VIDEO QUALITY METRICS – AN ANALYSIS FOR LOW BIT RATE VIDEOS

Ee Ping Ong, Mei Hwan Loke, Weisi Lin, Zhongkang Lu, and Susu Yao

Institute For Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613
Email: {epong, mhloke, wslin, zklu, ssyao}@i2r.a-star.edu.sg

ABSTRACT

Objective video quality metrics reported in the literature have so far been focused on TV signals with large frame sizes, full TV frame rates, and high compressed bit rates. The aim of this paper is to provide an analysis of 4 objective video quality metrics tested on low bit rate videos with small frame sizes, various frame rates, and low bit rates. The video quality metrics used were the NTIA Video Quality Metric, Modified Watson's Digital Video Quality metric, Video Structural Similarity Metric, and the Perceptual Video Quality Metric. The test videos consist of H.264 compressed videos with CIF and QCIF frame sizes, at various bit rates (24kbps-384kbps) and frame rates (7.5fps-30fps). Here, results and analysis derived from the comparison of the metrics will be provided.

Keywords: video quality metric, video coding, TV signals, multimedia videos, subjective test

1. INTRODUCTION

The growing digital video industry brings the need for a standardized benchmark to measure the objective quality of digital videos. A reliable video quality metric (VQM) would be useful in optimizing video coding algorithms and examining hypothetical reference circuit (HRC). Presently, there are metrics that give acceptable performance for digitally compressed high to medium bit rate video sequences [2,9], but they are either not illustrated or fail for measuring the visual quality of low bit rate videos. This is a realistic problem as there is increasingly more applications that operate at low bit rates (e.g. 3G mobile videophone applications). A good VQM should reflect the human visual system (HVS), which has limited sensitivity in lower spatial and temporal frequencies [1,2]. The difficulty of creating a good VQM lies in the nonlinear behavior of HVS, and deciding what parameters to be extracted for measurement. Due to the lack of VQM that has been explicitly illustrated to work well on low bit rates, lower frame rates, and small frame size videos, the Video Quality Expert Group (VQEG) is in the progress of investigating and consolidating contributions for video quality metrics designed for such multimedia videos [8].

In this paper, we analyze the performance of 4 objective video quality metrics on low bit rate videos. The video quality metrics that will be discussed are the National Telecommunication and Information Administration Video Quality Metric (NTIAVQM) [3], a modified Watson's Digital Video Quality metric (MWDVQM) [4], the Video Structural Similarity metric (VSSIM) [5], and the Perceptual Video Quality Metric (PVQM) [11]. The paper is organized as follows: Section 2 gives a brief overview of the metrics, Section 3 provides experimental tests conducted, results, and analysis, and finally Section 4 gives a conclusion of the comparison study being made.

2. OVERVIEW

2.1. National Telecommunication and Information Administration Video Quality Metric

The NTIAVQM [3] has 5 video quality models that extract different parameters to be compared. For this paper, only the "Videoconferencing" video quality model is being examined in detail as this model has been optimized for low bit rate videos. This VQM extracts parameters from both the processed and the original video sequences and compares the features. Parameters which this VQM used were features to measure spatial impairments, distortions in chrominance signals, localized contrast information, and distortions in motion flow. Feature extraction involves applying a perceptual filter, dividing the video sequence into spatial-temporal regions and extracting the needed parameters. The second part uses a comparison function to compare the extracted parameters, followed by spatial and temporal collapsing.

2.2. Modified Watson's Digital Video Quality Metric

This MWDVQM, based on Watson's Digital Video Quality metric [6,7], computes the visibility of artifacts expressed in the DCT domain via DCT coefficients. This algorithm first converts the processed and reference video sequences to YOZ color space, and then performs DCT transformation.

The DCT coefficients are converted to units of local contrast, which is defined as the ratio of the AC amplitude to the temporally low-pass filtered DC amplitude. The local contrasts are subjected to spatial contrast sensitivity functions for the static and dynamic frames, and the DCT coefficients are converted to just noticeable differences. The video sequences are subtracted to produce a difference sequence, and this is subjected to a contrast masking in a maximum operation and a weighted pooling mean distortion.

2.3. Video Structural Similarity Metric

The Video Structural Similarity metric (VSSIM) [5] measures the luminance, contrast, and structure of signals between two video sequences. For discrete signals, luminance is estimated as the mean intensity. The contrast is estimated as the standard deviation after removal of luminance from the signal. The 3 components are subjected to their respective comparison functions before being pooled into an overall similarity measure equation.

