

# No Reference Video Quality Metric for 4K-UHD Videos Based on H.264/AVC Bitstream Features

J.B. Xu, X.H. Jiang

Information Engineering School  
Communication University of China  
Beijing, China

**Abstract**—With the wide-spread use of digital videos, quality considerations have become essential, and industry demand for video quality measurement standards is rising. In this paper, we proposed a no reference video quality metric for 4K-UHD videos in the compressed domain. As video coding parameters which can reflect the degree of distortion caused by compression can be extracted in compressed domain, we firstly extracted some feature parameters from 4K-UHD compressed videos. Then the parameters were applied to train the proposed model with the corresponding subjective score using multi-linear algorithm. The experimental results show that our model can achieve a good performance for the 4K-UHD video streams with the subjective score.

**Keywords**—4K Ultra High Definition; subjective video quality assessment; no reference; objective video quality assessment; compressed domain

## I. INTRODUCTION

With the advent of video coding, display and network technologies, there have been many demands for high resolution video services in the areas of TV broadcasting, Internet, personal media and etc. Recently, rich video contents of Ultra High Definition (UHD) in picture resolution have drawn attention as an essential element for beyond-HDTV video services. Some commercial products such as beam projectors, monitors and TV sets that support 4K-UHD resolution of  $3,840 \times 2,160$  have already been introduced in the markets [1]. ITU-R has recently announced a recommendation on UHD TV in collaboration with experts from television industry, broadcasting organizations and regulatory institutions in its Study Group 6 [2]. Similar to its precedent recommendations [3]–[7] for SDTV and HDTV, the recommendation on UHD TV is laid out in two levels: the first level of UHD TV with picture resolution of about 8 megapixels ( $3,840 \times 2,160$  image system) which is called 4K-UHD TV in this paper; and the second level with about 32 megapixels ( $7,680 \times 4,320$  image system), called 8K-UHD TV. With the huge amounts of UHD TV data, high data compression which will induce distortion is inevitable. So how to evaluate the quality of UHD TV is becoming an important research topic.

Assessment methods used to evaluate video quality can be divided into two categories: subjective and objective. Subjective methods are wildly used, but they have obvious disadvantages. It is too inconvenient and too expensive to be conducted in real-time applications. The goal of objective

measurement research is to design algorithms that can automatically assess the video quality in a perceptually consistent manner. The objective methods can be classified further into three classes regarding their dependency on available reference information: full-reference metric (FR), reduced-reference (RR) and no-reference (NR). In general, it is believed that FR and RR metrics can achieve better quality predictions due to the reference information. In some applications, however, observers can only obtain the distorted medium, so NR metric are urgently needed. As video coding parameters which reflect the degree of distortion caused by compression can be extracted in compressed domain, we can obtain the objective quality by analyzing these parameters. This has the advantage of much lower bandwidth of data to be analyzed compared to metrics looking at the fully decoded video, resulting in much lower processing requirements.

In recent years there are several NR metrics proposed based on the compressed domain. Various features can be abstracted from the compressed domain directly, such as quantization parameters, bitrate, etc. In [8] Eden estimates the PSNR of interlaced HDTV video sequences with H.264/AVC bitstream features whereas Slanina et al. in [9] estimate the PSNR for videos in CIF resolution. Rossholm and Lovstrom not only estimate PSNR, but also other video quality metrics for videos in CIF resolution from the bitstream in [10]. In [11], Lee et al. use bitrate, QP and deblocking filter parameters for quality prediction of QCIF resolution videos, but no different coding structures were considered. Another approach is the combination of bitstream features and features extracted from the decoded video sequences in a hybrid metric as proposed for interlaced HDTV by Sugimoto et al. in [12].

In this paper, we proposed a no reference video quality metric for 4K-UHD bitstreams which are coded by H.264/AVC [14] video coding standard. The rest of the paper is organized as follows: Section 2 describes the subjective video quality assessment method and the results. Section 3 provides the proposed bitstream-based quality assessment algorithm. After presenting and discussing the results in section 4, we outlined the conclusion of the study and direction for future work.

