Quality Measurement Modeling on Scalable Video Applications

Sung Ho Jin, Cheon Seog Kim, Dong Jun Seo, and Yong Man Ro*
Image and Video Systems Lab.
Information and Communications Unversity (ICU)
Deajeon, Korea

Abstract— **For various mobile applications, measuring the grade of the video quality is needed in order to guarantee the optimal quality of video streaming service. As H.264/AVC scalable video coding (SVC) has been emerged and developed to support full scalability including spatial, temporal, and signal-to-noise ratio (SNR) scalability, each of which shows different visual effect, it is necessary to measure video quality with full scalability. In this paper, we develop a novel video quality metric allowing full scalability through the subjective quality assessment. Experimental results show that the proposed quality metric has high correlation with subjective quality and is useful to determine the video quality of SVC.**

Keywords—SVC; QoS; video quality metric; full scalability

Topic area—multimedia communication

I. INTRODUCTION

Nowadays, network environments have been changing in wide spectrum. With the conversion of heterogeneous networks, multimedia consumption environments have also become diverse, e.g., digital multimedia broadcasting (DMB), mobile video streaming, etc. There is a need for video streaming techniques adapting to the change of the consumption environments to maintain quality of service (QoS) [1]. For an effective multimedia service, it is obvious that QoS has to be affected by a suitable coding technique, an available network bandwidth, capability of devices, etc. Especially, an efficient video coding technique is an important factor to guarantee high-quality video services.

Currently, H.264/AVC scalable video coding (SVC) is being progressed in JVT (joint video team) to satisfy the various demands of increased multimedia content consumption [6]. SVC supports temporal, SNR, and spatial scalability with high coding efficiency, so that it can perform adaptive video streaming on various consumption environments. SVC compresses a raw video into multiple bitstreams consisting of one base bitstream and enhancement bitstreams so that it can extract and transmit adequate bitstreams from the coded bitstreams without re-encoding for video adaptation. The quality of SVC video is decided by the extracted bitstreams.

*Prof. Yong Man Ro is the director of the Image and Video Systems Lab. at ICU, (e-mail: yro@icu.ac.kr, Tel. +82-42-866-6289)

The quality metric in SVC has two important conditions. The first one is that because QoS is decided by a variable coding parameter combination consisting of various frame rates, qualities, and spatial resolutions, the metric provides optimal extraction parameters in terms of three aspects. The second is that the metric should contain the characteristics of human-perception in scalable video environments.

Measurement of video quality is performed by subjective and objective methods. Subjective quality method is based on human perception that comes from asking a degree of quality to assessors. It is close to the human perception; however it is very costly and time consuming. On the other hand, objective method, such as peak signal-to-noise ratio (PSNR) or mean square error (MSE), has large difference for human perceptions, but it gives no resources consuming as well as makes measurement in real-time.

Since objective metrics do not reflect human perception effectively, it is necessary to devise a quality metric of human visual properties [1]. Conventional work for video quality mainly focus on the analysis of codec difference [2], the relationship between frame rate and quality in given bitrate budget [3], video quality study using semantic concept [1], and human vision-based study [4]. There also exist various works analyzing coding characteristics such as frame rate, SNR, spatial resolution, and bit rate budget. For instance, a quality metric analyzes the effect of frame rate and bit rate under the low-rate and low-resolution condition [7]; [8] specifies various features of contents for a quality metric; in [9], a quality metric using frame rate and motion feature is suggested. Most of studies are limited to apply temporal and SNR scalability, even though mobile applications are widely used. Several studies combining both the methods, therefore, are being progressed. Standard organizations such as ITU-R, WP 6Q, ITU-T, VQEG also are developing quality metrics.

In this paper, therefore, we propose a new quality metric reflecting variable frame rate, SNR, and spatial resolution.

II. SUBJECTIVE ASSESMENT ON VIDEO QUALTIY

A. DSCQS Test

Double stimulus continuous quality scale (DSCQS) not only shows the best performance among the subjective assessment methods recommended in ITU-T BT.500 [5], but also estimates a degree of impairment accurately. It minimizes contextual effects, which occur when the subjective

assessment of an image or video is influenced by the order and severity of impairments during the test session [5]. Currently, the method is also being widely used as a basis of measurement for standardizing objective quality methods. In this work, therefore, the DSCQS method is adopted for measurement and modeling of quality metrics. In Fig. 1 (a), assessors face each pair of both reference and impaired sequences twice (they are labeled *A* and *B*.), where the order of the pair (*AB* or *BA*) is determined randomly for each trial. The score of each trial is graded on a five-point scale and then normalized into "0" to "100" scale as shown in Fig. 1 (b).

