

# Novel No-Reference Video Quality Assessment Metric with Estimation of Dynamic Range Distortion

Yo-Han Kim\*, Junghyun Han\*, Hyuntai Kim\*, and Jitae Shin\*

\* School of Information and Communication Engineering, Sungkyunkwan University, Suwon, Korea

{dos95, kazikai, httsog, jtshin}@skku.edu

**Abstract**— With advances in video technology, many researchers are interested in video quality assessment to prove the advantage of the developed algorithm. Since human visual system is too complex to be formulated exactly, many researches are in progressing. But, video quality assessment has not enough accuracy now.

No-reference video quality assessment is suitable for various video streaming services like real-time streaming and broadcasting system, because it don't require additional network capacity for reference data to perform quality assessment.

In this paper, we propose a novel no-reference video quality assessment method with estimation of dynamic range distortion. To measure the performance, we obtain Mean Opinion Score (MOS) data by subject video quality test with the ITU-T P.910 Absolute Category Rating (ACR) method. And, we compared it with proposed algorithm using 3 uncompressed, and 360 encoded video sequences. Experimental results show that suggested algorithm has a high correlation with obtained MOS.

**Keywords**— No-reference video quality assessment, Dynamic range distortion, MOS

## I. INTRODUCTION

The recent advances in video technology have resulted in a widely utilization of video streaming services such as u-tube or podcast are widely used. Video service developers are interested in video quality metrics to prove the improvement of their technology. Existing quality metrics such as Mean Square Error (MSE) or Peak Signal Noise Ratio (PSNR) are not able to represent characteristics of human visual system. Thus new perceptual video quality metrics has been developed and standardized by ITU-T SG9 and SG12.

The objective video quality assessment is classified into three types by necessity of reference data: Full-Reference (FR), Reduced-reference (RR), and No-Reference (NR) video quality assessment method. Using FR or RR method, we need additional network capacity of reference data to compare with source information. Because NR method needs no redundancy, we can easily apply NR method to any video streaming service, irrespective of sender or receiver side. However, NR method has less accuracy than FR and RR methods.

NR method can be divided into No-Reference Pixel (NR-P), No-Reference Bit-stream (NR-B), and Hybrid method. NR-P uses decoded video pixel values, NR-B performed with bit-stream parameters. Hybrid method utilizes both decoded video pixel and bit-stream information.

In this paper, we propose a new parameter with estimation of dynamic range distortion to increase assessment accuracy.

The remainder of this paper is organized as follows. Section 2 presents the related work. In Section 3, proposed method is presented. Section 4 shows the experimental result. Finally, Section 5 presents conclusion.

## II. RELATED WORKS

To evaluate how much objective quality assessment reflects human visual characteristics, subject video quality assessment is performed. Our measurement follows the Absolute Category Rating (ACR) standardized in ITU-T P.910 "Subjective Video Quality Assessment Methods For Multimedia Applications" [1]. This method shows video to testers with time pattern illustrated in Figure 1. Testers determine the quality of shown video by grading five-level score (i.e., Excellent, Good, Fair, Poor and Bad). Then, the score is translated to numeral score 5~1. Gathered data are used to measure the accuracy of proposed NR video quality metrics comparing with reference method.

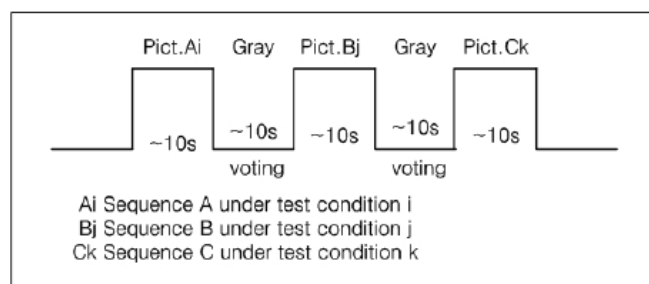


Figure 1. ACR Time Pattern for MOS Test

Since objective video quality metrics are standardized by ITU-T and Video Quality Expert Group (VQEG), many algorithms have been proposed.