## II. SUBJECTIVE VIDEO QUALITY ASSESSMENT

In this paper, we used three 4K-UHD sequences: ParkJoy, DucksTakeOff and Mobile. The first two sequences are from SVT high definition multi format test set in the 2160p 4K-

UHDTV format [15]. And the Mobile sequence is from [16] which is recorded by the RED Epic 4K camera. The first frame of each content is shown in fig.1. The sequences provide different types of content. The ParkJoy sequence shows a camera pan with a scene in the park, which is fairly critical. The DucksTakeOff sequence provides very fast motion and waves in water, which is fairly unpredictable for a motion compensation [9]. All source sequences were stored as raw video files, progressively scanned, with YUV 4:2:0 colour sampling, and 10 bits per sample. For the subjective video quality assessment on H.264 encoded 4K-UHD videos in this paper, x264 was used to encode the three 4K-UHD source sequences. More details on the configuration can be found in table 1. Each content has been coded at eight different coding rates: 12 Mbps, 15 Mbps, 18 Mbps, 23 Mbps, 28 Mbps, 32 Mbps, 36 Mbps and 40 Mbps.

During subjective tests, a group of people are arranged to watch a set of video clips and rate their quality. A professional, 56-inch LCD reference monitor, TVLogic-LUM 560W was used to display UHD video content.

And we adopted the Double-Stimulus Continuous Quality Scale (DSCQS) method for subjective quality assessment recommended by ITU-R BT.500-11 [5], which is thought to be especially useful for evaluation of a new system. Table 2 shows the Mean Opinion Score (MOS) obtained from all test sequences.

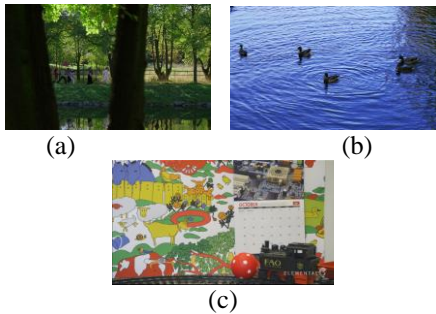


FIGURE I. THE FIRST FRAME OF EACH SEQUENCE: (A) PARKJOY, (B) DUCKSTAKEOFF, (C) MOBILE.

TABLE I. X264 ENCODING SETTINGS

Parameter	Value
Profile	High
Level	5.1
Reference Frames	5
Search Range	64
B-Frames	2
Intra period	1s

TABLE II. MOS Obtained From All Test Sequences

Sequence bitrate(Mbps)	DucksTake Off	ParkJoy	Mobile
12	74.67	55.72	79.22
15	85.33	68.22	86.89
18	85.56	69.33	87
23	88.17	81.56	92.17
28	92.61	82.5	92.62
32	93.44	85.44	93.44
36	94.44	88.17	97.56
40	95.89	88.33	97.59

### III. NO REFERENCE VIDEO QUALITY ASSESSMENT METRIC

#### A. Feature Extraction

In order to build our metric, we first need to extract features from the H.264/AVC [14] bitstreams that describe the properties of the encoded video sequence. Therefore, four video feature parameters which can reflect the compression degree and complexity degree are selected: Quantization parameters, bitrate, Skip-macroblock and motion vector entropy.

##### 1) Quantization Parameters (QP)

Forward quantization of H.264/AVC is denoted by equation (1).

$$Z_{ij} = \text{round}(Y_{ij} / Qstep) \quad (1)$$

where  $Y_{ij}$  is a transform coefficient and  $Qstep$  is the quantization step size.  $Z_{ij}$  represents a quantized coefficient.  $Qstep$  can have one of 52 levels in the H.264/AVC standard.  $Qstep$  is determined by QP. As QP increases by 6,  $Qstep$  doubles. This quantization parameter is closely related to the bitrate and video quality. We calculate the average value of quantization parameters over a frame.

