PERCEPTUAL QUALITY EVALUATION ON PERIODIC FRAME-DROPPING VIDEO

Zhongkang Lu†, Weis Lin‡, Boon Choong Seng§, Sadaatsu Kato§, Eeping Ong†and Susu Yao†

†Institute for Infocomm Research, Agency for Science, Technology and Research,
21 Heng Mui Keng Terrace, Singapore 119613;
‡School of Computer Engineering, Nanyang Technological University,
Nanyang Avenue, Singapore 639798.
§Research Laboratories, NTT DoCoMo, Inc., Kanagawa, 239-8536 Japan.

ABSTRACT

In general, low and very-low bitrate video communication systems cannot achieve the full frame rate (25Hz for PAL or 30Hz for NSTC), which brings temporal distortion. The influence of reduced frame rate on subjective quality evaluation is a very important topic for both perceptual quality metrics and communication system optimization. In this paper, we present our work on the numerical modeling of the influence. The work includes two parts: first we measured the detectability and annoyance of periodic frame dropping's effect on perceptual visual quality evaluation under different content and frame size conditions. Then, a simple and effective feature is proposed to represent the content of video in temporal quality evaluation. The high Pearson and Spearman correlation results between the MOS and proposed model, as well as the results of other two error metrics, confirm the success of the selected temporal content-quality feature.

Index Terms— Visual Quality, Periodic Frame-dropping, Temporal Content-Quality Feature

1. INTRODUCTION

The distortion of a very-low bitrate compressed video generally includes two parts: spatial distortion and temporal distortion. Considering the spatial and temporal information are processed in different cortex in human brain, the subjective evaluation of the distortion can be modeled by a nonlinear combination of spatial distortion and temporal distortion. Up to now, most of current video quality metrics only consider spatial distortion. Moreover, because the temporal distortion in very-low bitrate compressed video belongs to suprathreshold distortion condition [1], temporal CSF doesn't work well in the application [2] [3]. This work is to isolate the subjective evaluation of temporal distortion from spatial distortion, and to establish a model to accurately reflect the relationship among subjective evaluation of temporal distortion, frame rate and content of video.

Much research showed that the perception of continuous motion is very complex. It concerns a manifestation of complex functions. For example, the perception of first-order (luminance-defined) motion and second-order (contrast-defined) motion are processed in different areas in the human brain, and the associated processing mechanisms are very different [5] [6]. Substantial psychological and physiological research also showed that the sensitivity of apparent motion perception can be analyzed by a number of spatiotemporal energy models [7], which suggests that the motion content in video plays an important role on temporal quality evaluation.

This paper aims at the modeling of the influence of periodic frame-dropping on visual quality perception. The model is established under two conditions:

1. no-reference condition, which means the information of original sequences are not used; and

2. prediction condition, which means the input to the model is the motion representation of original sequences and the target frame rate.

The proposed models not only can be used into the design of visual quality metrics (first condition), but also on rate-distortion optimization for very-low bitrate video communication (second condition).

The work includes two parts: first is a subjective experiment to measure the detectability and annoyance on periodic frame dropped video sequences, without the introduction of spatial distortions. Then, based on the subjective experiment results, a simple and effective feature is found to represent the content of video in temporal quality evaluation. The paper is organized as following: section 2 gives an introduction of the subjective experiment. The proposed temporal content-quality feature is presented in section 3. Experimental results and analysis are given in section 4. Section 5 is the conclusions and future works.

2. SUBJECTIVE VIEWING EXPERIMENT

The subjective viewing experiment is to measure the detectability and annoyance of frame dropping's effect on perceptual

1-4244-1437-7/07/$20.00 ©2007 IEEE

Table 1. The characteristics of monitor used in the subjective testing experiment.

Monitor size:	2.4 inch
Display area:	26.72mm × 48.96mm
Pixel pitch:	0.153mm × 0.153mm
Resolution:	QVGA(320 × 240)
Number of colors:	262,144 (18bit)
Refresh rate:	60Hz
Luminance:	150.6cd/m2
Contrast:	180.25:1

visual quality evaluation under different motion and size conditions. In total eleven standard video test sequences are chosen in the test. They are: “Bus”, “Coastguard”, “Container”, “Goldfish”, “Hall monitor”, “Mobile”, “Mother daughter”, “Paris”, “Stephan”, “Table”, and “Tempete” . We can see all these sequences cover almost all kinds of motion types, i.e.,. from very slow motion (”Container”) to fast motion (”Bus” and ”Stephan”), rigid motion (”Mobile”) to non-rigid motion (”Goldfish”). The original sequences all include 260 frames at frame rate of 30Hz without interlacing. They are all 8-second long with 10 frames of redundancies at beginning and ending.

