Lightweight Quality Metric Based on No-Reference Bitstream for H.264/AVC Video

Yo-Han Kim[1] and Jitae Shin[1*] and Hokyom Kim[2]
[1] College of Information and Communication Engineering, Sungkyunkwan University
Suwon 440-746, Republic of Korea
[2] Creative and Challenging Research Division, Electronics and Telecommunications Research Institute
Daejeon 305-700, Republic of Korea
[e-mail : {dos95, jtshin}@skku.edu, hokykim@etri.re.kr]
*Corresponding author: Jitae Shin

Received February 13, 2012; revised April 9, 2012; accepted April 22, 2012; published May 25, 2012

Abstract

This paper proposes a quality metric based on a No-Reference Bitstream (NR-B) having least computational complexity for the assessment of the human-perceptual quality of H.264 encoded video. The proposed NR-B method performs a modeling of encoding distortion with three bit-stream information (i.e. frame-rate, motion-vector, and quantization-parameter) that can be directly extractable from the encoded bitstream and does not require additional complex processing of final pictures. From performance evaluation using 165 compressed video sequences, the experiment results show that the proposed metric has a higher correlation with subjective quality than is achieved with other comparable methods.

Keywords: No-reference, perceptual quality, H.264/AVC.

This work was supported by the R&D program of the Ministry of Knowledge Economy (MKE), Korea entitled “Study on Ubiquitous Infra, Core Technology on Enhanced QoS/ QoE.”

http://dx.doi.org/10.3837/tiis.2012.05.008

1. Introduction

Traditionally, Mean Square Error (MSE) or Peak Signal-to-Noise Ratio (PSNR) are the prevalent means used to measure objective video quality. However, PSNR has a poor relationship with the human visual system (HVS) [1]. Therefore, subjective video quality assessment is used to measure human perceptual video quality rather than traditional objective video quality, e.g. PSNR, but it requires a lot of time and human resources. Moreover, it can't be executed in real time. In ITU-T SG12, the quality of experience (QoE) is defined as the overall acceptability of an application or service perceived subjectively by the end-user. Service providers want to enhance the QoE that the end-user feels. So the objective assessment of perceptual video quality is an important factor in QoE. The Video Quality Expert Group (VQEG) has been launched, with the aim of standardization of a video quality assessment method for human perceptual quality.

Perceptually objective video-quality assessment means to automatically compute quality scores that show highly correlation with those given by human observers. There are three video quality methods used for perceptually objective video-quality assessments. The first of these, the Full-Reference (FR) method, compares a processed video signal to the original signal. The second is the Reduced-Reference (RR) method. In order to reduce the amount of data needed for the calculation at the receiver, the RR method extracts features from the sender side and received signals. Thus, for the measurement of picture quality, only these features have to be considered, instead of the entire original signal as in FR. The third of these, the No-Reference (NR) method, tries to determine the video quality using only extracted information from received pictures or bit-streams. The NR method can be categorized into two types. One is the No-Reference Pixel (NR-P) type, which uses the processed video sequences (PVS) to check for encoding errors, such as blocking and blurring artifacts. The other No-Reference Bit-stream (NR-B) method uses the bit-stream. In this method, NR extracts the motion vector and encoding parameters from the bit-stream and uses the information of packet loss. Generally, the FR method and RR method have high correlation in comparison with the NR method. But the FR and RR methods are not suitable for real time streaming service. The FR and RR methods have an overhead when estimating end-user's quality, because they transmit some information of the original signal. On the other hand, NR does not require additional data. Methods of objective video quality assessment are shown in **Fig. 1**.

