Video Quality Metric for Consistent Visual Quality Control in Video Coding

Long Xu†, ‡, King Ngi Ngan†, Song Nan Li†, Lin Ma†
†Electronic Engineering Department, the Chinese University of Hong Kong, Hong Kong, China
‡School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China
{xulong, knngan, snli, lma}@ee.cuhk.edu.hk

***Abstract*—The visual quality consistency is one of the most important issues in video quality assessment (VQA). When people view a sequential video, they may have an unpleasant perceptual experience if the visual quality of video frames is inconsistent even though the average visual quality of the video is not too bad. Thus, the consistent visual quality control is mostly expected in real-time video communication. Additionally, in conventional video communication, the channel bandwidth and buffer resources are limited. The unfair distribution of encoding resources among video frames would result in not only inconsistent visual quality but also other types of spatial distortions. In this paper, a new objective visual quality metric (VQM) is firstly proposed for measuring the video quality in video coding. It makes full use of the information of video coding without extra computational complexity. Secondly, a visual quality control algorithm is proposed to ensure the consistent visual quality of video coding under the given channel and buffer resources. Finally, the experimental results indicate that the proposed VQM is consistent well with the human visual system (HVS). In addition, the consistent visual quality, better rate-distortion efficiency, accurate bit control and compliant buffer can be achieved by the proposed visual quality control algorithm.**

I. INTRODUCTION

In recent years, there has been an increasing interest in visual quality assessment that measures the perceptual visual quality of images and videos. Since humans are the ultimate receiver of the visual signal, the most accurate way of assessing image/video quality is to ask humans for their opinions of the quality of an image or video, which is known as the subjective visual quality assessment. Researchers usually need to perform the subjective experiments to validate the objective Visual Quality Metrics (VQMs) and compete with benchmarking metrics. The subjective scores are first gathered from the subjective experiments to compare with the values computed from the objective VQMs. Then, the correlation between the subjective scores and objective metrics is analyzed to conclude whether the proposed objective metrics are good at measuring the human perception of image/video quality. The methods of measuring such a correlation include Linear Correlation Coefficient (LCC), Spearman's Rank Order Correlation Coefficient (SROCC) Root Mean Square Prediction Error (RMSE). According to the recommendation of the International Telecommunication Union (ITU) [1][2], a number of viewers are asked to rate the images or videos in the subjective experiments, and the scores of the viewers are processed as the Mean Opinion Score (MOS) or the Difference Mean Opinion Score (DMOS). The subjective experiment is actually fundamental to rank a proposed objective metric. The publically available databases of subjective scores and test material were reported in [3][4][5] for quality degradation of compression and error-prone channels. Most researchers developed their objective metrics based on these available subjective quality databases.

The objective quality metric aims at automatically predicting human perceptual behavior in evaluating image or video quality. It is convenient and computationally efficient in the real-world applications. Traditionally, Mean Squared Error (MSE) / Peak Signal to Noise Ratio (PSNR) were used to evaluate image or video quality. Almost all image/video compression standards use MSE/PSNR to measure the visual quality of the compressed signal. However, MSE/PSNR does not correlate well with the Human Visual System (HVS) [6]. Thus, a host of image/video quality metrics have been proposed in the last decade. These quality metrics can be generally categorized into two classes [7]. One focused on psychologically modeling the human perception of the visual signal. The signal was decomposed into multiple channels to simulate the tuning properties of the HVS. Luminance masking, contrast sensitivity function and contrast masking models are typically used to obtain visibility thresholds for each channel. The state-of-the-art HVS-model-based video metrics, such as Moving Pictures Quality Metric (MPQM) [8], Perceptual Distortion Metric (PDM) [9] and the Sarnoff JND [10] vision model, filter the videos using one band-pass and one low pass filter along the temporal dimension. The Digital Video Quality (DVQ) metric [11] and the scalable wavelet based video distortion index [12] utilize a single low pass filter along the temporal dimension. In [13], a VQM is developed based on the spatio-temporal distortions through a temporal analysis of spatial perceptual distortion maps. The other class of quality metrics is based on the image structural

The work was partially supported by the Research Grants Council of Hong Kong (Project CUHK415712) and National Science Foundation of China (61202242).

similarity. There exist strong relations between pixels in nature images which show the obvious structural information to the HVS. Since the HVS is highly adapted for extracting structural information from the scene, a measure of the structural similarity can provide a good approximation to the perceived image quality. In [14][15], Wang, et. al proposed an objective metric named Structural SIMilarity index (SSIM) to rank the image quality, which was further extended into multiple-scale SSIM (MSSIM) [16] to adapt to the variable viewing conditions. In [17], a statistical model of human motion perception [18][19] was used to perceptually weight the spatial and temporal pooling processes. In [20], a computationally efficient motion compensated SSIM (MC-SSIM) along temporal trajectories was proposed for video quality assessment.

To provide a VQM for visual quality control in general CBR encoding, the state-of-the-art algorithms [17][21][22] are investigated, which are computationally efficient and highly correlated with video compression operation. A new objective VQM is proposed by referring to [17][21][22]. Then, a new VQM-based Rate-Distortion (R-D) model is constructed for measuring the visual quality consistency . Moreover, a new visual quality control algorithm is developed to output consistent visual quality based on the proposed VQM-based R-D model and the previously proposed window-level rate control algorithm [23].