Two test conditions are used is the experiment, with both CIF and QCIF formats. The QCIF sequences are obtained by down-sampling from CIF sequences. Six test points are selected for every test sequence in each picture format: 30Hz, 15Hz, 7.5Hz, 6Hz, 5Hz, and 3Hz. The five test points with the discontinuity caused by the frame dropping process were obtained from the 30Hz video sequences by retaining only the first frame and discarding all the other (n-1) frame(s) for every n consecutive frames (n = 2, 4, 5, 6, and 10). In the subjective viewing, the discarded frame(s) are replaced by the first frame in the group. Please note that spatial distortions are not introduced into the test sequences so that the video quality measured in the experiment is purely caused by frame dropping.

Double-Stimulus Impairment Scale Method (DSIS) [8] was used to evaluate the subjective quality. The semantic of each of the 5 grade is ”Imperceptible”, ”Perceptible, but not annoying”, ”Slightly annoying”, ”Annoying”, and ”Very annoying” from score 5 to 1.

The test was carried out in a normal lab environment. The background luminance is set up to a comfortable level by the viewer. All test sequences were stored and played back on a PC station. Specially designed display devices in the form of mobile phones at QVGA resolution are used. The characteristics of the monitor is shown in table 1

No instructions are given to the viewers on the viewing distance. The viewers can choose the viewing distance that they feel comfortable. All these setups are targeted at normal personal viewing environment on mobile or hand-held devices. In the test, the CIF sequences (352 × 288) were cropped to the QVGA size (320 × 240) because of the limitation of the display. The right 32 pixels and the bottom 48 pixels were cropped from the CIF sequences. QCIF sequences were displayed and evaluated in the original sizes. For every test sequence, 23 subjects are employed for each test point and each test condition. The viewers don’t have any knowledge or experience on image processing technologies. Mean Opinion Score (MOS) are therefore obtained.

3. TEMPORAL CONTENT-QUALITY FEATURE

The results of the subjective viewing experiment are shown in figure 1(a) and (b), which compare the MOS values against PSNR values. The PSNR values are obtained by:

$$MSE = \frac{1}{MNT}\sum_{t=0}^{T-1}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\|I^2(i,j,t) - \hat{I}^2(i,j,t)\| \quad (1)$$

$$PSNR = 20 \cdot \log_{10}(\frac{255}{\sqrt{MSE}}) \quad (2)$$

where $I(i,j,t)$ and $\hat{I}(i,j,t)$ denote the original and distorted video sequence, respectively. T, M and N are the video dimensions. Moreover, the curves shown in the figures are fitted by the 4-parameter exponential logistic function recommended in [9].

Besides the figures reflect the fact that PSNR cannot correclty represent the influence of periodic frame dropping on perceptual quality evaluation, we also can see from the figures that the content of videos dominate the subjective viewing experiment results. In this paper, a simple and effective feature is proposed to represent the content of video in temporal quality evaluation. We name it AMMF (Average Maximal Motion by Frame), which can be represented as an average of every frame’s maximal motion offset values. Its mathematic expression is below:

$$AMMF = \sum_i \frac{\max(MotionMap_i)}{N-1} \quad (3)$$

where $MotionMap_i$ is the motion vector map between effective frames i and $i+1$, N is the number of connected frames, and the motion vector map is estimated by M. J. Black’s optical flow algorithm [10].

The choosing of AMMF as the temporal content-quality feature implies that the highest motion region in video plays an important role on subjective temporal quality evaluation. Please note that because of local constraints applied on Black’s motion estimation algorithm, AMMF represents the motion strength of the highest motion region in video more than the real maximal motion.

Based on the selected feature, an optimized non-linear optimized logistic function is also proposed:

$$\begin{aligned} MOSp &= 5 - (a_1 + a_2 * AMMF) * \\ & [log(30) - log(fr)]^{a_3 + a_4 * AMMF} \end{aligned} \quad (4)$$

where fr denotes the frame rate.

Moreover, we can see that the size of display is not included in the model. The reason is that only two frame sizes are used in the subjective viewing experiment. The number of test point on frame size is not enough to establish a reliable model.

4. EXPERIMENTAL RESULTS AND ANALYSIS

Figure 1 compares PSNR and the proposed model in scatter plots. We can see from them that the model based on AMMF outperforms PSNR.

