The Collaborative Evaluation Method of Coded Picture Quality Using the Internet

(インターネットを利用した大人数による主観画質評価)

Seiichiro Hangai (member)[†], Takayuki Hamamoto (member)[†] and Hidetaka Takano[†]

Abstract In order to avoid tiresome and tedious evaluation of coded pictures, collaborative subjective evaluation method using Internet is newly proposed. Except for observation conditions, evaluating environment including evaluation time, picture presentation sequence and evaluation word is matched with that of ITU-R recommendation. Experiments open to the students in our university was continued for 4 months on the Internet to get 100 voluntary testers. According to experimental result using 14 coded pictures, the order of pictures sorted by the collaborative MOS agreed with that by the MOS of the ITU-R recommendation when 60 voluntary testers collaborated. It was also found that we could predict the MOS of the ITU-R recommendation from the collaborative MOS with 80 voluntary testers or more.

Key words: MOS, image quality, WSNR, Internet

1. Introduction

Recently, not only standard picture coding schemes such as JPEG/MPEG but also original coding schemes have been opened to the public as a free software or a shared software through the Internet. User, however, cannot know the coding efficiency based on the subjective measure, and often saves important pictures with visible distortions. Although the subjective evaluation method recommended by ITU-R¹⁾ gives reliable picture quality, the recommendation defines many conditions such as environmental, lighting, displaying and observation conditions. These conditions disturb evaluating coded picture quality simply. On the other hand, in accordance with the spread of the Internet, many pictures distorted by coding noise are easily disclosed on a browser and taking questionnaires from many persons in the world becomes possible.

From this background, we propose a new subjective evaluation method for coded pictures based on the collaboration of many indefinite volunteers²⁾. After a few months pass, the Mean Opinion Score (MOS) is informed to the client who uploads the coded picture to the server system. In this paper, the evaluating process and the experimental results are given.

Fig. 1 Collaborative picture evaluation system.

2. Collaborative Evaluation on Network

2.1 Picture Uploading

Fig. 1 shows the relationship between the client, the tester and the constructed server. The client opens the homepage (<http://peserv.hanlab.ee.kagu.sut.ac.jp/>) and uploads an original picture and a coded picture with some interactions. After getting the pictures, the server calculates the conventional SNR and the predictive MOS derived from WSNR³⁾, and informs the client of the results soon.

2.2 Tester Inviting and Evaluation Flow

The server opens the evaluated pictures to the public through Inter- or Intra- net with informing related news groups. Someone in the news group accesses the homepage. If he/she is interested in the picture evalua-

Received February 7, 2000; Accepted March 7, 2000

[†] Faculty of Engineering, Science University of Tokyo
(1-3, Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan)

Fig. 2 Evaluation flow.

tion experiment, he/she becomes a tester and starts the evaluation.

Fig.2 shows a flow of the proposed evaluation method. After checking the operation of JAVA Applet in a voluntary tester's browser, this evaluator gets display data such as size and resolution of a monitor by the Applet automatically. The evaluator calculates an optimum distance between the tester and the display monitor. Then, it requests the tester to enter Name and E-mail address. This step is essential for removing unreliable data. The number of colors of the display monitor, which is not got by the Applet, is also requested to enter.

Most important process is to adjust the brightness and the contrast of the monitor because they strongly affect subjective scores. **Fig.3** shows sinusoidal gratings of 3cpd displayed on the monitor for brightness and contrast adjustment. The dotted line presents the detectable border when the brightness of the monitor is adjusted adequately. The contrast is also adjusted using the BW reversed gratings.

After reading the explanation of "how to evaluate coded pictures", the tester starts for evaluation based on the ITU-R sequence¹⁾.

2.3 Subjective Score Processing

Fig.4 shows an example image displayed in a browser. Five pairs of pictures from 14 pairs of pictures (origi-

Fig. 3 Sinusoidal gratings for monitor adjustment.

Fig. 4 Example of displayed picture (left) and input form(right).

nal and coded) are selected for the evaluation randomly. The tester enters a subjective score by clicking button at the right input form and submits after finishing the evaluation of five pictures. From the start of evaluation to the end, it takes five minutes.

Mean Opinion Score is calculated in the server and informed the client when the number of testers reaches every tens.

3. Experiment

In order to invite voluntary testers for subjective picture evaluation, we announced our home page address (<http://peserv.hanlab.ee.kagu.sut.ac.jp>) into news group in our university. It took about 1 month for collecting 60 testers, 2 months for 80 testers and 4 months for 100 testers. **Fig.5** shows the relationship between MOS of 14 distorted pictures and the number of testers. **Table 1** summarizes kinds of the distortions.

As shown in Fig. 5, 4 (milk-as6, girl-as8, milk-xx6, girl-xxa) of 14 pictures are stable in order of MOS with different number of testers. This means that the invisible distortion does not affect subjective scores even

Table 1 Tested pictures and distortion.

No.	Picture	Distortion
1	aerial-av7	ADCT-VQ ⁽⁵⁾ , 0.7bpp
2	aerial-xxc	Random noise, 13.1 dB
3	barbara-as9	ADCT ⁽¹⁾ , 0.9bpp
4	barbara-av8	ADCT-VQ ⁽⁵⁾ , 0.8bpp
5	barbara-xxa	Random noise, 31.3dB
6	girl-as8	ADCT ⁽¹⁾ , 0.8bpp
7	girl-av5	ADCT-VQ ⁽⁵⁾ , 0.5bpp
8	girl-av8	ADCT-VQ ⁽⁵⁾ , 0.8bpp
9	girl-pe5	VQ-PE ⁽⁶⁾ , 0.5bpp
10	girl-pe9	VQ-PE ⁽⁶⁾ , 0.9bpp
11	girl-xxa	Random noise, 27.8dB
12	milk-as6	ADCT ⁽¹⁾ , 0.6bpp
13	milk-av4	ADCT-VQ ⁽⁵⁾ , 0.4bpp
14	milk-xx6	Random noise, 26.7dB

if the evaluation is done on the tester's monitor. And, low qualitative pictures, MOS of which are less than 2.6, are also evaluated correctly. However, MOS of the other 10 pictures fluctuates by the number of testers. **Fig.6** shows the relationship between variance of MOS and the number of testers. The variance of MOS evaluated under ITU-R recommendation is 0.01⁽⁷⁾. Therefore we can determine the superiority of picture quality when the number of testers is 60, and predict the final MOS of picture quality when the number of testers becomes 80.