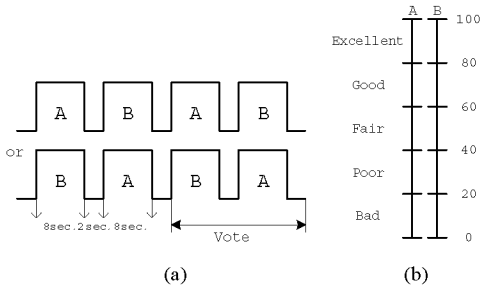


Figure 1. The scheme of DSCQS method: (a) presentation sequences and (b) rating scales

The main goal of DSCQS is to obtain the differential mean opinion score (DMOS) between the reference and test sequences averaged by all of the viewers. The task is to assess the degradation of the test sequences for the references. A score of "0" and "-100" on the subjective DMOS denotes the highest and the most impaired quality, respectively.

Before the subjective test, the assessors need to fully understand the test and the DSCQS method. Due to the volume of the test and assessors' fatigue, this subjective assessment is divided into three sessions. In each session, a demonstration is firstly held, showing test sequences with various qualities including the best and worst qualities from high to low resolution, respectively. The sequences are successively presented for 2 seconds each to prevent severe bias from careless rating out of 100 scale rating. When each session comes to end, assessors should take a rest for a minute. Again they repeat the first step and begin with the next session.

B. Subjective Assessment Analysis

The original source is used as the reference sequence and the test sequences are impaired forms of the original one in aspects of temporal, SNR, and spatial resolution. Here, the original source has a resolution of CIF, a frame rate of 30 frames per second (fps), and a length of 8 seconds. The test sequences are encoded with three different frame rates of 15, 7.5, and 30 fps, for temporal scalable variation. For each frame rate, three values of the quantization parameter (QP) are selected; 32, 37, 42 for 30 fps; and 23, 29, 35 for 15 and 7.5 fps which provide assessors with discriminative video qualities. The 15 and 7.5 fps sequences repeat frame overlapping and dropping once and three times, so that the total number of frames in every sequence becomes 240 frames.

Since this work focuses on mobile applications, we use 6 spatial resolutions between CIF and QCIF, to measure video quality in spatial domain. Each video sequence with different spatial resolution maintains an aspect ratio of 1.22:1 to reveal an identical visual effect. The viewing distance is fixed to five times of the original video height. 162 sequences are used for the training of video quality metrics and three baseline sequences are encoded by JSVM 5.5 [6]. When measuring the perceptive degree of the impairment on all the test sequences, 18 undergraduates participated as subjects in this test. All assessors have normal visual power and color vision.

Prior to the test, we consider the characteristics of content itself. To minimize interference by decrease of the spatial resolution, we choose training sequences that have similar spatial and textural details. The degree of spatial and textural details is measured by an MPEG-7 edge histogram [10]. The "*Football*," "*Foreman*," and "*Paris*" sequences have values of 0.15, 0.13, and 0.15, respectively. If the average textural value of one sequence is high, we regard the sequence as a video with detailed textures, and vice versa. To analyze motion characteristics more accurately, we additionally select the sequences which have different kinds of motion characteristics, e.g., low and fast motions. The MPEG-7 motion activity is used for obtaining the motion features to select the training sequences. Motion activity values of "*Football*," "*Foreman*," and "*Paris*" are 3.73, 2.70, and 1.34, respectively. If motion activity is greater than 2.0 it is categorized as fast motion, otherwise it has slow motion.

From the assessments, the relationships between spatial resolution or PSNR and the subjective DMOS are obtained. Under the condition of fixed QP (SNR) and frame rate, most assessors tend to prefer the sequences with high spatial resolution to low spatial resolution in every test. Moreover, with the fixed spatial resolution and frame rate, there exist high correlations between the PSNR and the subjective DMOS. But, the PSNR shows a low correlation with the subjective DMOS in the case where the frame rate, quality, and spatial resolution are applied jointly. Therefore, this result shows that a PSNR-based quality metric cannot reflect full scalability.