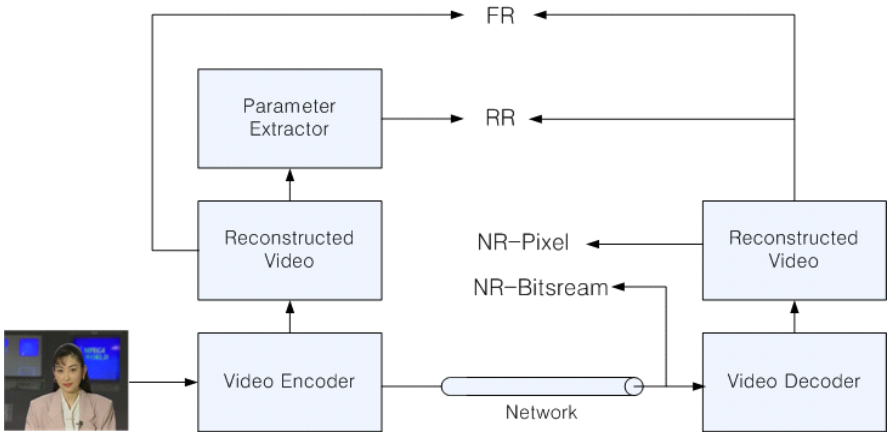


Fig. 1. Objective video quality assessment methods

In this paper, we propose a perceptual video-quality metric based on the NR-B method, which is the simplest one among all the objective video quality assessment methods shown in **Fig. 1**. However, it is known that the accuracy of NR-based methods is poor, in comparison with FR and RR, because of least available information. Therefore, it is worthwhile if an NR-based metric has similar performance to those based on FR and RR, because the NR-based metric requires least information to estimate the perceptual video quality. In particular, the NR-B method has the lowest complexity, because it uses the encoding parameters received directly from the bitstream.

The remainder of this paper is organized as follows. In Section 2, we review the previous work in perceptual video quality assessment. In Section 3 we introduce the subjective video quality assessment, which is compared with the proposed objective video quality metric for accuracy. Section 4 proposes a light-weight NR-B metric, and Section 5 presents experiment results.

2. Related Work

Regardless of FR, RR, or NR, typical pixel-based methods estimate the video quality from measuring features, such as blur [2][3], blockiness [4] or motion jerkiness, but the calculation of feature quantities causes high computing complexity. Lin and Kuo [5] provide a survey of perceptual visual quality metrics using image features.

Quality assessment based on the FR method has been well investigated. Seshadrinathan and Bovik [6] evaluated 'Speed SSIM (Structural SIMilarity)', 'V-VIF (Video-Visual Information Fidelity)', and 'Temporal MOVIE (MOtion-based Video Integrity Evaluation)' models. These models show a higher correlation than other methods. Speed SSIM uses the SSIM index in conjunction with statistical models of visual speed perception. V-VIF is an extension model of VIF for video. Also, Temporal MOVIE is one of the MOVIE indexes, which consider motion-based temporal distortion. These three methods have high computation complexity, because they are based on FR and pixel processes. Furthermore they are designed without consideration of video compression characteristics. An FR algorithm was proposed by Feghali et al. [7], using the frame rate (Fr), PSNR, and motion vector (MV). They treated PSNR as the most important element, and used Fr and MV in order to compensate for difference between the PSNR and measured perceptual quality.

For RR and NR methods, Wang and Bovik [8] provide a generic framework using a visual quality assessment model that is mainly focused on image, not video. A quality metric in [9] is expanded to various spatial resolution, using PSNR, Fr, MV and width of picture. But, it also requires additional transmission data due to characteristics of the RR method. Le Callet et al. [10] suggest metrics based on RR and NR using neural network architectures with frequency, temporal, and blocking features. However, their scheme has high complexity, due to the neural network algorithm, and the achieved correlation coefficient is not high.

Among NR approach, NR-P methods require additional pixel processing to extract the features of blur, blockiness, etc. after decoding and then they are much more complex than NR-B methods. A quality estimation model in NR-P was proposed by Kawano et al. [11] using blockiness and blur derived from decoded videos, but the correlation with subjective test was poor. A NR video quality monitoring (NORM) algorithm of Naccari et al. [12] estimates the PSNR due to channel errors by using decoded frames, received MVs, coding modes, and prediction residuals in video sequences coded with H.264/AVC. A rule-based quality estimation was proposed by Oelbaum et al. [13] for H.264/AVC video using spatial

activity, continuity features in addition to blur and blocking effect in pixel domain, but it also required complex models for feature extraction. The other NR-P approaches applied to scalable video were found in Zhai et al. [14] for the video quality metric of scalable video with elements of blockiness, blur, and motion jerkiness, and Eichhorn and Ni [15] performed subjective tests in order to evaluate priority of the video quality layers in scalable video.