Furthermore, to evaluate the accuracy of the equation 4, four functional parameter sets are fitted. They are:

1. Functional Parameter Set 1 (FPS1): The input of equation 4 for the set of parameters are $AMMF$ values estimated from frame-dropped CIF sequence and the framerate. The information of original sequence is not included.
2. Functional Parameter Set 2 (FPS2): The inputs are same to Functional Parameter Set 1, except the frame-size of the sequences are QCIF.
3. Functional Parameter Set 3 (FPS3): The input of equation 4 for the set of parameters are $AMMF$ values estimated from original 30Hz CIF sequence and the target framerate of frame-dropped sequence. The $AMMF$ value of frame-dropped sequence is not used.
4. Functional Parameter Set 4 (FPS4): The inputs are same to Functional Parameter Set 3, except the frame-size of the sequences are QCIF.

Obviously, the first two sets of functional parameters can be used for no-reference visual quality evaluation; and the latter two sets of parameters are trained for visual quality prediction, which can be a part of content-based rate-distortion optimization in video compression. Comparisons of the fitting of the four functional parameter sets on the results of subjective experiment are shown in table 2. Four metrics are used to measure the similarity between original MOS values and predicted MOSp values. They are:

1. Pearson Correlation Coefficient (PCC)
2. Spearman Correlation Coefficient (SCC)

Table 2. Comparisons of the results of four metrics by fitting the four functional parameter sets on equation 4 (*with AMMF values*).

	FPS1	FPS2	FPS3	FPS4
PCC	0.9679	0.9752	0.9721	0.9740
SCC	0.9948	0.9948	0.9948	0.9948
RMSE	0.2276	0.2210	0.2426	0.2219
MAD	0.1716	0.1610	0.1842	0.1596

3. Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{\sum (MOS - MOSp)^2}{n}} \quad (5)$$

where $MOSp$ denotes the predicted MOS value by fitting.

4. Mean Absolute Difference (MAD):

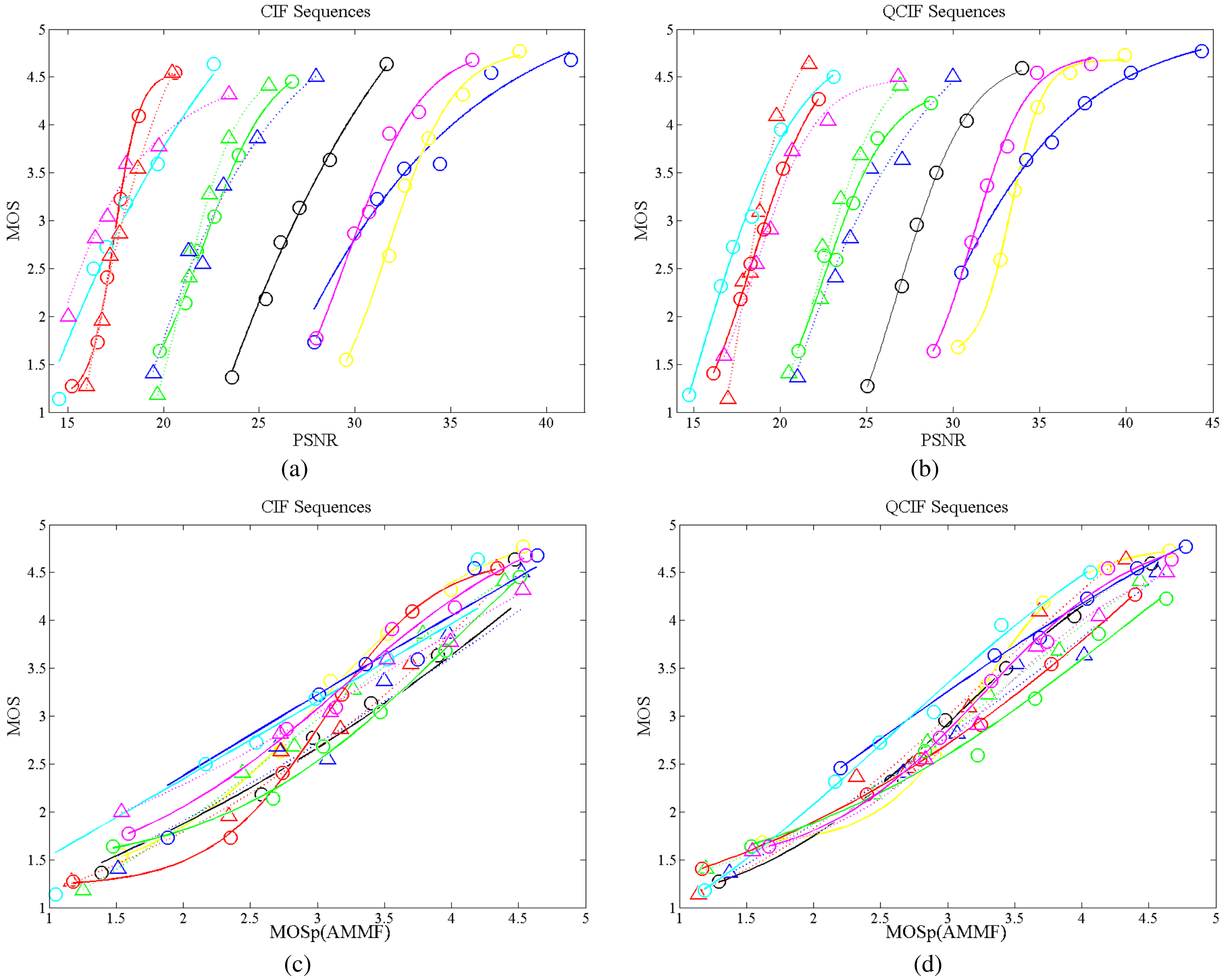
$$MAD = \frac{|MOS - MOSp|}{n} \quad (6)$$