In terms of the frame rate, the case of full frame rate (30 fps) is preferred over half- and quarter-frame rates. The half-frame rate scores a high preference regarding a sample with a slow motion. However, the viewers avoid watching the sequences with the quarter-frame rates due to frame-skipping, regardless of slow motion. In order to investigate the effects of motion characteristics and spatial resolution, we perform an analysis where the frame rate is fixed and where there are different motion sequences.

III. Quality Metric Model For Full Scalablity

When a quality metric has a linear property with subjective quality, it can be regarded as human perception-reflected metric. In order to develop a quality metric in scalable video environments, it should analyze the parameters that affect to a scalable video. In this work, we describe the modeling of the quality metric as two steps: the first step is to seek the metric between temporal and SNR scalability and the other is to add spatial scalability. The first step uses the frame rate, SNR, and motion characteristics and yields the quality metric containing temporal and SNR scalability. Since the human perception for the full scalability depends on the motion property of video [9], this modeling uses the motion feature, which is one of the key parameters to lead the quality metric. The second step accepts and adds the variation of spatial resolution into the metric obtained from the previous step, and thereby the quality metric

for full scalability is established.

The aim of this work is to seek the quality metric model to accord with the experimental results on subject quality, *SQ*. For a range [0, 100] of the *SQ* experimental results, the values of the DMOS are shifted by one hundred for calculation convenience and spatial resolution reflection. This leads to

$$SQ = DMOS + 100 . \quad (1)$$

Then, we obtain and compare the proposed quality metric model with (1) to verify the effectiveness of the model.

A. Quality metric with temporal and SNR scalability

In general, a good quality metric has a high correlation with subjective quality regardless of the variation of the consumption circumstances. In [9], Feghali et al. show a quality metric which adopts frame rate, SNR, and motion characteristics as

$$QM = PSNR + m^{0.38}(30 - FR) , \quad (2)$$

where *m* is motion feature and *FR* is frame rate. From the idea of a linear combination between PSNR and frame rate [7], PSNR is used as the first parameter in (2). The subjective quality depends on variation of the frame rate and motion so that they can be the rest of the parameters. The frame rate reduction (30-*FR*) and the motion feature establish the linear mapping between the PSNR and the perceived quality. To measure the degree of motion complexity, the equation uses the motion vector magnitudes between every consecutive two frames in source sequences [7]. These parameters prove that the correlation coefficients between the quality metric of (2) and subjective quality are high on the subjective assessment. In this work, therefore, a new video quality metric (*VQM*) is derived and modeled from (2).

In the above context, the important factor is not to find the exact value of the motion vector magnitudes, but to obtain the range of the magnitudes [9]. Therefore, we apply MPEG-7 motion activity, which is the closest to the perceived motion speed, as a motion parameter. The motion activity is based on the standard deviation of the motion vector magnitudes which are used as a good measure of the perceived motion intensity. In [10], the intensity level of the motion activity is used for a subjective perception on motion intensity of video.

Reasoning by analogy with (2), it is expected that the video quality metric is easily obtained with fixed spatial resolution, in the form of

$$VQM_{(Temp,PSNR)} = \alpha PSNR + \beta M_A (30 - FR) , \quad (3)$$

where M_A and *FR* denote motion activity value and frame rate, respectively. To find coefficients α and β of the quality metric for temporal and SNR scalability, we firstly analyze the training sequences with 3 frame rates, and a quantization parameter under the condition that a sequence with each frame rate has the fixed spatial resolution. To reflect the spatial resolution afterwards, we perform linear regression analysis using the least squares method to fit the *SQ* from (3). Then, α=2.3 and β=-0.16 are determined to enable the video quality metric to guarantee the highest correlation with *SQ*. Regarding only the PSNR, the average correlation coefficients between $VQM_{(Temp,PSNR)}$ and assessed subjective quality are augmented from 0.68 to 0.92 for "*Football*," from 0.69 to 0.90 for "*Foreman*," and from 0.93 to 0.95 for "*Paris*," for each of 6 different spatial resolutions, respectively. This shows that Equation (3) is suitable to represent the actual perceptual quality on temporal and PSNR scalability.