There were also several NR-B methods that use only bitstream information. An encoding error estimation model of Brandão and Queluz [16] utilized discrete cosine transform (DCT) coefficients and quantization step for estimating PSNR. In order to packet loss effect, a NR quality assessment scheme for networked video of Yang et al. [17] used information about lost packets, frame type, and bit-rate in order to estimate distortion caused by packet loss, but it was limited in that it applied only to MPEG-4 encoded QCIF (quarter common intermediate format, low-resolution) videos.

Most of the previous research for perceptual and objective video quality formulated complex estimation models which apply complicated operations on video quality elements such as blur, blockiness, resolution, Fr, PSNR, and MV. In the paper, we propose a lightweight NR-B quality metric as a simplest method with no additional complexity for feature extraction like NR-P. It offers simple and direct operation, using bitstream parameters only that are readily extractable; therefore the proposed metric has lower complexity than existing schemes. Also we formulate a relationship among parameters, in order to increase the accuracy of the proposed objective video quality metric.

3. Referenced Subjective Quality Test

An experiment is performed to obtain a mean opinion score (MOS) for subjective quality in order to reflect user QoE. Our subjective test method follows the Absolute Category Rating (ACR) of ITU-T P.910. This method specifies that opinion providers are asked to evaluate the quality of the sequence shown after each presentation. The time pattern and scoring for the stimulus presentation is illustrated in **Fig. 2**. We use 9 test sequences, of which four are used for training as shown in **Fig. 3**, and five for evaluation as in **Fig. 7**. We made 225 encoded sequences from 9 raw videos that have a length of 8 seconds, with a resolution of QCIF. Encoding of the H.264/AVC is performed with the conditions of the quantization parameter (QP) being 28 / 32 / 36 / 40 / 44, using 1.875 / 3.75 / 7.5 / 15 / 30 frames per second (fps). We use a baseline profile, group of picture (GOP) length of 16, and an error concealment of the frame copy method. We used the detailed test process from Cano et al. [18].

20 students participated in our subjective test, and they were not related in any other way with this research.

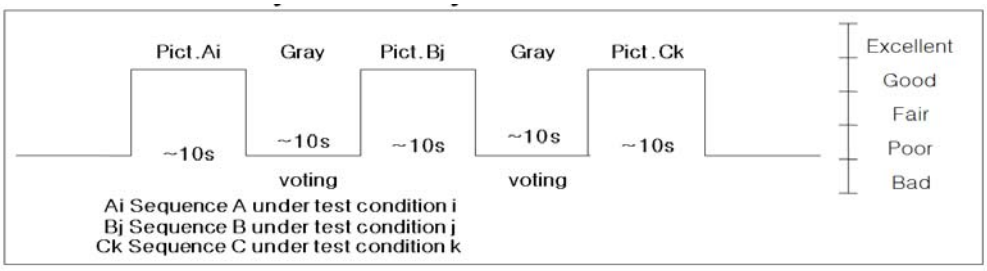


Fig. 2. The Scheme of the ACR method

Fig. 3. Training sequences

To measure the quality of the compressed video, we can use Differential MOS (DMOS) obtained from our performed subjective test or LIVE video quality database [19]. From the VQEG Hybrid / Bit-stream Group test plan [20], the DMOS value is calculated using the following formula.

$$DMOS = 5 + MOS(PVS) - MOS(SRC) \tag{1}$$

where PVS is the decoded video sequence and SRC is the original video sequence. The DMOS is a subjective quality metric with a procedure known as hidden reference removal, because it is not affected by the quality of the original source video. The value of DMOS is then normalized to a scale from 0 to 1, and is converted to the distortion of the compressed video, D_m, which can be defined as

$$D_m = 1 - \{DMOS - 1\}/4 \ . \tag{2}$$

D_m represents the metric of subjective video quality, which will be a reference for comparison with the proposed video quality metric.

4. Proposed Quality Metric based on NR-B

Compressed video shows different qualities according to the video features and encoding parameters. Generally, fast motion video has a lower quality than slow motion video. Decreasing the frame rate accelerates the quality degradation.

Moreover, *QP* has a close correlation with distortion. For an NR-B video quality assessment, we have to estimate the video quality using only compressed video bitstream and transmitted parameters, because there is no available information about the uncompressed video at the receiver.

