14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

QOS MAPPING FOR FINE GRANULAR SCALABILITY WITH BASE LAYER SCALING

Masayuki Inoue, Katsuhiko Fukazawa, Shigetaro Iwatsu, and Yoshiyuki Yashima

NTT Cyber Space Laboratories, NTT Corporation
1-1 Hikari-no-oka, Yokosuka-shi, Kanagawa, 239-0847 Japan
E-mail:{inoue.m, fukazawa.katsuhiko, iwatsu.shigetaro, yashima.yoshiyuki}@lab.ntt.co.jp

ABSTRACT

Scalable video coding is receiving increasing attention. However, few papers address QoS mapping with regard to scalable video coding. This paper uses Fine Granular Scalability with base layer scaling that offers application-level QoS control. An experiment is conducted that uses the DSIS method to subjectively assess Fine Granular Scalability. The results show that the maximum of the mean grading point fell as the base-layer Size value decreased. Our results also show that Fine Granular Scalability provides only SNR scalability, not subjective image quality scalability. A multiple regression analysis shows that we can establish QoS mapping between user and application-levels by using both application-level QoS parameters and a feasible human factor. Furthermore, we show that FGS can adjust to not only a wide range of channel capacities but also a wide variety of users by using the indicated QoS mapping.

1. INTRODUCTION

Streaming video over the Internet has been rapidly adopted over the past few years. Also, many different types of networks are now converging on the all-IP network and the Internet. Moreover, network capacities are increasing radically. To offer successful streaming video services, the Internet must have some kind of QoS control at each level of the protocol stack and we have to remember that users have a wide variety of profiles. Tasaka considered the QoS at each level of the protocol stack. He identified six kinds of QoS: (1) physical-level QoS, (2) node-level QoS, (3) network-level QoS, (4) end-to-end-level QoS, (5) application-level QoS, and (6) user-level QoS (or perceptual QoS)[1]. Among the six levels of QoS, the user-level QoS is the most important since the users are the true recipients of the service. However, many studies on QoS mapping treat only the transfers up to the end-to-end level QoS[2, 3, 4, 5, 6, 7, 8].

This paper uses MPEG4 FGS (Fine Granular Scalability)[9] with base layer scaling to implement application-level QoS control, because of its ability to adjust to a wide range of network channel capacities. However, FGS provides only SNR scalability, not subjective image quality scalability. This is unfortunate since the latter is more important than the former; the users are the true determiners of service acceptance. Furthermore, the control of bitrate over FGS bitplane is not user friendly, because users can not imagine the subjective image quality from the bitrate.

No paper has considered in detail how to map the application-based QoS level to user-satisfaction level when using FGS. We rectify this omission by conducting an experiment in which the DSIS(double-stimulus impairment scale) method[10] is used to subjectively assess FGS. We consider QoS mapping relations between application-level parameters, such as FGS codec parameters, and user-satisfaction level parameters, such as subjective image quality grading points[11]. Furthermore, this paper describes how to control the FGS bitplanes to support a wide variety of users through the human factors, i.e. user profiles.

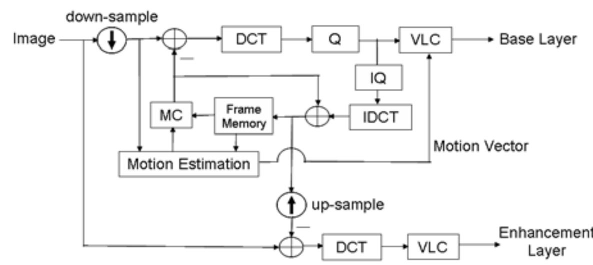

Figure 1: FGS encoder with base layer scaling.

2. FINE GRANULAR SCALABILITY WITH BASE LAYER SCALING

FGS is a method of encoding a sequence as a base layer and an enhancement layer(or FGS layer). The enhancement layer can be truncated during or after encoding (reducing both the bitrate and the decoded quality) to give highly flexible control over the transmitted bitrate.