Fig.7 and **Table 2** show the relationship between the collaborative MOS obtained from 100 voluntary testers on the Internet and the true MOS obtained from 31 testers based on the ITU-R recommendation. The experimental conditions is as follows.

- Picture size: 8x8 inch.
- Viewing distance: 7H
- Presentation order of pictures: randomized
- Presentation sequence: as shown in Fig.2
- Tested pictures: as shown in Table 1

From the result, it is found that the pictures with the collaborative MOS of 3 or more correspond to the true MOS of 4 or more. We suppose that this is caused by the difference of the distance between the tester and the display monitor. Although a recommended distance is 7H (H is the height of the monitor), many testers on the internet may watch the pictures within 4H (H is the height of pictures in this case) and can find the distortions in detail. Furthermore, we think that testers were not sure that the picture, in which any distortion was not detected, had good quality truly or not. Although the ideal relationship between tru-MOS and colb-MOS is propotional like a dotted line shown in Fig.7, the actual relationship is offset by 0.5 like a solid line.

Fig. 5 Relationship between MOS and number of testers.

Fig. 6 Relationship between variance of MOS and number of testers.

Fig. 7 Relationship between true-MOS and colb-MOS.

4. Conclusion

We have proposed a new subjective evaluation method based on scores from voluntary testers on the Internet. Because of the experimental process, it takes

Table 2 Evaluated pictures and their MOS.

No.	Picture	true-MOS	colb-MOS
1	aerial-av7	3.9	4.0
2	aerial-xxc	1.1	1.1
3	barbara-as9	4.2	3.4
4	barbara-av8	4.2	3.3
5	barbara-xxa	4.2	3.5
6	girl-as8	4.4	4.2
7	girl-av5	4.3	4.0
8	girl-av8	4.6	4.1
9	girl-pe5	3.8	2.8
10	girl-pe9	4.2	4.1
11	girl-xxa	3.2	2.3
12	milk-as6	4.4	4.3
13	milk-av4	4.3	3.9
14	milk-xx6	2.7	2.5

4 months to get the subjective score from 100 voluntary testers. From experimental result using 14 coded pictures, the order of pictures sorted by the collaborative MOS agrees with that by the MOS of the ITU-R recommendation in case of 60 voluntary testers collaboration. It is also found that the score from testers became slightly worse than that of the ITU-R recommendation. We think that this tendency is caused by the difference of viewing distance and the lack of testers' sureness in scoring. Although the collaborative MOS is less reliable than the MOS obtained by the ITU-R recommendation, the proposed method requires non special environment for evaluation, and we can get a row subjective score of coded picture from many persons. Although the current system evaluates the quality of still pictures, the method is applicable to moving pictures. In the case of moving pictures, the picture size and the number of frames should be restricted in order to reduce the data size for uploading or downloading.

[References]

- 1) ITU-R: "Recommendation 500-2", Geneva, 11, pp.165-168 (1982)
- 2) S.Hangai, T.Hamamoto, H.Takano: "The Collaborative Evaluation of Coded Picture Quality using Internet", Proc. of PCS, pp.435-437 (1999)
- 3) J.Okamoto, S.Hangai, K.Miyauchi: "A Study on Subjective Evaluation Method for Coded Moving Picture Quality", Proc. of PCS, 2, pp.519-523 (1996)
- 4) N.Doi, H.Hanyu, M.Izumita, S.Mita, Y.Eto, H.Imai: "Adaptive DCT Coding of the Home Digital VTR", GLOBECOM88, pp.1073-1079 (1988)
- 5) K.Aizawa, H.Harashima, H.Miyakawa: "Adaptive Discrete Cosine Transform Coding with Vector Quantization for Color Image", ICASSP86, pp.985-988 (1986)
- 6) S.Ito, Naito, Sato, T.Utsunomiya: "Vector Quantization of Predictive Error of Image Signal", IEICE Trans., J70-B, 3, pp.337-345 (1987)
- 7) H.Takano, T.Hamamoto, S.Hangai: "The Mass Collaborative Picture-Quality Evaluation on the Network", Tech. report of IEICE, 98, 136, pp.47-53 (1998)

Seiichiro Hangai received the B.E., the M.E. and the Dr.E. in electrical engineering from Science University of Tokyo in 1975, 1977 and 1981, respectively. Currently, he is an Associate Professor at the department of electrical engineering of Science University of Tokyo. His current research interests are image processing and speech processing.

Takayuki Hamamoto received the B.E. and the M.E. in electrical engineering from Science University of Tokyo in 1992, and 1994, respectively, and the Dr.E. in electrical engineering from the University of Tokyo, in 1997. Currently, he is a lecturer at the department of electrical engineering of Science University of Tokyo. His current research interests are image processing and computational image sensors.

Hidetaka Takano received the B.E. and the M.E. in electrical engineering from Science University of Tokyo in 1997, and 1999, respectively. In 1999, he joined Toshiba Information Systems (Japan) Corporation.