##### 2) Bitrate

Compared with the method performed in spatial images, the method in compressed domain can acquire bitrate from the streams. It has a remarkable advantage. The bit rate can determine the image quality in a large extent.

##### 3) Skip-Macroblock (Num\_Skip)

When using inter prediction coding, H.264/AVC allows using the “skip” macroblock in the flat areas. With the skip-macroblock increasing, the number of macroblocks which do not need to be coded is rising. That means the encoder will produce less bits. So the video with fierce movement will get the lower subjective score under the same bitrate.

##### 4) Motion Vector Entropy (MVE)

The direction of motion vector can be expressed as below:

$$\theta_k(x, y) = \arctan \frac{mv_{ky}}{mv_{kx}} \quad (2)$$

where  $\theta_k(x, y)$  is the direction of kth macroblock,  $mv_{kx}$  is the corresponding horizontal motion vector, and  $mv_{ky}$  is the corresponding vertical motion vector.

In order to reduce the calculation complexity, the direction of movement divided into 36 different regions, i.e., the [0,360] range is equally divided into 36 different regions. Then, we calculated the probability  $P(i)$  of the ith region:

$$p(i) = \frac{k(i)}{\sum_{j=1}^{36} k(j)} \quad (3)$$

where  $k(i)$  is the macroblock number that the direction of the movement belonging to the  $i$ th angle area.

The motion direction entropy (MVE) can be representing as follows.

$$MVE = -\sum_{i=1}^{36} p(i) \log p(i) \quad (4)$$

### B. The Mathematical Model

After extracting the features from the bitstreams, we employ data analysis methods to generate a model to estimate the MOS. In this approach, the problem can be presented as an observation matrix,  $X = [x_1 x_2 \dots x_N]$ , where  $x_1, x_2, \dots, x_N$  are a number of feature vectors that has been generated with different video bitstreams. Each feature vector  $x_n$  consists of extracted parameters denoted  $x_{1,x2}, \dots, x_K$ . The corresponding quality measures for the different video content, MOS, then correspond to the desired  $Y = [y_1 y_2 \dots y_N]$ .  $X$  and  $Y$  can be viewed as training data for a classification, mapping or regression problem. It is desired to find a function  $Z = f(x)$  that maps the given values in  $x$  to a specific value  $Z$  [10]. There are several different models solving the problem, which are more or less computational complex. Since a low complex solution is required in order to have the possibility for an implementation in a real-time assessment, multi-linear regression is selected, as shown below:

$$ObjectiveScore = \alpha \times QP + \beta \times Bitrate + \gamma \times num\_skip + \eta \times MVE + \delta \quad (5)$$

## IV. EXPERIMENTAL RESULTS

The total number of videos is 24 which are ParkJoy, DucksTakeOff, Mobile and there corresponding distorted videos. In order to get a prediction model, we perform the multi-linear regression function. The resulting coefficients in this model are shown in equation (6).

$$ObjectiveScore = -0.88 \times QP + 19.78 \times Bitrate - 9.92 \times num\_skip + 11.446 \times MVE + 107.01 \quad (6)$$

The prediction results of our metric are presented in fig.2 and table 3. Besides the Pearson and Spearman rank order correlation coefficients, we also provide the root mean squared error (RMSE) between predicted and the corresponding subjective score. For comparison, we included the results of the well-known full-reference video quality metrics: SSIM [13] and PSNR. The results show that the proposed metric outperformed the SSIM and PSNR metric. But we only covered three different content and note that the performance for unknown content might be worse.

## V. CONCLUSION

In this paper, we proposed a video quality assessment model for 4K-UHD video streams in the compressed domain. We extracted parameters from the H.264/AVC bitstreams for 4K-UHD videos and used multi-linear algorithm to build a prediction model in combination with MOS. Our results show that the accuracy of the estimation is good. Still, the metric can be further improved in future work by either including a larger data set or by considering a larger set of different encoding parameters.