Figure 1 shows a block diagram of an FGS encoder with base layer scaling. In the Base Layer, the Image(current frame) is down-sampled to form the lower spatial resolution image, subtracted from the down-sampled previous frame, transformed with DCT, quantized, and finally encoded. The quantized coefficients are inverse-quantized, and these inverse-quantized coefficients are transformed with IDCT. The transformed image is up-sampled to form the prediction for the original resolution image in the enhancement layer. The prediction image is subtracted from the original image and transformed with DCT to give a set of difference coefficients. The difference coefficients for each block are encoded as a series of 'bitplanes'. First, the residual coefficients are reordered in a zigzag scan. The highest-order bits of each coefficient are encoded first (the MS(Most Significant) bitplane) followed by the next highest-order bits and so on until the LS(Least Significant) bits have been encoded.

3. EXPERIMENT METHODOLOGY

We examined the relation between user and application-levels by assessing the quality of decoded videos in terms

Table 1: Questionnaire.

human factors	questions
gender	male or female?
age	twenties or thirties?
imagery	Please imagine yourself seeing an image. Which type of image do you imagine, TV, MOVIE, DVD, or others?
interest	Do you have an interest in image quality?

Table 2: Coding specifications.

item	mode
Base Layer	MPEG4 ASP(YUV)
Enhancement Layer	MPEG4 FGS(Y)
I-VOP	1 in every 30 frames
Quantization	H.263

of the QoS parameters at the application-level.

3.1 Experimental Model

In this paper, we examine the effect of subjective image quality using the simplified model. For QoS mapping between user-level and application-level, we consider that subjective image quality consists of three elements: 'encoder', 'decoder', and 'user'. In this case, we use 'Size value(Table 3)' and 'Base bitrate(Base Layer bitrate)' as the parameters for encoder-element, 'FGS bitplane' and 'FGS bitrate (Enhancement Layer bitrate)' as the parameters for decoder-element, and 'Gender', 'Age', 'Imagery', and 'Interest' as the user-element parameters, which are human factors as is shown in Table 1. For application-level QoS, we adopt 4 parameters: 'Size value', 'Base bitrate', 'FGS bitplane' and 'FGS bitrate'. For user-level QoS, we adopt the single parameter of 'grading point'[10].

3.2 Subjects

The 24 subjects, 12 males and 12 females, had no previous experience in evaluating images. To obtain information on human factors, we used the questionnaire shown in Table 1.

3.3 Assessment Environment

Each subject had his/her own LCD(Liquid Crystal Display) monitor and a grading input device. The distance from the LCD to the subject was 4H, where H means image height which equals 25 cm.

3.4 Methodology

To conduct the assessment from the user-level QoS point of view, we used the DSIS method. The reference video (the original video), and the test decoded-video were presented only once. After each presentation, the subject pushed a button to identify the grading scale with regard to perceivable image degradation. Grades 5, 4, 3, 2, and 1 mean 'imperceptible', 'perceptible, but not annoying', 'slightly annoying', 'annoying', and 'very annoying', respectively. We used three

Table 3: Size values for down and up-sampling.

down-sampling → up-sampling	Size values
CIF → QQCIF → CIF	1
CIF → QCIF → CIF	2
CIF → CIF → CIF	4

Table 4: Parameter sets for Susie.

Size values	FGS bitplanes
1	1(M),3,5,7,9(L)
2	1(M),2,3,5,7(L)
4	1(M),2,3(L)

(M):MS bitplane (L):LS bitplane

Table 5: Parameter sets for Foreman and Carphone.

Size values	FGS bitplanes
1	1(M),3,5,7,9(L)
2	1(M),2,4,6,8(L)
4	1(M),2,3(L)

(M):MS bitplane (L):LS bitplane

original videos (CIF size): 'Susie(frame no.:1-150)', 'Foreman(frame no.:91-240)', and 'Carphone(frame no.:91-240)'. Table 2 shows the coding specifications. Each original video was coded as per Table 2. Table 4 and Table 5 show decoding parameter sets for the test decoded-videos. Decoding was performed using the parameter sets of Size values and FGS bitplanes. The three original videos were multiplied by the 13 parameter sets to yield 39 test decoded-videos. In this paper, an FGS bitplane value of 1 means that only the MS bitplane of the coded video was decoded. A Size value of 1 and an FGS bitplane value of 9 means that all data up to and including the LS bitplane was decoded.

4. RESULTS AND DISCUSSION

The experiment yielded 936 results from the 24 subjects and 39 test decoded-videos.