We extract MV information (*MV*), frame rate (*Fr*), and quantization parameter (*QP*) from the encoded bitstream. Before formulating the distortion, we normalize *MV* and *QP*. The values of *MV* are obtained with every inter-coded Macro Block (MB), and the magnitude of the normalized *MV* is defined as follows

$$MV_{NORM} = \frac{\sum_{i=0}^{M}\sum_{j=0}^{N}\sqrt{x_{i,j}^2 + y_{i,j}^2}}{F_{tot} \times B_{tot} \times \sqrt{(h\times 4)^2 + (w\times 4)^2}} \times Fr \tag{3}$$

where F_{tot} is the number of frames of the sequence, and B_{tot} is the total number of MBs of

a frame picture. The h and w are the height and width of the sequence, and $x_{i,j}$ and $y_{i,j}$ are the motion coordinates of the MB, respectively,.

In H.264 compression, the QP value ranged from 0 to 51, and increasing QP by 6 doubles the size of the quantization step that is proportional to video quality.

So we normalize QP as

$$QP_{NORM} = \log_{\sqrt{6}} QP \tag{4}$$

Fig. 4 illustrates the subjective distortion D, according to QP_{NORM}.

To estimate the quality of the compressed video, we use the Differential MOS (DMOS) metric, so the estimated DMOS D_e is

$$D_e = aQP_{NORM} + b \tag{5}$$

where a is the gradient of QP_{NORM} and b is the offset of the linear equation. **Fig. 5** shows the variance of a according to Fr and MV_{NORM} for four typical training video sequences chosen to show distinct picture activities, i.e. different MV_{NORM} patterns.

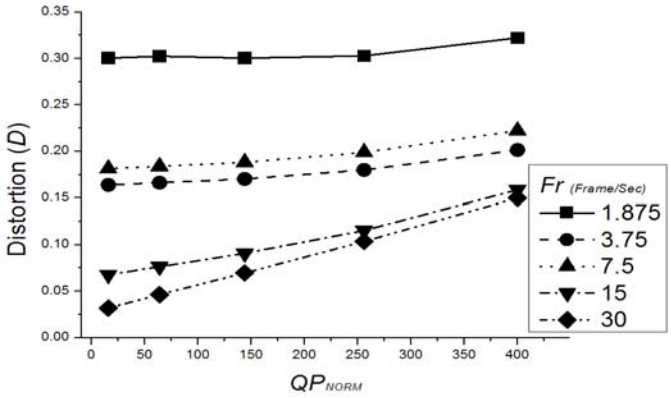


Fig. 4. Distortion versus QP_{NORM} of the AKIYO sequence

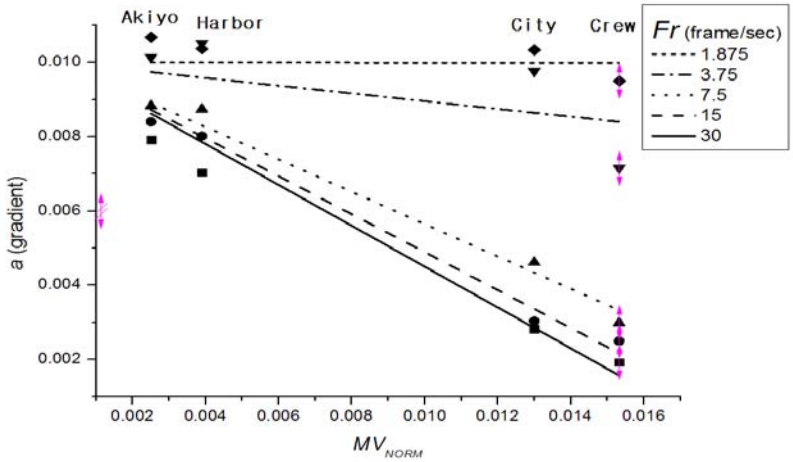
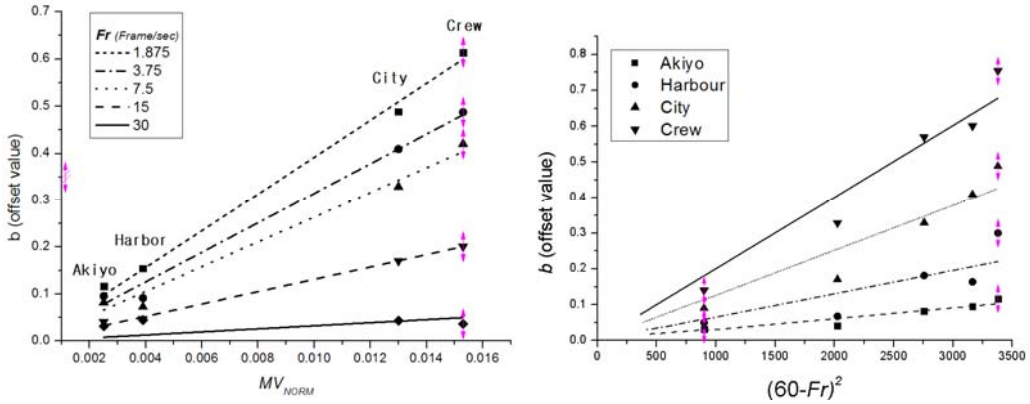


