Question(s):	VQEG	**Meeting, date:**	Sep. 10-14, 2007
Study Group:	**Working Party:**	**Intended type of document** (R-C-D-TD):C	
Source:	NTT (Nippon Telegraph and Telephone Corporation), Japan		
Title:	Performance of parametric bitstream model using ES layer information		
Contact:	Mr. Keishiro Watanabe NTT Japan		Tel: +81-422-59-7204 Fax: +81-422-59-5671 Email: watanabe.keishiro@lab.ntt.co.jp
Contact:	Mr. Jun Okamoto NTT Japan		Tel: +81-422-59-6526 Fax: +81-422-59-5671 Email: okamoto.jun@lab.ntt.co.jp

Please don't change the structure of this table, just insert the necessary information.

Summary

As a trial of construction of the parametric bitstream (w/o PVS) model in the hybrid perceptual / bitstream project, we conducted subjective video quality assessment of coding distortion of HD-H.264 and investigated the relationship between ES layer parameters and subjective video quality. We report the results in this contribution.

1 Introduction

There was an agreement to include the parametric bitstream (w/o PVS) model, which estimates subjective video quality by using only bitstream information, in the scope of the hybrid perceptual / bitstream project at the Paris VQEG meeting on May 2007. The parametric bitstream model will require less computation effort than existing objective video quality assessments that use decoded video signals. However, this model would also be a less reliable method because we cannot use decoded video signals for quality estimation. Therefore, we conducted the subjective video quality assessment of coding distortion of HD-H.264 to check the level of estimation accuracy of the parametric bitstream model.

2 Discussion

2.1 Overview of subjective video quality assessment

We show an overview of a subjective video quality assessment below.

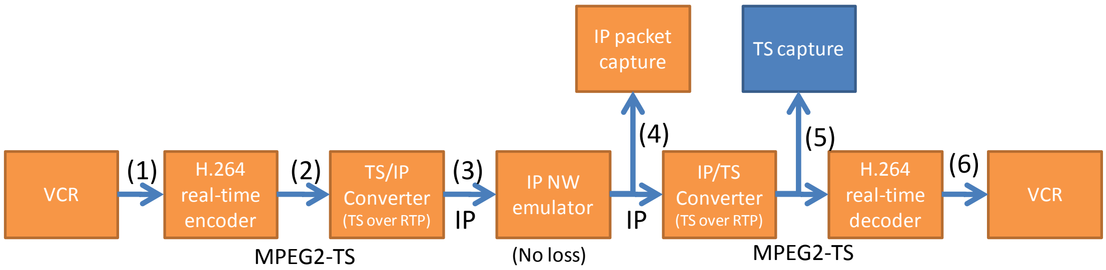

Figure 1: Experimental system

We prepared degraded video sequences according to the following procedure by using the experimental system shown in Figure 1 (Numbers in Figure 1 correspond to following procedure numbers).

(1) A video signal from a VCR was input into an H.264 real-time encoder. Input video sequences (recommended sequences in ITU-R BT.1210) are shown in Table 1.

(2) The video signal from the VCR was encoded at various bit rates on an H.264 real-time encoder and a H.264 elementary stream on MPEG2-TS was sent to a TS/IP converter. The bit-rate conditions are shown in Table 2. Each scene length was 10 seconds.

(3) The MPEG2-TS from the encoder was embedded into an IP packet stream at the TS/IP converter. This IP packet stream was sent to an IP network emulator.

(4) The IP packet stream from the TS/IP converter was sent to an IP/TS converter at the IP network emulator, but the IP network emulator did not perform any action. We captured the IP packet stream at the same time.

(5) Embedded MPEG2-TS was extracted from the IP packet stream sent from the IP network emulator at the IP/TS converter. The extracted MPEG2-TS was sent to an H.264 real-time decoder.

(6) A H.264 elementary stream in the MPEG2-TS was decoded to a video signal at the H.264 real-time decoder. This video signal was recorded at a VCR and used as a degraded video sequence in the subjective video quality assessment.

Table 1: Input video sequence (recommended in ITU-R BT.1210)

No.	Video sequence
1	Streetcar
2	Opening Ceremony

3	Crowded Crosswalk
4	Boy and Toys
5	Buildings along the Canal
6	Baseball
7	Summertime Tanning
8	Flamingos

Table 2: Bit rate conditions

No.	Bit rate (Mbps)
1	20
2	18
3	16
4	14
5	12
6	10
7	8
8	6
9	4
10	2

The subjective video quality assessment was conducted by using degraded video sequences shown in procedure (6) above. Viewing conditions are shown in Table 3. The grading scale was DSIS method (Table 4). There were 24 subjects. We derived DMOS by averaging the scores of all subjects.

Table 3: Viewing condition

Video format	HD (1440 x 1080, 59.94fps (interlace))
Monitor	16:9 32inch CRT
Viewing distance	3H
Illumination intensity	200 lx

Table4: Relationship between value and description on DSIS method

Value	Description
5	Imperceptible
4	Perceptible, but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

2.2 Analysis of bitstream

For bitstream analysis, we used an elementary stream of H.264. This elementary stream was demultiplexed from MPEG2-TS, which is extracted from an IP packet stream captured in procedure (4) of section 2.1. In particular, we paid attention to some ES layer parameters, which express the scene characteristics (difficulty of coding).