2.3. Perceptual Video Quality Metric

The Perceptual Video Quality Metric (PVQM) [11] operates in the spatial domain (directly on the video frames) and uses both local and global features for quality measurements. The local features include colour masking, spatial-textural masking, and temporal masking. The global features include content-richness fidelity and block-fidelity. The local masking features are combined into a visibility threshold value and together with sequence difference and global features are pooled to give an objective video quality rating.

3. RESULTS AND DISCUSSIONS

3.1. Test Conditions

A total of 90 QCIF and CIF video sequences were generated from 12 reference sequences ("Coast Guard", "Container", "Foreman", "Japan League", "News" and "Tempete"). They were subjected to H.264 video coding, with different bit rates (24kbps to 384kbps) and frame rates (7.5Hz to 30Hz). Each of the video sequences consists of 250 frames. The VQMs were run on each of the video sequence under test.

3.2. Subjective Test Method

The subjective video quality tests of the test video sequences have been carried out as the tests conducted for the evaluation of JVT video sequences [10], using Double-Stimulus Impairment Scale variant II (DSIS-II) subjective test method and performed by 20 subjects. The decoded sequences with frame rate lower than 30 fps are displayed with repeated frames on the 30 Hz display device when performing the subjective test.

3.3. Performance

Two performance measures are used for comparison: Pearson correlation coefficient (r_p), and Spearman rank-order correlation coefficient (r_s). Ideally, r_p and r_s should be 1. These measures are obtained from the logistically-fitted data [9].

Table 1 shows the Pearson and Spearman correlation results of the 3 VQMs. The upper bound (UB) and lower bound (LB) of the Pearson correlation were calculated with a confidence interval of 95%.

	r_p	r_p UB	r_p LB	r_s
PSNR	0.701	0.793	0.578	0.676
NTIAVQM	0.747	0.826	0.639	0.722
MWDVQM	0.676	0.775	0.545	0.629
VSSIM	0.593	0.713	0.440	0.599
PVQM	0.916	0.944	0.875	0.920

Table 1: Performance of the VQMs

Figure 1-4 shows the scatterplot of subjective ratings versus the NTIA, MWDVQM, VSSIM, and PVQM video quality ratings respectively with respect to (a) frame rates and (b) bit rates.

It can be seen that PVQM gives the best performance, followed by the NTIAVQM, PSNR, MWDVQM, and finally the VSSIM method. It seems that those methods designed for low bit rate videos perform much better than PSNR, while those not designed for low bit rate videos perform poorer than PSNR. The NTIAVQM does not take the measurement of chrominance components into its computations whilst the other metrics do.

While the MWDVQM showed a less linear relationship, it was noted in [4] that the training database used was small compared to the databases used for the other algorithms. The VSSIM is a computationally-simple method but does not address some problems, such as vertical distortions being more noticeable than horizontal distortions.

In addition, it seems that the 2 best methods, namely NTIAVQM and PVQM, seem to have poorer correlation to subjective ratings at 7.5 Hz frame rate (as seen by a wider dispersion of the 7.5 Hz data) as compared to 30 Hz frame rate. Meanwhile, it is difficult to see distinct pattern from the bit rate plots as there are insufficient 384 kbps data that span across different subjective ratings. However, PVQM seems to provide very good results at 24 kbps bit rate (vs NTIAVQM).

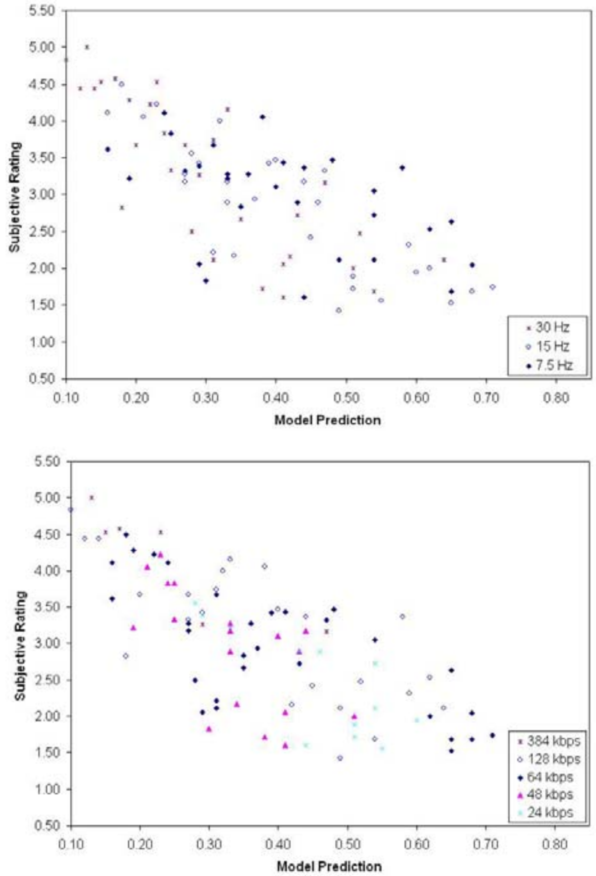
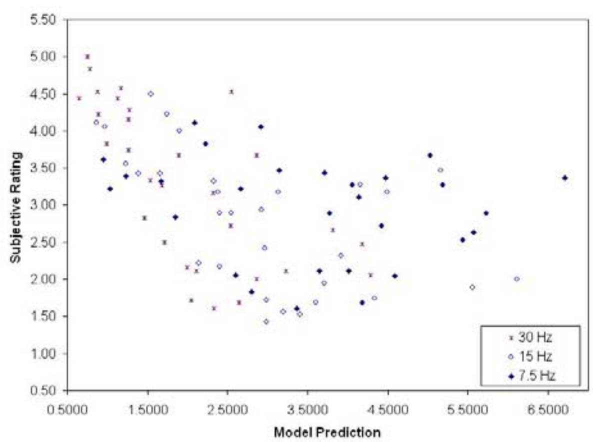