Among the four metrics, the values of Pearson Correlation Coefficient and Spearman Correlation Coefficient is higher, the two sets of data are more similar; and for the latter two metrics, smaller values means closer. Considering the work is a part of no-reference visual quality metric for very-low bitrate compression video and content-based rate-distortion optimization control, the improvement is critical to the performance of these two applications.

5. CONCLUSIONS AND FUTURE WORKS

The paper first introduces the subjective viewing experiment to measure the detectability and annoyance of periodic frame dropping's effect on perceptual visual quality evaluation under different content and frame size conditions. Based on the subjective experiment results, a simple and effective temporal content-quality feature, AMMF, is selected to model the relationship among video content, frame rate and temporal visual quality. The high correlation results between the MOS and predicted MOSp confirm the success of the selected feature.

The work presented in the paper is useful for visual quality metric design, perceptual definition of spatiotemporal rate-distortion optimization for very-low bitrate video compression, and active perceptual rate control for narrow-band communication applications. Among the functional parameter sets, Sets 1 and 2 can be used for the first application, and Sets 3 and 4 can be used for the other two applications.



(a)

(b)

(c)

(d)

Fig. 1. Scatter plots of MOS versus prediction models: (a) PSNR on CIF sequences; (b) PSNR on QCIF sequences; (c) MOSp(AMMF) on FPS1; and (d) MOSp(AMMF) on FPS2;

6. REFERENCES

[1] Zhongkang Lu, W. Lin, X. Yang, EePing Ong, and Susu Yao, “Modeling visual attention’s modulatory aftereffects on visual sensitivity and quality evaluation,” *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1928–1942, Nov. 2005.

[2] R. R. Pastrana-Vidal, J.-C. Gicquel, C. Colomes, and H. Cherifi, “Sporadic Frame Dropping Impact on Quality Perception,” in *SPIE conference on Human Vision and Electronic Imaging IX, Proceedings of SPIE Vol. 5292*, B. E. Rogowitz and T. N. Pappas, eds., (San Jose), Jan. 2004.

[3] R. R. Pastrana-Vidal, J.-C. Gicquel, C. Colomes, and H. Cherifi, “Frame Dropping Effects on User Quality Perception,” in *5th International Workshop on Image Analysis for Multimedia Interactive Services*, April 2004.

[4] ANSI, “American National Standard for Telecommunications - Digital Transport of Video Teleconferencing/Video Telephony Signals - Performance Terms, Definitions, and Examples,” 1996. T1.801.02-1996.

[5] T. Ledgeway and A. T. Smith, “Evidence for Separate Motion-Detecting Mechanisms for First- and Second-Order Motion in Human Vision,” *Vision Research* **34**, pp. 2727–2740, Oct. 1994.

[6] A. M. Derrington, H. A. Allen, and L. S. Delicato, “Visual Mechanisms of Motion Analysis and Motion Perception,” *Annual Review of Psychology* **55**, pp. 181-205, Feb. 2004.

[7] E. H. Adelson and J. R. Bergen, “Spatiotemporal Energy Models for the Perception of Motion,” *Journal of the Optical Society of America A* **2**, pp. 284–299, Feb. 1985.

[8] ITU-R, “Methodology for the Subjective Assessment of the Quality of Television Pictures,” 2002. ITU-R Rec. BT. 500-11.

[9] VQEG (Video Quality Export Group), “VQEG Subjective Test Plan,” Feb. 1999.

[10] M. J. Black and P. Anandan, “The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields,” *Computer Vision and Image Understanding* **63**, pp. 75–104, Jan. 1996.