Because of complexity of human visual system, perceptual quality assessment uses various parameters that are a mirror of human visual system. And, weighted coefficients of each

parameter are calculated with experimental way. All of parameters and coefficients are combined with linear equation, and those coefficients are calculated by recursive analysis based on the least squares method.

Psytechnics model [2] shown in Figure 2 is contributed to ITU-T SG9 for RRNR project. This model makes 16 parameters after performing seven analysis modules. The acquired parameters are committed to linear equation. seven kinds of analysis are listed as follows: “Temporal distortion analysis”, “Temporal complexity analysis”, “Blockiness analysis”, “Blurriness analysis”, “Image property analysis”, “Activity analysis”, and “Structure analysis”.

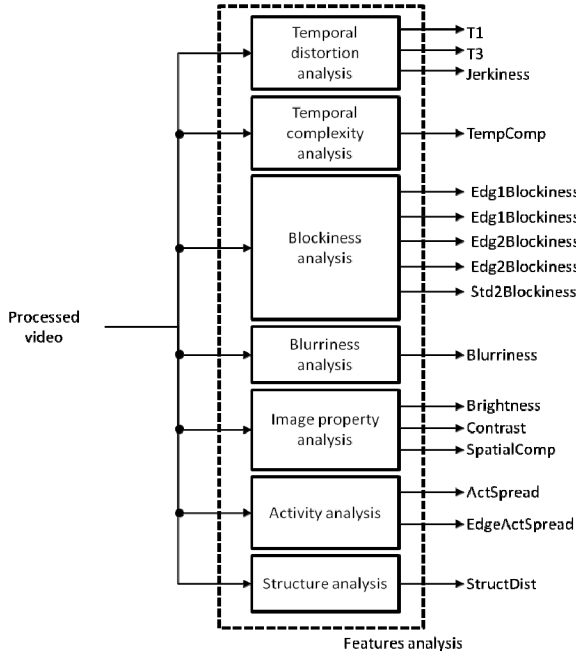


Figure 2. Psytechnics Model Contributed at ITU-T

### III. PROPOSED METHOD

In this paper, we offer new parameters that reflect distortion well from human visual system. And, we compute the weighted coefficient of parameter experimentally to find literal and linear equation. New parameter is obtained by estimation of dynamic range distortion. When we receive the video sequences, the scene is captured and converted to digital signal by quantization. Thus, digital signal has constraint range. For example, 8-bit digital signal expresses a brightness or a color within 0~255 range. To solve this problem, a camera controls dynamic range to fit range of digital signal. Since dynamic range control is performed after the scene is captured, the scene has dynamic range distortion. As shown in Figure 3.

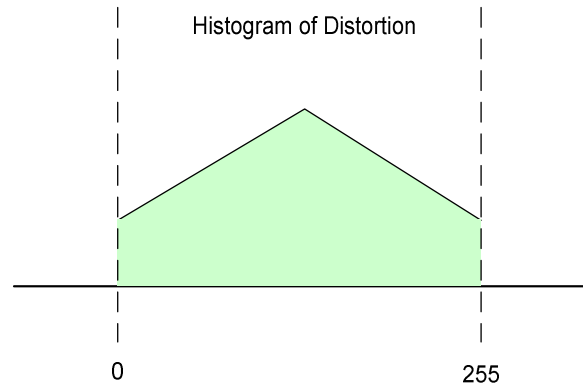


Figure 3. Dynamic Range Distortion

If dynamic range is controlled correctly, edge which has value 0 or 255 is appeared rarely in the histogram of scene. But if incomplete dynamic range is occurred, edge values are appeared incessantly. So, we focus on estimating distortion with histogram analysis

We perform the estimation of dynamic range distortion to get higher accuracy as shown in Figure 4.