Figure 1: Scatterplot of NTIAVQM: (a) with respect to frame rate; (b) with respect to bit rate



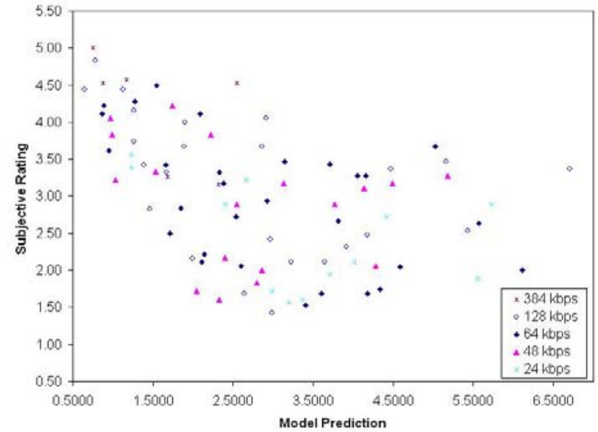


Figure 2: Scatterplot of MWDVQM: (a) with respect to frame rate; (b) with respect to bit rate

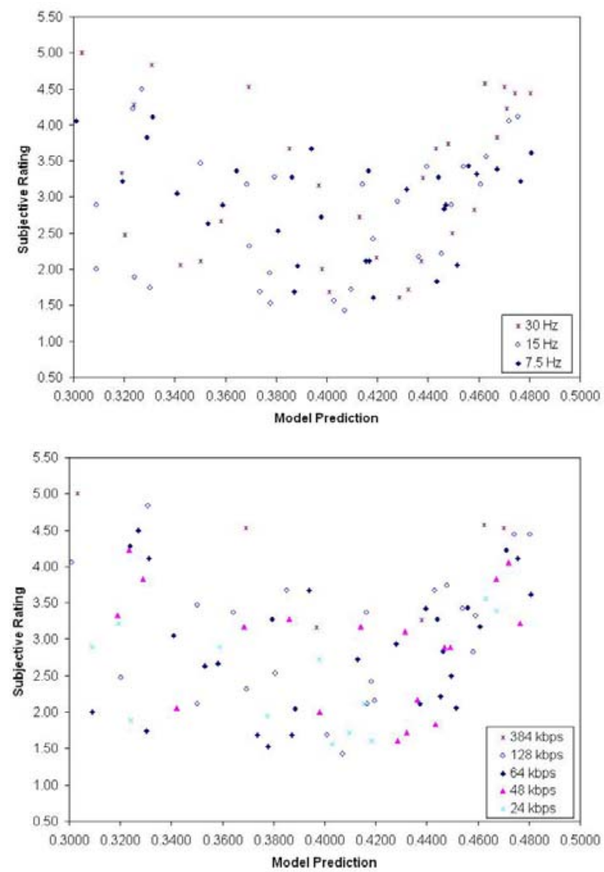


Figure 3: Scatterplot of VSSIM: (a) with respect to frame rate; (b) with respect to bit rate

