 1.2-Gbps HDTV sequences should be converted into an analog video signal with a digital/analog (D/A) converter. The analog video signal should then be converted into a digital bit stream with an analog/digital (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 seconds. 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
- PSNR data and average error rates for five video sequences
- Additional data, if any, robustness against other attacks

3.8 Contest Categories

- Highest Tolerance
  This category targets entries with the highest compression ratio for the five video sequences under the conditions of the IHC standards, ver.3. No error should occur during detection.
- Highest Image Quality This category targets entries with the highest average PSNR. No error should occur during detection. A subjective assessment will be made if necessary.

4 Watermark Criteria for Audio (ver.3)

4.1 Host signals

Sixteen-bit linear quantization, a sampling frequency of 44.1 kHz, and stereo format should be used. Audio clips from SQAM[4] (CD Tracks 27, 32, 35, 40, 65, 66, 69, and 70) should be used repetitively for a duration of 60 s.

4.2 Payload

Ninety-bit payloads per 15 seconds of the host signal should be embedded, meaning that 360 bits per 60 seconds should be embedded. No error correction scheme should be used. 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.

[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/`

- 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 128-kbps joint stereo signal and decompress it as the degraded signal. The arithmetic mean of 8 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.

4.4 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 DA and AD conversions. Three of the seven optional attacks are required. Changing the parameters and/or their values and/or the embedding algorithm is prohibited.

Mandatory

- MP3 128 kbps joint stereo (LAME ver. 3.99.3[8])
- A series of attacks that mimic DA and AD conversions

Optional

- Gaussian noise addition (overall average SNR 36 dB)
- Bandpass filtering 100 Hz — 6 kHz, −12 dB/oct. (filter coefficients are available on 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])

4.5 A Series of The Attacks That Mimic DA And AD Conversions

The following signal processing steps mimic attacks using DA and AD conversions.

- Additive Gaussian noise at −80 dB (relative to 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 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 "blind detection." Key data and embedding parameters that do not depend on the host signal can be used for 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 seconds for each simulation should be used for extracting the payload, which is intended to simulate a clipping attack on the stego audio. The bit error rate (BER) is defined as the number of mismatched bits between the embedded and extracted payloads relative to the 180 bits that are embedded into 15 to 45 seconds of the stego audio. 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%.

4.7 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 second samples (a 10% increase from the original) of the stego audio. The detection period is 49.5 seconds.

4.8 Information Required for Submission

- Embedding and detection algorithms
- ODGs between the original PCM host signals and the stego signals
- ODGs between the original PCM host signals and the MP3-coded stego signals
- BERs obtained from combinations of the all stego signals for two mandatory and three 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`