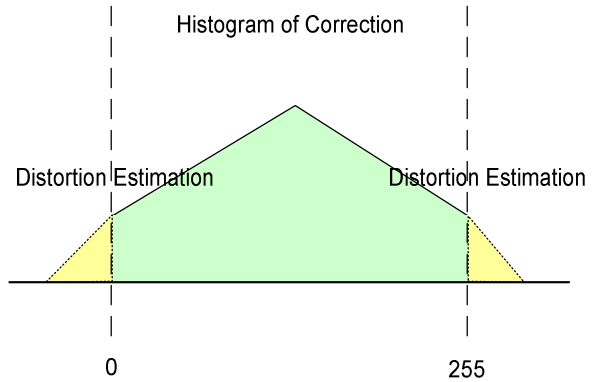


Figure 4. Proposed Method

Formulations for distortion are as below.

Let  $D_{\min}$  and  $D_{\max}$  are dynamic range distortion from low edge and from high edge, respectively. Then

$$D_{\min_n} = \frac{1}{n} \sum_{a=-1}^{-n} \{(0-a)^2 \times V_{\min} \times 2^{0-a}\} \quad (1)$$

$$D_{\max_n} = \frac{1}{n} \sum_{a=256}^{255+n} \{(255-a)^2 \times V_{\max} \times 2^{255-a}\} \quad (2)$$

where

$V_{\min}$  and  $V_{\max}$  are the number of 0 and 255 value pixels, respectively.  $n$  indicate the range of distortion estimation that is chosen between 4 and 16 with experiential value. Total dynamic range distortion  $D_n$  is formed as below

$$D_n = (D_{\max_n} + D_{\min_n}) / 2 \quad (3)$$

Since dynamic range distortions  $D_n$  is existed in a luminance and two chrominance data ( $Cb$ ,  $Cr$ ), there are three dynamic range distortion parameters to be calculated:  $D_{n,Y}$ ,  $D_{n,Cb}$ ,  $D_{n,Cr}$ .

We can define our proposed video quality assessment metric named  $eMOS$  as follow

$$\begin{aligned} eMOS = & W_1 \times Br + W_2 \times Mv + W_3 \times Ct \\ & + W_4 \times (Cb + 2Cr) + W_5 \times Ed \\ & + \sum_{i=4}^{16} W_{i+2} \times D_{i,Y} + \sum_{i=4}^{16} W_{i+15} \times (D_{i,Cb} + 2D_{i,Cr}) \end{aligned} \quad (4)$$

Where  $W_x$  is weight factor of each parameters.

$eMOS$  with some experiential values is expressed as below

$$\begin{aligned} eMOS = & 1.6 \times Br + 0.6 \times Mv + 4.4 \times Ct \\ & + 0.1 \times (Cb + 2Cr) + 0.5 \times Ed \\ & - 2.0 \times D_{4Y} - 0.3(D_{4Cb} + 2D_{4Cr}) \end{aligned} \quad (5)$$

where  $Br$  is brightness,  $Mv$  is motion analysis parameter,  $Ct$  is contrast,  $Ed$  is edge parameter,  $Cb$  and  $Cr$  is chrominance parameters.

For low complex computation, parameters having low weight factor value are excluded from equation.

#### IV. SIMULATION RESULTS

We use three video sequences for objective and subjective video quality assessments as shown as Figure 5.

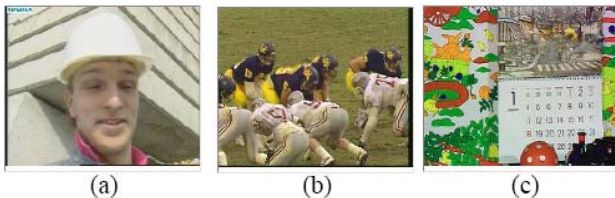


Figure 5. Test Sequences (a) foreman, (b) football, and (c) mobile

Sequence (a) named Foreman has medium motion and simple background, (b) Football has largest motion, and (c) Mobile has complex background and smallest motion. These

three sequences represent various type of video. We extend three sequences to 360 sequences with various types of error or transfer situations for reliability. Variable conditions for 360 sequences are listed at Table 1.