4.1 Fine Granular Scalability Performance

Figure 2-Figure 4 show the mean grading point of each set of Size values and FGS bitplanes. ANOVA(analysis of variance) showed that there were significant differences in the Size value($F(2,920)=1592$, $p < 0.05$) and FGS bitplanes($F(8,920)=256$, $p < 0.05$), where F is F-ratio. We observe that the larger the base-layer Size value is, the larger the maximum of the mean grading point is.

Figure 5-Figure 7 show the mean PSNR of each set of Size values and FGS bitplanes. ANOVA showed that there were significant differences in the Size value($F(2,920)=23421$, $p < 0.05$) and the FGS bitplanes($F(8,920)=6770$, $p < 0.05$). The tendency is that the bigger the FGS bitplane is, the bigger the mean PSNR is, which is different from the trend in Figure 2-Figure 4.

Figure 8-Figure 10 show the mean FGS bitrate of each set

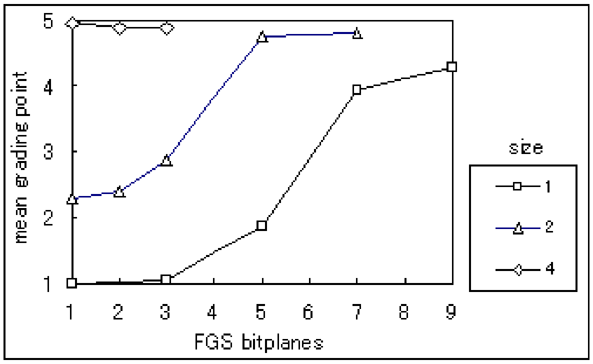

Figure 2: Comparison of grading point for Susie.

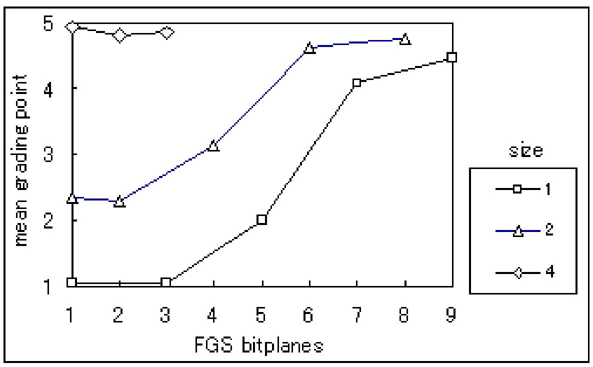

Figure 3: Comparison of grading point for Foreman.

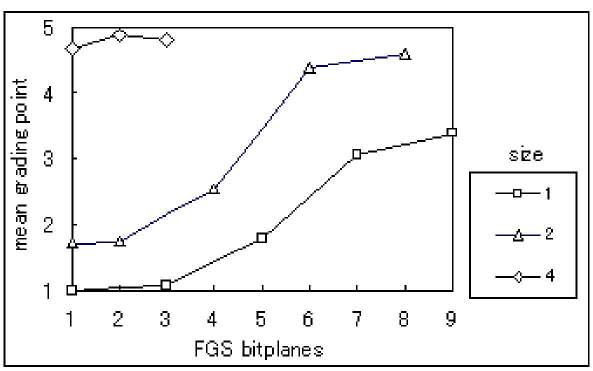

Figure 4: Comparison of grading point for Carphone.

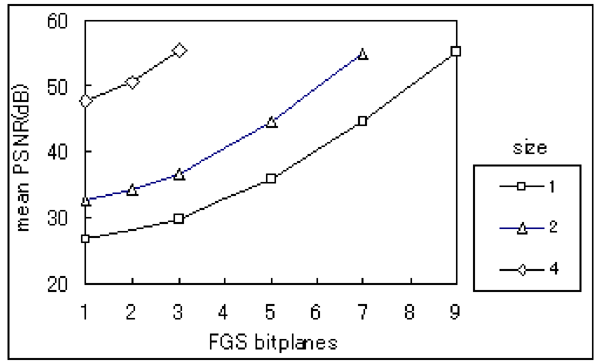

Figure 5: Comparison of PSNR for Susie.