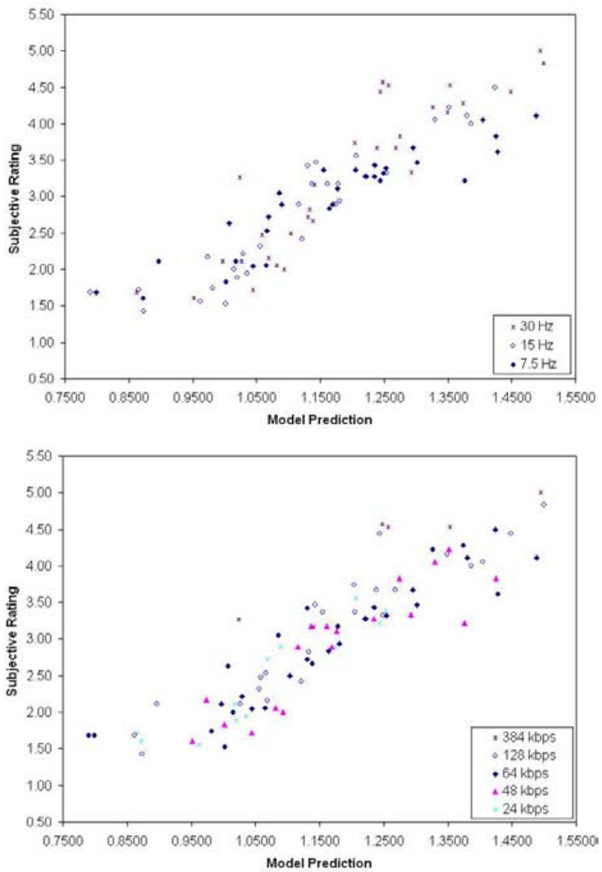


Figure 4: Scatterplot of PVQM: (a) with respect to frame rate; (b) with respect to bit rate

The test sequences used were all low-bit rate sequences. The video metrics designed for TV signals (high bit rate videos) do not perform as well in the low bit rate range as they do in the mid-high bit-rate range because different artifacts dominate in the lower bit-rate ranges. For example, noise would be less noticeable compared to blocking artifacts in the low bit-rate video. A characteristic of low bit-rate videos is that it results in serious blocking artifacts (e.g. MPEG-4) because block motion detection can straddle objects, and adjacent blocks may have very different estimated motion vectors. In the case of H.264 compressed videos, the low bit rate videos look blurred across block boundaries, mainly because of the use of in-loop filter.

4. CONCLUSION

This paper has presented a description of 4 objective video quality metrics, followed by a comparison of their results and conclusions that can be drawn from the comparison. This study suggests that current state-of-the-art objective video quality metrics that so far have been designed for quality measurement of TV signals with large frame sizes, full TV frame rates, and high compressed bit rates may not work well on multimedia videos with low bit rates, various frame rates, and small frame sizes. The experiments here show that quality metrics designed specifically for low bit rates perform better than PSNR when applied on low bit rate videos while those quality metrics not designed for such application performed poorer than PSNR. It is encouraging that a major step has been taken by the Video Quality Expert Group (VQEG) which is in the progress of investigating and consolidating contributions for video quality metrics designed for such multimedia videos [8].

5. REFERENCES

[1] Miyahara, M., Kotani, K. and Algazi, V.R., “Objective picture quality scale (PQS) for image coding”, *IEEE Transactions on Communications*, Vol. 46(9), 1998, pp. 1215-125.

[2] Video Quality Expert Group, “Final Report from the VQEG on the validation of Objective Models of Video Quality Assessment, Phase II”, *www.vqeg.org*, August 2003.

[3] Stephen Wolf, Margaret Pinson, “Video Quality Measurement Techniques”, *NTIA Report 02-392*, June 2002.

[4] Feng Xiao, “DCT-based Video Quality Evaluation”, *MSU Graphics and Media Lab (Video Group)*, Winter 2000.

[5] Zhou Wang, Alan Conrad Bovik, Ligang Lu, “Video Quality Assessment Based on Structural Distortion Measurement”, *IEEE Signal Processing: Image Communication, Vol 19, No 2. pp. 121-132,* February 2004.

[6] A.B. Watson, “Towards a perceptual videoquality metric”, *Human Vision, Visual Processing, and Digital Display VIII*, 3299, 139-147.

[7] A.B. Watson, James Hu, John F McGowan III, “DVQ: A digital video quality metric based on human vision”, *Journal of Electronic Imaging*, Vol. 10(1), 20-29.

[8] VQEG (Video Quality Expert Group), “Multimedia Group Test Plan”, Draft Version 1.13, July 2006.

[9] VQEG (Video Quality Expert Group), “Final Report from the VQEG on the validation of Objective Models of Video Quality Assessment”, *www.vqeg.org*, March 2000.

[10] Baroncini, V., ISO/IEC JTC 1/SC29/WG11, 4240, Sidney, July 2001.

[11] Ong, E.P., Lin, W., Lu, Z., Yao, S., “Colour Perceptual Video Quality Metric”, *IEEE International Conference on Image Processing*, Genoa, Italy, 11-14 Sep 2005.