B. Quality metric with spatial scalability

One problem with (3) is that it does not reflect spatial resolution variation, because every analysis is performed under the fixed spatial resolution. To contain the variation of spatial resolution effects, we conduct an analysis based on variation effects under the fixed frame rate, fixed SNR (QP) condition. The experimental conditions of 32 QP, 30 fps, and 6 different resolutions (from CIF to QCIF) for "*Football*," "*Foreman*," and "*Paris*" sequences are analyzed.

Several previous studies have attested to high spatial resolution preference [8], [9]. Even though it is intuitively clear that assessors will prefer a high to a low spatial resolution size, we experiment with 3 sequences to find correlation between subjective quality and spatial resolution. As a result, the tendency of assessors' preferences shows the form of a sigmoid shape in every sequence. The effect of spatial resolution variation has no relation to different motion characteristics. So, we use the spatial resolution term as an independent parameter towards variable motion characteristics.

The experimental results on spatial resolution and assessed subjective DMOS form the shape of the sigmoid function.

$$VQM_{(Spatial)} = \frac{\delta}{1 + e^{-\phi\left(\frac{x-216}{7.2}\right)}} + \gamma , \quad (4)$$

where *x* denotes the height of the spatial resolution. Then, we perform non-linear regression to get the optimal coefficients for (4): δ=44.75, ϵ=0.25, and γ=-42.

C. Quality metric with full scalability

By linearly combining both (3) and (4), we establish a video quality metric supporting full scalability with the four parameters, PSNR (QP), frame rate, motion activity, and spatial resolution size (height).

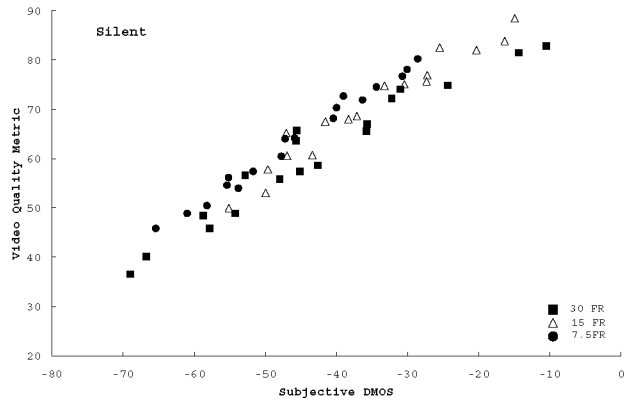
$$VQM_{(Temp,PSNR,Spatial)} = 2.9PSNR + 0.22M_A(30 - FR) + \frac{33.12}{1 + e^{-0.25\left(\frac{x-216}{7.2}\right)}} X_3 + 54.08 . \quad (5)$$

Through the comparison of the average correlation coefficients, the proposed $VQM_{(Temp,PSNR,Spatial)}$ in (5) performs more effective than previous quality metric *QM* in (1) by supporting spatial resolution. For training sequences,

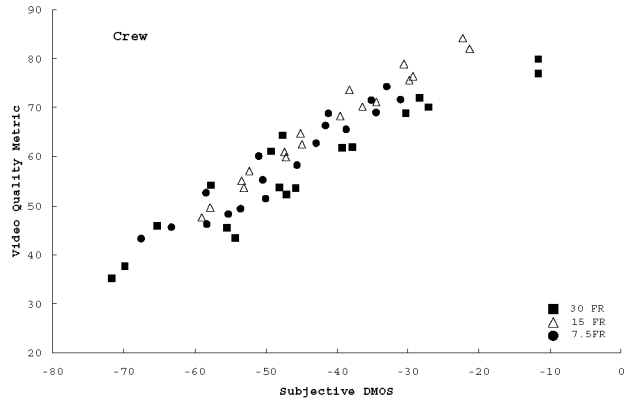
$VQM_{(Temp,PSNR,Spatial)}$ yields the average correlation coefficients of the assessed subjective DMOS higher than 0.9. On the other hand, *QM* is not reflected accurately if the spatial resolution variation is very low.