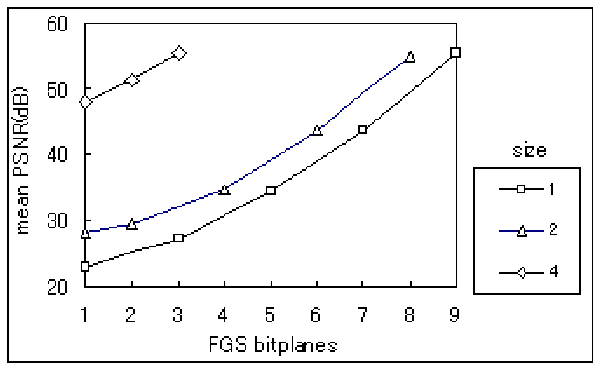

Figure 6: Comparison of PSNR for Foreman.

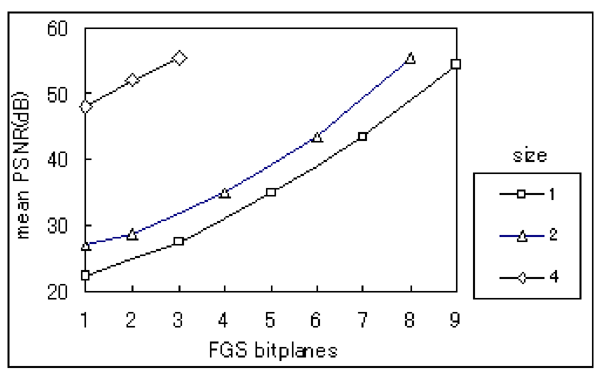

Figure 7: Comparison of PSNR for Carphone.

of Size values and FGS bitplanes. ANOVA showed that there were significant differences in the Size value($F(2,920)=4316$, $p < 0.05$) and the FGS bitplanes($F(8,920)=8274$, $p < 0.05$). We observe the tendency in which the bigger the FGS bitplane is, the bigger the mean FGS bitrate is. Comparing FGS bitplane 7 with FGS bitplane 9 at Size value of 1, 9's bitrate is more than twice that of 7. However, Figure 2-Figure 4 show that 7 and 9 have almost the same mean grading point. This means that the FGS bitplane value impacts SNR but not subjective image quality. Therefore, it is important to determine the QoS relation between application-level and user-level.

4.2 QoS mapping Application-level to User-level

We carried out principal component analysis using the 'Size value', 'FGS bitplane', 'Base bitrate', and 'FGS bitrate'. The result was that the first two principal components had characteristic-values of more than one; accordingly, we adopted the first and second principal components. Table 6 shows the principal component loading. Based on these results, we adopted two sets of QoS parameters as predictor variables in multiple regression analysis. Owing to multicolinearity, we put Size value and FGS bitplane in one set

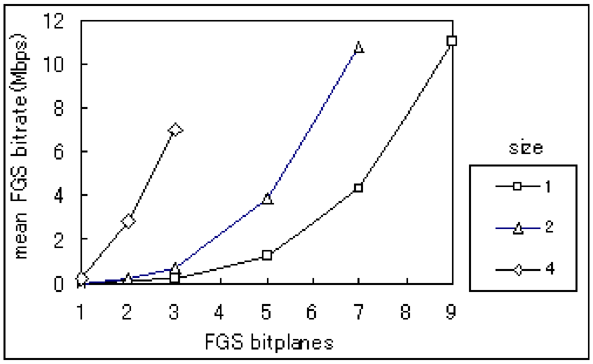

Figure 8: Comparison of FGS bitrate for Susie.

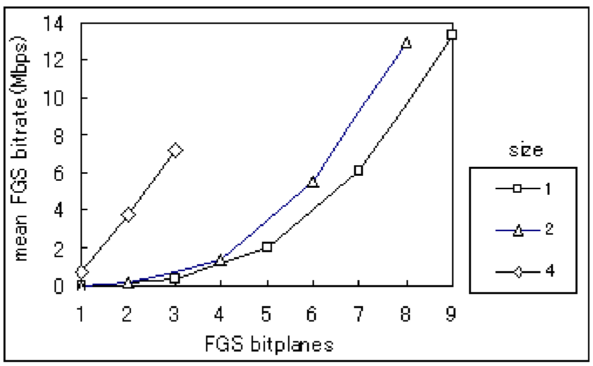

Figure 9: Comparison of FGS bitrate for Foreman.