2.3 Analysis of relationship between result of experiment and ES layer parameters.

We tried to predict subjective video qualities from some ES layer parameters, which were derived by using video sequences shown in Table 1. We compared predicted subjective video qualities with experimental which were measured in subjective assessment. The results are shown in Figure 2. The correlation coefficient between predicted and experimental video qualities was 0.97 and the RMSE was 0.32. This result is only for reference because degraded conditions are not uniform perfectly, but we can hope that the accuracy of estimating subjective video quality is suitable for our purpose.

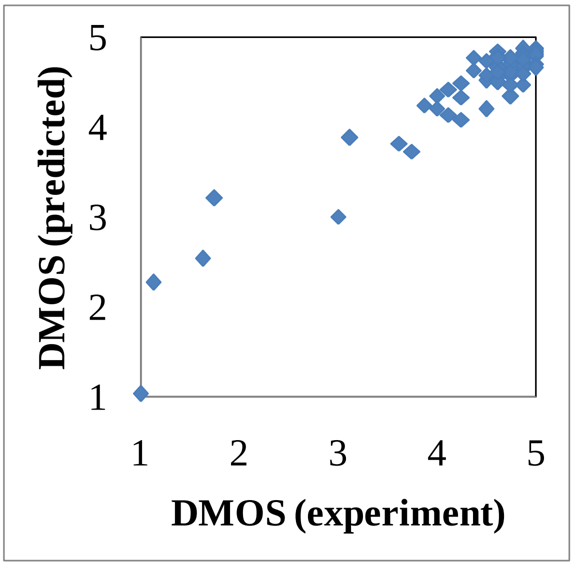

Figure 2: DMOS (predicted) vs. DMOS (experiment)

Moreover, we tried to adapt this quality assessment model to unknown video sequences. An unknown video sequence list is shown in Table 5. Results of the estimation are shown in Figure 3.

Table 5: Video sequence (for verification) (recommended in ITU-R BT.1210)

No.	Video sequence
9	European Market
10	Harbour Scene
11	Whale Show
12	Soccer Action
13	Green Leaves
14	Japanese Room
15	Ice Hockey
16	Weather Report

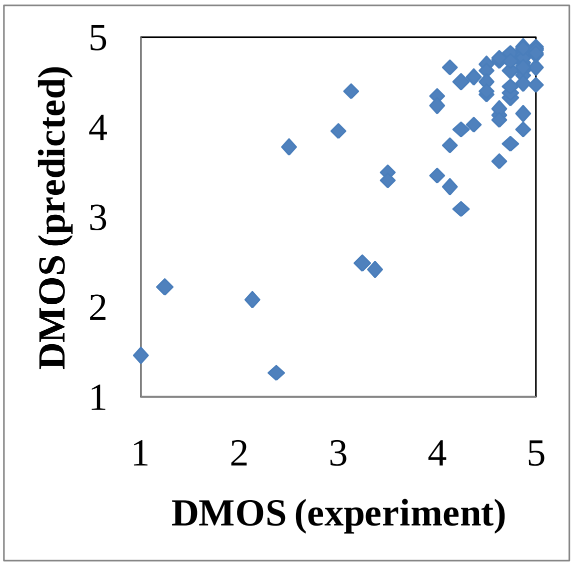

Figure 3: DMOS (predicted) vs. DMOS (experiment)
(Unknown video sequence)

The correlation coefficient was 0.94 and the RMSE was 0.46. This result was not good, but a rough trend could be estimated though we used the very simple ES layer parameters, and obtaining those ES layer parameters was very easy.

2.4 Analysis of relationship between subjective video qualities and bit-rate.

For the purpose of comparison, we tried to estimate DMOS from only bit-rate. The parameters of this parametric bitstream model were derived by using video sequences shown in Table 1 and target video sequences for assessment were the same. The relationship between predicted and experimental video qualities is shown in Figure 4. The correlation coefficient was 0.88 and the RMSE was 0.61. This score was very bad; especially there was the region which ranges from 1.1 to 4.6 (DMOS (experiment)) at the same bit-rate by the effect of scene characteristics.

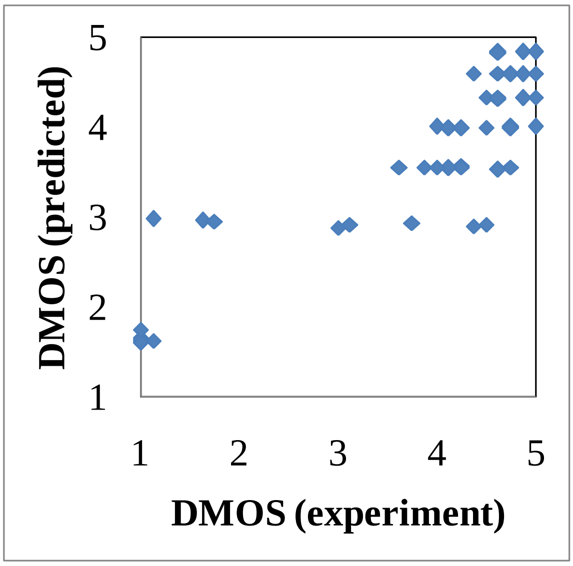

Figure 4: DMOS (predicted) vs. DMOS (experiment)

2.5 Conclusion.

We constructed the simple parametric bitstream model for estimating coding distortion of H.264 by utilizing syntax information in elementary stream, which can be obtained easily from a bitstream. This parametric bitstream model could estimate the rough trend of subjective video quality. We also constructed the simple parametric bitstream model using bit-rate parameter, but this kind of parameters cannot consider scene characteristics and the estimation accuracy was very bad as a result.

3 Proposal

We propose to use syntax information in elementary stream, which can consider scene characteristics (coding difficulty) in the parametric bitstream model.