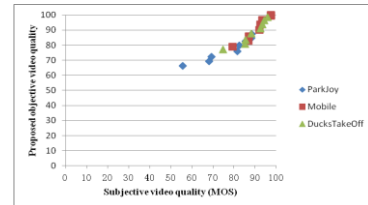


FIGURE II. PREDICTION RESULTS OF THE PROPOSED NO-REFERENCE METRIC

TABLE III. Performance of the Quality Prediction

Metric	Spearman	Pearson	RMSE
Proposed metric	0.990435	0.942768	9.770236
PSNR	0.493913	0.445677	28.15779
SSIM	0.468696	0.391835	43.4313

## ACKNOWLEDGEMENT

This work was supported by the National High Technology Research and Development Program 863(No.2012AA011702).

## REFERENCES

- [1] Sung-Ho Bae, Jaeil Kim, Munchurl Kim, Sukhee Cho, and Jin Soo Choi, "Assessments of Subjective Video Quality on HEVC-Encoded 4K-UHD Video for Beyond-HDTV Broadcasting Services", IEEE Transactions on Broadcasting, 2013 ,pp.209-222.
- [2] Draft New Recommendation--Parameter Values for UHDTV Systems for Production and International Programme Exchange, ITU-R BT.2020, International Telecommunications Union, May 2012.
- [3] Subjective Assessment Methods for Image Quality in High-Definition Television, ITU-R BT.710, International Telecommunications Union, Nov. 1998.
- [4] Subjective Assessment of Standard Definition Digital Television (SDTV) Systems, ITU-R BT.1129, International Telecommunications Union, Feb. 1998.
- [5] Methodology for the Subjective Assessment of the Quality of TV Pictures, ITU-R BT.500-11, International Telecommunications Union, Jun. 2002.
- [6] Parameter Values for the HDTV Standards, ITU-R BT.709-5, International Telecommunications Union, Apr. 2002.
- [7] Test Materials to Be Used in Subjective Assessment, ITU-R BT.1210, International Telecommunications Union, Feb. 2004.
- [8] A. Eden, "No-reference estimation of the coding PSNR for H.264-coded sequences", IEEE Trans. Consum. Electron., vol. 53, no. 2, pp.667 -674, May 2007.
- [9] M. Slanina, V. Ricny, and R. Forchheimer, "A novel metric for H.264/AVC no-reference quality assessment", EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services, 2007, pp.114-117.

- [10] A. Rossholm and B. Lovstrom, "A new video quality predictor based on decoder parameter extraction", International Conference on Signal Processing and Multimedia Applications, 2008, pp.285-290.
- [11] S.-O. Lee, K.-S. Jung, and D.-G. Sim, "Real-time objective quality assessment based on coding parameters extracted from H.264/AVC bitstream", IEEE Trans. Consum. Electron., vol. 56, no. 2, pp.1071-1078, May 2010.
- [12] O. Sugimoto, S. Naito, S. Sakazawa, and A. Koike, "Objective perceptual video quality measurement method based on hybrid no reference framework", Image Processing (ICIP), 2009 16th IEEE International Conference on, 2009, pp.2237-2240.
- [13] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity", IEEE Trans. Image Processing, vol. 13, no. 4, pp.600-612, April 2004.
- [14] Advanced video coding for generic audiovisual services, ITU-T Recommendation H.264, Version 1 2003, Version 2 2004, Version 3 2005.
- [15] [ftp://vqeg.its.bldrdoc.gov/HDTV/SVT\\_MultiFormat/2160p50\\_CgrLevels\\_Master\\_SVTdec05/](ftp://vqeg.its.bldrdoc.gov/HDTV/SVT_MultiFormat/2160p50_CgrLevels_Master_SVTdec05/).
- [16] <http://www.elementaltechnologies.com/resources/4k-test-sequences>.