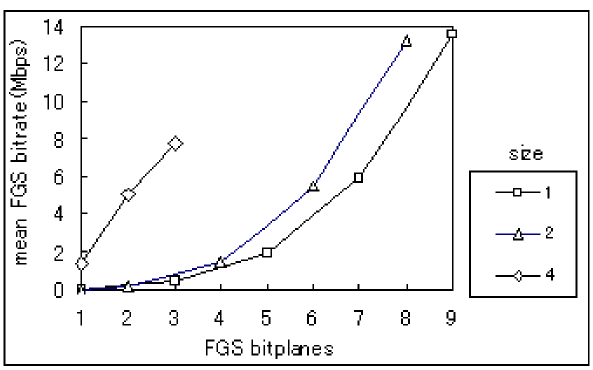

Figure 10: Comparison of FGS bitrate for Carphone.

and Base bitrate and FGS bitrate in the other. This yielded

$$GP = -1.003 + 1.258SZ + 0.425FP \quad (1)$$

$$GP = 1.447 + 0.365BB + 0.198FB \quad (2)$$

where GP, SZ, and FP stand for Grading point, Size value, and FGS bitplane, respectively, and BB and FB mean Base bitrate and FGS bitrate, respectively. The contribution rates adjusted for the degrees of freedom for (1) and (2), are 0.762 and 0.581, respectively. This means that regression equation (1) is a better model than regression equation (2). Therefore,

Table 6: Principal component loading.

	loading	
Image quality factor	first	second
Size value	0.970	-0.107
FGS bitplane	-0.354	0.913
Base bitrate	0.972	-0.059
FGS bitrate	0.109	0.977

Table 7: Questionnaire results.

Sub. No.	gender	age	imagery	interest
1	male	30s	TV	yes
2	male	20s	others	no
3	female	20s	TV	no
4	female	30s	TV	yes
5	male	30s	TV	no
6	male	30s	DVD	yes
7	female	20s	DVD	yes
8	female	20s	MOVIE	no
9	female	30s	TV	no
10	male	30s	TV	no
11	male	20s	TV	no
12	male	20s	TV	yes
13	female	30s	DVD	no
14	male	20s	MOVIE	yes
15	female	20s	TV	no
16	female	20s	TV	yes
17	female	30s	TV	yes
18	male	20s	TV	yes
19	female	20s	DVD	no
20	female	30s	MOVIE	yes
21	female	30s	TV	no
22	male	30s	TV	yes
23	male	20s	TV	yes
24	male	30s	TV	yes

we can accurately estimate subjective image quality using the two application-level QoS parameters of Size value and FGS bitplane.

4.3 Human Factor

The questionnaire results are shown in Table 7. Since these results are qualitative, we transformed them into quantitative values to permit multiple regression analysis, as is shown in Table 8. The human factors 'gender', 'age', 'imagery', and 'interest' were assigned dummy variables 'g', 'a', 's1,s2,s3', and 'k', respectively. We then performed multiple regression analysis using the dummy variables as the predictor variables. We took the human factors of 'gender', 'age', 'imagery', and 'interest' into account as follows: The multiple regression equation that considers g, a, s1, s2, s3, and k became

$$\begin{aligned} GP &= -1.398 + 1.258SZ + 0.425FP + 0.194g + 0.001a \\ &\quad + 0.442s1 + 0.488s2 + 0.479s3 - 0.256k \end{aligned} \quad (3)$$

Table 8: Dummy variables.

Sub. No.	gender	age	imagery			interest
-	g	a	s1	s2	s3	k
1	1	0	1	0	0	1
2	1	1	0	0	0	0
3	0	1	1	0	0	0
4	0	0	1	0	0	1
5	1	0	1	0	0	0
6	1	0	0	0	1	1
7	0	1	0	0	1	1
8	0	1	0	1	0	0
9	0	0	1	0	0	0
10	1	0	1	0	0	0
11	1	1	1	0	0	0
12	1	1	1	0	0	1
13	0	0	0	0	1	0
14	1	1	0	1	0	1
15	0	1	1	0	0	0
16	0	1	1	0	0	1
17	0	0	1	0	0	1
18	1	1	1	0	0	1
19	0	1	0	0	1	0
20	0	0	0	1	0	1
21	0	0	1	0	0	0
22	1	0	1	0	0	1
23	1	1	1	0	0	1
24	1	0	1	0	0	1