Fig. 5. a versus QP_{NORM}.



(a) (b)

Fig. 6. Offset versus (a) MV_{NORM} and (b) $(60\text{-}Fr)^2$

As can be seen in **Fig. 5**, a low motion sequence has a lower variation range of a than a high motion one. It means that Fr is directly proportional to a. It indicates that higher MV_{NORM} causes a lower a value. As results, we can describe the gradient, a, as

$$a = \alpha_1(0.01 - \alpha_2 MV_{NORM}) \times Fr \tag{6}$$

where α_1 and α_2 are some constant values.

Next, we can depict the value of the offset b. **Fig. 6** describes this offset according to the MV. We find that MV_{NORM} and $(60\text{-}Fr)^2$ is in proportion to the offset, b. So we can express b in the form

$$b = \beta_1 MV_{NORM} \times (60 - Fr)^2 + \beta_2 \tag{7}$$

where β_1 and β_2 are some constant values.

Finally, the estimated distortion D_e, from Eqns. (5), (6) and (7) can be formulated as

$$\begin{aligned} D_e = {} & w_1 QP_{NORM} \times Fr - w_2 MV_{NORM} \times QP_{NORM} \times Fr \\ & + w_3 MV_{NORM} \times (60 - Fr)^2 + w_4 \end{aligned} \tag{8}$$

An experiment of subjective quality testing using the four training sequences shown in Fig. 3, which were selected because of having different features of picture activities, was performed with different parameters of QP, Fr, and MV. Using these experimental data, and with regression analysis based on the least squares method, we obtain a set of coefficient values (w_1, w_2, w_3, w_4) as (1.04, -66.5, -0.0140, 0.363) with the four training sequences of **Fig. 3**.

5. Experimental Results

In order to evaluate the performance of the proposed NR-B metric, we firstly validate the proposed method with five QCIF test sequences as low quality (LQ) video, as shown in **Fig.**

7. For a rule of generality, these test sequences were not used in the estimation of the weighting in Eqn. (8). Secondly we compare the proposed one with the performance data from [19] using other referenced methods. In the LIVE Video Quality Database [19], performance results of 10 video sequences with high quality (HQ) are provided, as shown in **Fig. 8**. Those sequences are high quality videos with a resolution of 768x432, and 25 or 50 fps sequences. Each sequence is compressed using H.264, and the compression rates vary from 200 Kbps to 5 Mbps, according to Seshadrinathan et al. [21]. We calculated the Pearson correlation (PC) coefficient mentioned in the VQEG Group Test Plan [20]. In **Table 1**, the proposed metric shows similar PC coefficients for each sequence with the 'Temporal MOVIE' using a pixel-based FR method. However, the proposed metric shows better correlation in the diversity of total video sequences.

Fig. 7. Low Quality Test Sequences for Evaluation.