Table 1. Test Sequence Conditions

| Variable       | Value                       |
|----------------|-----------------------------|
| Resolution     | QVGA, QCIF                  |
| FPS            | 15, 30 Frame/Sec            |
| QP             | 28, 32, 36                  |
| Error Position | Front, Middle, End (of GOP) |
| Error length   | 1,2,3                       |
| GOP            | 15                          |

We perform subjective video quality assessment with 20 students of Sungkyunkwan University who have no concern with the development of proposed algorithm. Then, we compare  $eMOS$  with obtained subjective quality data. Pearson correlation metric is used for reliability measurement. The experimental results are as below.

Table 2. Simulation results

| Model             | Pearson Correlation |
|-------------------|---------------------|
| Psytechnics [4]   | 0.815               |
| w/o new parameter | 0.795               |
| Proposed $eMOS$   | 0.847               |

Table 2 says proposed method has higher correlation with subject quality than reference methods, figure 6 and 7 shows distribution map of reference and proposed scheme.

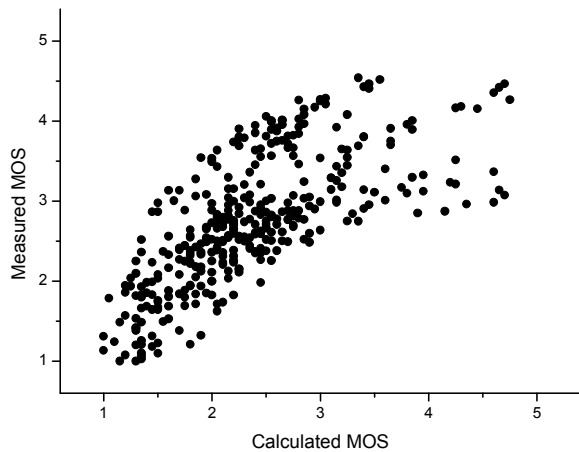


**Figure 6. Distribution map of Reference method without new parameters**

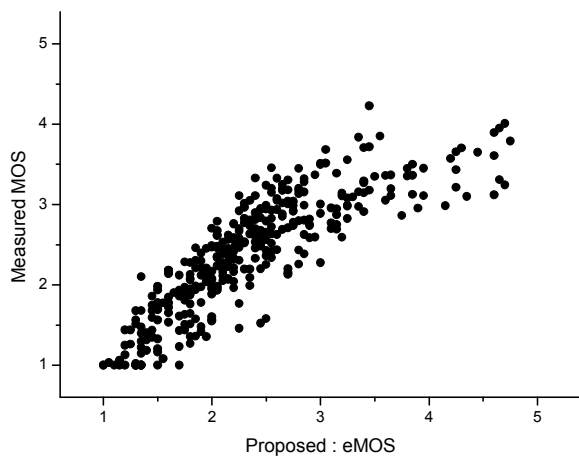


**Figure 7. Distribution map of Proposed method**

## V. CONCLUSION

In this paper, we proposed new quality reflecting parameters based on human visual system and, performed the video quality assessment with these parameters. We obtained higher correlation between subjective and objective video quality metrics with experiments, and it means that accuracy of quality assessment is increased without overhead for reference data.

Even if we acquired higher accuracy metric for no reference video quality assessment, there is still lower accuracy compared to the results from using FR or RR metrics [5]. In a future work, we need to research new parameters from transferring information. Those might give us more expectable results.

## ACKNOWLEDGMENT

This work was supported by the National Research Foundation (NRF) grant funded by the Korea government (MEST) (2009-0071289).

## REFERENCES

- [1] ITU-T "Subjective Video Quality Assessment Methods For Multimedia Applications," P.910
- [2] Psytechnics Ltd "Psytechnics no-reference video quality assessment model" *ITU-T SG9 Meeting*, COM9-C190-E, 2008.
- [3] VQEG "Hybrid / Bitstream Group test plan" JVT-Q.069 , 2007
- [4] Psytechnics Ltd "Results of VQEG Multimedia Phase I Validation Test for no-reference models" *ITU-T SG9 Meeting*, COM9-C192-E, 2008.
- [5] ITU-T Recommendation J.144 "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference,"
- [6] T. Liu, Y. Wang, et al "Subjective quality evaluation of decoded video in the presence of packet losses", *IEEE, ICASSP*, 2007