The contribution rate adjusted for the degrees of freedom for (3) became 0.768. Regression equation (3) has a slightly better fit than regression equation (1). We make the following observations from (3). First, GP increases as Size value and FP increase. Next, the coefficients of s1, s2, s2, and k are slightly larger than those of the others. This means that the human factors of ’imagery’ and ’interest’ impact GP more than the others. Also, we can transform equation (3) as follows:

$$GP = BGP + FGP \tag{4}$$

$$BGP = -1.398 + 1.258SZ \tag{5}$$

$$\begin{aligned} FGP &= 0.425FP + 0.194g + 0.001a + 0.442s1 \\ &\quad + 0.488s2 + 0.479s3 - 0.256k \end{aligned} \tag{6}$$

where BGP and FGP stand for Base grading point and FGS grading point, respectively. Furthermore, we can also transform equations (5) and (6) into equations (7) and (8), respectively.

$$SZ = (BGP + 1.398)/1.258 \tag{7}$$

$$\begin{aligned} FP &= (FGP - (0.194g + 0.001a + 0.442s1 \\ &\quad + 0.488s2 + 0.479s3 - 0.256k))/0.425 \end{aligned} \tag{8}$$

Equation (7) indicates that Size values can be determined if you know the Base grading point required by the user. Equation (8) means that the FGS bitplanes can be adjusted through the FGS grading point required by the user and the user’s human factors. This indicates that we can control Size values and FGS bitplanes in compliance with the users’ specified subjective image quality.

Note that it is easier to control the subjective image quality control via the FGS bitplane than by bitrate, because users can not imagine the subjective image quality from a bitrate. Equation (8) also indicates that we can control the FGS bitplane term so as to well support a wide variety of users according to the human factors, i.e. user profiles. In consequence, FGS can be adjusted freely to support not only a wide range of channel capacities but also a wide variety of user demands.

5. CONCLUSIONS

We conducted an experiment using the DSIS method to subjectively assess Fine Granular Scalability. As a result, we observed that the maximum of the mean grading point fell as the base-layer Size value decreased. By using multiple regression analysis, we obtained a QoS mapping relationship between user and application-levels through the use of both application-level QoS parameters and the human factors. Furthermore, we showed that we can control Size values and FGS bitplanes in compliance with the users’ specified subjective image quality. Finally, using the QoS mapping so identified, we showed that FGS is a good approach to supporting not only a wide range of channel capacities but also a wide variety of user demands.

REFERENCES

[1] S. Tasaka and Y. Ishibashi, “Mutually compensatory property of multimedia QoS,” in Conf. Rec. IEEE ICC2002, pp.1105–1111, 2002.

[2] K. Nahrstedt and J. M. Smith, “The QoS broker,” IEEE Multimedia, vol. 2, no. 1, pp.53–67, Spring 1995.

[3] J.-F. Huard and A. A. Lazar, “On end-to-end QOS mapping,” in Proc. IWQoS’97, May 1997.

[4] J.-F. Huard and A. A. Lazar, “On QOS mapping in multimedia networks,” in Proc. COMPSAC’97, Aug. 1997.

[5] K. Nahrstedt and J. M. Smith, “Design, implementation and experiences... ,” IEEE J. Sel. Areas in Commun.,, vol. 17, no. 7, pp.1263–1279, Sep. 1996.

[6] P. Francis-Cobley and N. Davies, “Performance implications of QoS mapping...,” in Proc. IEEE Int. Conf. ATM (ICATM’98), pp.529–535, Jun. 1998.

[7] L. A. DaSilva, “QoS mapping along the protocol stack: Discussion and preliminary results,” in Conf. Rec. IEEE ICC2000, vol. 2, pp.713–717, Jun. 2000.

[8] T. Yamazaki and J. Matsuda, “Adaptive QoS management for multimedia...,” IEICE Trans. Commun., vol.E82-B, no.11, pp.1801–1807, Nov. 1999.

[9] W. Li, “Overview of fine granularity scalability in MPEG-4 video standard, ” IEEE Trans. Circuits Syst. Video Technol., vol.11, pp.301–317, 2001.

[10] ITU-R Recommendation BT.500-11:“Methodology for the Subjective Assessment of the Quality of Television Pictures”, 2002.

[11] M. Inoue, K. Jinzenji, K. Fukazawa, S. Iwatsu, and Y. Yashima, “A Study on QoS Mapping for SNR Scalable Video Coding,” in Proc. IWAIT2006, S03-7, pp.105–110, 2006.