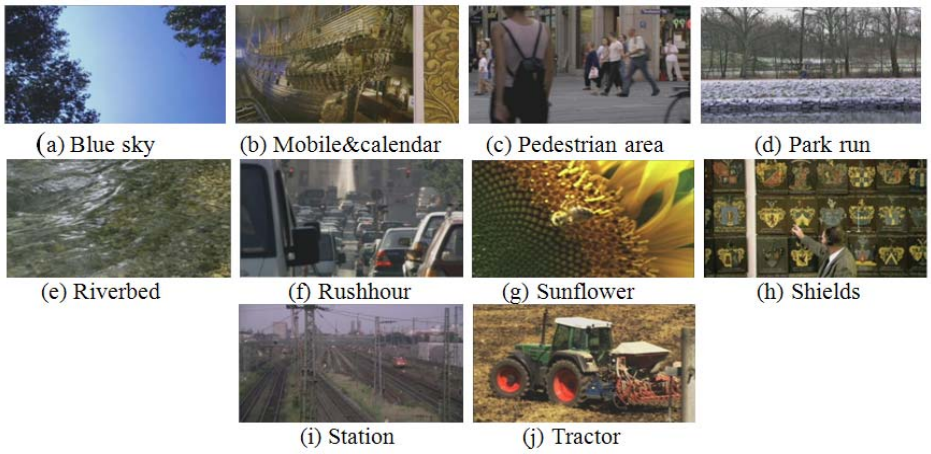


Fig. 8. High Quality Test Sequences

Next, we run the experiment, and compare the proposed metric with other comparable methods shown in [21] for high quality videos as well. **Table 2** shows that the Proposed NR-B method has a higher PC in diverse video sequences of **Fig. 7** than other comparable methods, in terms of subjective DMOS.

Table 1. PC coefficient evaluated from LQ video sequences

Video	Temporal MOVIE [6]	Proposed NR-B
News	0.945	0.805
Foreman	0.931	0.950
Football	0.905	0.969
Mobile	0.874	0.949
Soccer	0.713	0.716
Total LQ Video	0.781	0.894

Table 2. PC coefficient evaluated from HQ video sequences

	Temporal MOVIE [6]	Proposed NR-B
Total HQ	0.756	0.758

The reason for the lower total value in the Temporal MOVIE is that PC trend lines are scattered according to different sequences, but the proposed one is not. The graphical explanation is shown in **Fig. 9**.

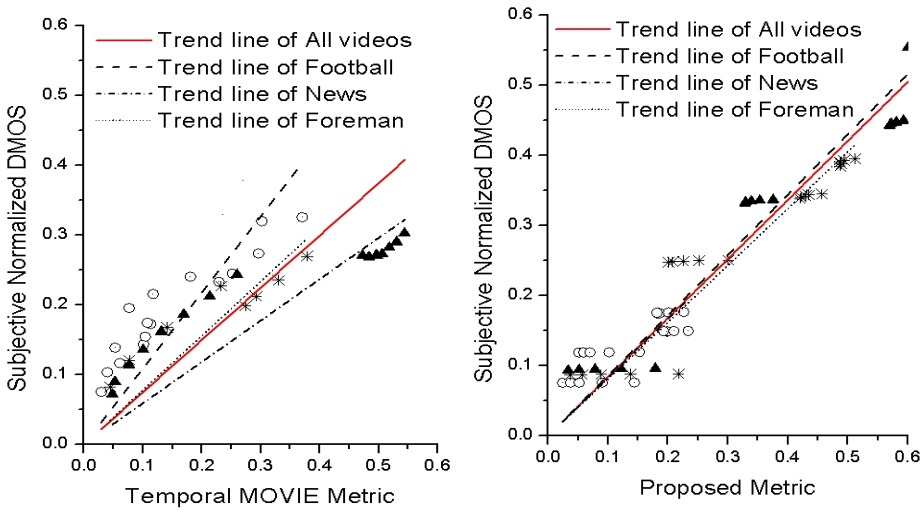


Fig. 9. Linear PC trend line of Low quality videos of (a) Temporal MOVIE (b) Proposed NB-R

In summary, the total distribution between the subjectively normalized DMOS score, D_m, and the proposed NR-B quality metric, De, is illustrated in **Fig. 10**. 125 evaluation and 125 training compressed sequences of low quality are differently encoded from the 10 sequences shown in **Fig. 3** and **Fig. 7**. 40 compressed high quality sequences are encoded from the 10 sequences shown in **Fig. 8**.

Fig. 10 shows that the NR-B quality metric provides a trend line having high correlation with the subjective normalized DMOS score. Its PC coefficient averages 0.827 over a wide range of video resolution.