IV. PERFOMANCE EVALUSTION

So far, the proposed video quality metric $VQM_{(Temp,PSNR,Spatial)}$ with full scalability has been derived and modeled. To evaluate the effectiveness of the new video quality metric, we test it with two sequences: "*Crew*" with a fast motion and "*Silent*" with a slow motion of which motion activities are 3.12 and 1.88, respectively. The two sequences are subdivided into 108 video sequences that follow the conditions of Table I. In other words, every encoding and decoding condition on these sequences is exactly identical to the modeling (training) experiment. The number of assessors is fifteen; and there are a few assessors who took part in both subjective tests on modeling and testing of the quality metric. The test sequences also show a similar pattern compared to the training sequences in terms of the variation of spatial resolution.



(a)



(b)

Figure 2. Video quality metric with subjective DMOS (a) "Crew" and (b) "Silent"

Fig. 2 shows the scatter plots for the test sequences which draw the relationship between the proposed quality metric and subjective DMOS linearly. It is obvious that the metric gives good correlation on human-perception for video quality.

Then, we calculate and compare correlation coefficients between the video quality metric and the subjective DMOS. The average correlation coefficients from the new *VOM* show that "*Crew*" is 0.93 and "*Silent*" is 0.95 for the test sequences. The experimental results are improved in contrast with those of *QM* in (2). This fact proves that the new *VQM* is an appropriate video quality metric involving human perception, as shown in Table I.

TABLE I. CORRELATION COEFFICIENTS BETWEEN SUBJECTIVE DMOS AND THE PROPOSED METRIC ON TEST SEQUENCES FOR EVALUATION

Sequence	*QM*	$VQM_{(Temp,PSNR,Spatial)}$
Crew	0.41	0.93
Silent	0.48	0.95

V. CONCLUSION

In this paper, we addressed novel video quality measurement modeling for scalable video coding with full scalability. The result of our subjective assessment revealed that PSNR and other metrics without consideration of spatial resolutions were not suitable to estimate the quality of practical videos. To model a new video quality metric with full scalability, we analyzed temporal, SNR scalability first, and thereby found the relationship with spatial scalability. Motion feature was also employed for the metric. Experimental results showed that the proposed quality measurement modeling gave high correlation on human perception. Therefore, it could be expected that the proposed video quality metric would be used for video adaptation of SVC in mobile applications. Future points need to be studied for video sequences with either both slow and fast motion or higher spatial resolution.

REFERENCES

[1] Y.J. Jung, Y.S. Kim, D.Y. Kim, J.G. Kim, J.W. Hong, Y.M. Ro, "Analysis of human perception for semantic concept-based video transcoding," *Int. Work. Advanced Image Tech.*, pp. 251–256, Jan, 2005

[2] S. Jumisko-Pyykkö, J. Häkkinen,"Evaluation of subjective video quality of mobile devices," *in Proc. ACM Multimedia*, pp. 535–538, Nov. 2005

[3] E.C. Reed and F. Dufaux, "Constrained bit-rate control for very low bit-rate streaming-video applications," *IEEE Trans. Circuit Syst. Video Tech.*, vol. 11, no. 7, pp. 882–889, July 2001

[4] E. Ong, X. Yang, W. Lin, Z. Lu, S. Yao, "Perceptual quality metric for compressed videos," *IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol.2, pp. 581–584, Mar. 2005

[5] *ITU-R Recommendation BT.500-11*, "Methodology for the subjective assessment of the quality of television pictures," 2002

[6] *ISO/IEC JTC 1/SC29/WG11*, "Joint Draft 5: Scalable Video Coding," JVT-R201, Bangkok, Thailand, Jan. 2006

[7] G. Hauske, R. Hofmeier, T. Stockhammer, "Subjective image quality of low-rate and low-resolution video sequence," *in Proc. Int. Work. Mobile Multimedia Comm.*, Oct. 2003

[8] M. Ries, O. Nemethova, B. Badic, M. Rupp, "Assessment of H.264 coded panorama sequences," *in Proc. Conf. Multimedia Services Access Networks*, pp. 6–9, Jun. 2005

[9] R. Feghali, D. Wang, F. Speranza, A. Vincent, "Quality metric for video sequences with temporal scalability," *in Proc. Int. Conf. Image Processing*, pp. 137–140, Sept. 2005

[10] S. Benini, L.-Q. Xu, R. Leonardi, "Using lateral ranking for motion-based video shot retrieval and dynamic content characterization," *in Proc. Fourth Int. Work. Content-Based Multimedia Indexing*, June, 2005