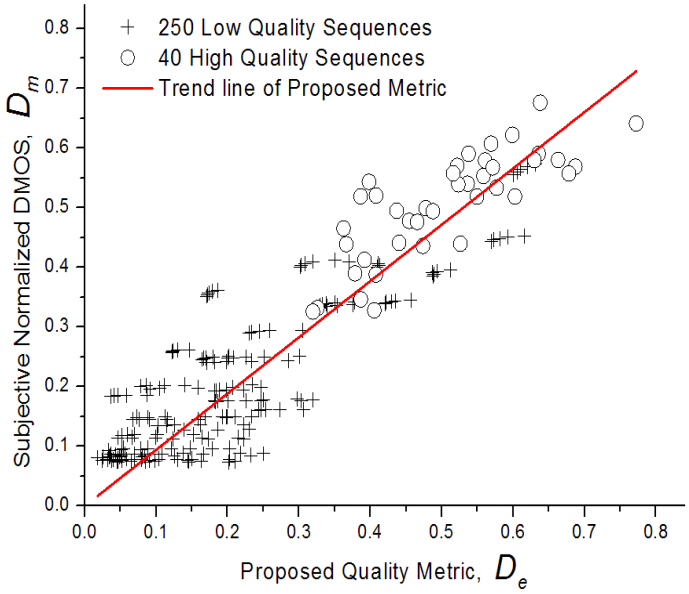


Fig. 10. Distribution map between D_e and D_m

6. Conclusion

A lightweight metric based on the NR-B method is proposed to represent perceptual video-quality. The proposed metric does not require complex models to manipulate the directly extracted parameters such as QP, Fr, and MV from received H.264/AVC encoded bitstreams, in comparison with FR, RR, and NR-P approaches, which do. The experimental results evaluated by both low- and high-quality video sequences show that the proposed quality metric has a high average correlation value of 0.827 in PC coefficient, and is slightly better than one of the well-established FR-based schemes, i.e. MOVIE. The proposed NR-B metric can be operated in real-time and with least burden to a legacy system from the quality monitoring at the receiver side, because of the simple extraction of the parameters used, and lightweight calculation of the quality metric.

Further work is required to evaluate whether the proposed scheme will be also effective under several different error loss conditions. The NR-based scheme is susceptible to different packet loss from error-prone wireless networks, and needs to be validated under well-known channel error patterns. Also, there are perceptual quality-assessment issues in recently prevailed three-dimensional (3-D) or stereoscopic images and videos.

References

[1] B. Girod, “What’s wrong with mean squared error?” in AB. Watson (Ed.), *Digital Images and Human Vision*, MIT Press, Cambridge, MA, 1993, pp.207–220. Article (CrossRef Link)

[2] P. Marziliano, F. Dufaux, S. Winkler and T. Ebrahimi, “A no-reference perceptual blur metric,” in *Proc. of IEEE ICIP*, pp. III-57-III-60, 2002. Article (CrossRef Link)

[3] R. Ferzli, and L.J. Karam, “A No-reference objective image sharpness metric based on the notion of just noticeable blur (JNB),” *IEEE Transactions on Image Processing*, vol.18, no.4, pp.717-728, Apr.2009. Article (CrossRef Link)

[4] S. Lee and S.J. Park, “A new image quality assessment method to detect and measure strength of blocking artifacts,” *Signal Processing: Image Communication*, vol.27, no.1, pp.31-38, Jan. 2012. Article (CrossRef Link)
[5] W. Lin, and C.J. Kuo, “Perceptual visual quality metrics: A survey,” in *Journal of Visual Communication and Image Representation*, vol.22, no.4, pp.297-312, May.2011. Article (CrossRef Link)
[6] K. Seshadrinathan and A. C. Bovik, “Motion tuned spatio-temporal quality assessment of natural videos,” *IEEE Trans. Image Process*., vol.19, no.2, pp.335-350, Feb.2010. Article (CrossRef Link)
[7] R. Feghali, D. Wang, F. Speranza and A. Vincent, “Quality metric for video sequences with temporal scalability”, In *Proc. of IEEE ICIP*, pp.11-14, Sept.2005. Article (CrossRef Link)
[8] Z. Wang and A.C. Bovik, “Reduced- and no-reference image quality assessment,” *IEEE Signal Processing Magazine*, vol.28, no.6, pp.29-40, Nov.2011. Article (CrossRef Link)
[9] C. S. Kim, S. H. Kim, D. J. Seo and Y. M. Ro, “Measuring video quality on full scalability of H.264/AVC scalable video coding”, *IEICE Trans. Commun*., vol.E91-B, 2008. Article (CrossRef Link)
[10] P. Le Callet, C. V. Gaudin, S. Pechard and É. CaillardAILLAULT, “No reference and reduced reference video quality metrics for end to end QoS monitoring”, *IEICE Trans. Commun.*, vol.E85-B, 2006. Article (CrossRef Link)
[11] Kawano, T.; Yamagishi, K.; Watanabe, K. and Okamoto, J., “No reference video-quality-assessment model for video streaming services,” in *Proc. of 18th International Packet Video Workshop*, pp.158–164, 2010. Article (CrossRef Link)
[12] M. Naccari, M. Tagliasacchi and S. Tubaro, “No-reference video quality monitoring for H.264/AVC coded video,” *IEEE Trans. On multimedia*, vol.11, no.5, pp.932-946, Aug.2009. Article (CrossRef Link)
[13]T. Oelbaum, C.Keimel and K.Diepold, “Rule-based no-reference video quality evaluation using additionally coded videos,” *IEEE Journal of Selected Topics in Signal Processing*, vol.3, no.2, pp.294–303, 2009. Article (CrossRef Link)
[14] G. Zhai, J. Cai, W. Lin, X. Yang and W. Zhang, “Three dimensional scalable video adaptation via user-end perceptual quality assessment,” *IEEE Trans. Image Process*., vol.19, no.2, pp.335-350, 2010. Article (CrossRef Link)
[15] A. Eichhorn and P. Ni, “Pick your layers wisely - a quality assessment of H.264 scalable video coding for mobile devices,” In *Proc. of IEEE International Conference on Communications*, 2009. Article (CrossRef Link)
[16] T. Brandão and M. P. Queluz, “No-reference quality assessment of H.264/AVC encoded video,” *IEEE Transactions on Circuits Systtem Video Technol*., vol.20, no.11, pp.1437-1447, Nov.2010. Article (CrossRef Link)
[17] Fuzheng Yang; Shuai Wan; Qingpeng Xie and Hong Ren Wu;, “ No-reference quality assessment for networked video via primary analysis of bit stream,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol.20, no.11, pp.1544– 554, Nov.2010. Article (CrossRef Link)
[18] M-D. Cano, F. Cerdan and S. Almagro, “Statistical analysis of a subjective QoE Assessment for VVoIP Applications”, *ETRI Journal*, vol.32, no.6, pp.843-853, Dec.2010. Article (CrossRef Link)
[19] LIVE Video Quality Database, http://live.ece.utexas.edu/research/quality/live_video.html.
[20] “Hybrid Perceptual/Bitstream Group Test Plan,” *Video quality experts group (VQEG)*, Draft Version 1.3, 2009.
[21] K. Seshadrinathan, R. Soundararajan, A. C. Bovik and L. K. Cormack, "Study of Subjective and Objective Quality Assessment of Video", *IEEE Trans. Image Process*., vol.19, no.6, pp. 1427–1441, 2010. Article (CrossRef Link)

Yo-Han Kim received B.S. and M.S degrees in electronic engineering from Ajou University, Korea in 2001 and 2003, respectively. He received his Ph.D. degree at Sungkyunkwan University, Korea, in 2012. Recently, he has joined Display Processing Lab, Samsung Electronics, Suwon, Korea from March 2012. His research interests include video communication and next generation networks.

Jitae Shin received his B.S. from Seoul National University in 1986, M.S. in Korea Advanced Institute and Technology (KAIST) in 1988, and his second M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, USA, in 1998 and 2001, respectively. He is an Associate Professor in the School of Electronic and Electrical Engineering, Sungkyunkwan University (SKKU), Korea. Before joining SKKU, he had worked for instrumentation & control systems of nuclear power plants at Korea Electric Power Corporation (KEPCO) and Korea Atomic Energy Research Institute (KAERI) from 1988 through 1996. His research interest includes video signal processing and multimedia transmission over future Internet or wireless/mobile networks focusing on QoS and multimedia network control/protocol issues. He is a member of IEEE and IEICE.

Hokyom Kim received his B.S. and M.S. degrees in electronics engineering from Yonsei University, Seoul, Korea, in 1983 and 1988, respectively. From 1983 to 1987, he was with the Hyosung Heavy Industries. From 1987 to 1988, he was with Samsung Advanced Institute of Technology (SAIT). He has joined in 1989 to Electronics and Telecommunications Research Institute (ETRI) as a principal member of research staff. His current research interests include convergence technologies of broadcast and telecommunications and broadband Public Protection and Disaster Relief (PPDR) communication system design.