The rest of this paper is organized as follows. In Section II, a new objective VQM is proposed to measure the visual quality of compressed video. Section III proposes a VQM-based R-D model and a new visual quality control algorithm for consistent visual quality control of video coding. Section IV shows the experimental results for evaluating the efficiencies of the proposed VQM and rate control algorithm. Finally, a brief conclusion is given in the last section.

II. THE PROPOSED VIDEO QUALITY METRIC

For the conventional real-time video encodings, the computational complexity is important for developing a VQA algorithm. In the reference software of H.264/AVC, SSIM has been integrated as an optional measurement of R-D performance of video coding. However, it does not have an explicit relation with Quantization Parameter (QP) in video compression. In this section, a new VQM is proposed taking into account the aspects of image texture, video content variation and temporal motion speed. The motivation of the proposed VQM is to make full use of the available information in the temporally Motion Estimation/Motion Compensation based video compression standards with very limited overhead computations.

In [21], a VQM was developed to evaluate compressed videos based on the spatial texture content and MSE as

$$VQM_m = \frac{1}{k_m} \times MSE_m \quad (m = 0,1,...,M-1) , \qquad (1)$$

where k_m is a weighting factor for the m-th MB, VQM_m is the visual quality level of the m-th MB and M is the number of MBs in a frame. Since the MSE change of low-detailed videos contributes more to visual quality change than that of high-detailed videos, k_m severs as a weighting factor in (1). In [21], k was computed by the spatial edge strength, so an edge detection process was employed and the computational efficiency was compromised. In this work, the temporal disparity between two consecutive frames is firstly computed. Secondly, two spatial disparities: $I(x,y)$-$I(x$-$2,y)$ and $I(x,y)$-$I(x,y$-$2)$ are calculated along horizontal and vertical directions respectively within a frame. The sum of squares of the two spatial disparity images produces a spatial disparity image. The combination of them produces a disparity image, which can approximate the edge strengths of image blocks as

$$k_m = \sum_{(x,y)\in \mathrm{MB}} \left|S(x,y)\right| + \left|T(x,y)\right| , \qquad (2)$$

where $S(x,y)$ and $T(x,y)$ represent the spatial and temporal disparity images respectively. Eq. (2) is much more computationally efficient than [21].

In [17], a motion information content and perception uncertainty were defined and used to assess video quality. A spatio-temporal weighting function was defined as

$$w = (\alpha \log v_r + \beta) - (\log v_g - \gamma \log c + \delta) , \qquad (3)$$

where v_r, v_g represent the relative and global motion, α, β, γ and δ are model parameters. The first component of the right part of (3) measures the motion information content, and the second component represents the perception uncertainty. The model indicates that an object with significant motion with respect to the background would be a strong surprise to the visual system. However, when the background motion is too large, the HVS cannot identify the objects with the same accuracy as in static background. In this section, a motion activity term is defined as

$$w_m = 0.5 \times \sqrt{\frac{\sum_{i,j} v_x(i,j)^2 + v_y(i,j)^2}{d(i,j)}} \qquad (4)$$

to measure the motion information content, where w_m is for the m-th MB, (v_x, v_y) is the motion vector of a 8×8 block in a MB and $d(i, j)$ is the distance from current frame to its reference. In practice, (4) is normalized by the maximum w_m from (4) for all MBs. To obtain human perception uncertainty, the global motion should be calculated. For a natural video, a scene usually lasts several seconds, where an almost same background remains for all frames of the scene. For simplicity, only the relative motion is considered in this work for the purpose of computationally efficiency. To have a smooth motion filed of an image, which is actually preferred by the HVS, (4) is further quantized as

$$0.8.*(w_m>=1.5 \ \& \ w_m<=0.5)+1.0*(w_m<1.5 \ \& \ w_m>0.5), \qquad (5)$$

which is based on our experience. In addition, it indicates the HVS is not sensitive to very fast motion and very low motion. To obtain an integral VQM, both (1) and (4) are incorporated into the local MSE measure as the weighting factors. Then, a new VQM is proposed at MB, frame and sequence levels respectively as

$$VQM_m = \frac{w_m}{k_m} \times MSE_m, \tag{6}$$

and

$$VQM_f = \sum_{m=0}^{M-1} \frac{w_m}{k_m} \times MSE_m, \tag{7}$$

and

$$VQM = \sum_{f=0}^{L-1} \left(\sum_{m=0}^{M-1} \frac{w_m}{k_m} \times MSE_m \right). \tag{8}$$

The efficiency of the proposed metric is evaluated in the section of experimental results. The two most important correlation scores, i.e., the Linear Correlation Coefficient (LCC) and Spearman Rank Order Correlation Coefficient (SROCC), are 0.74 and 0.72 respectively on the LIVE Video Quality Database [24][25]. The performance is better than PSNR, SSIM, MSSIM, VSNR and VIF, and inferior to MOVIE [7][24].

III. THE PROPOSED VISUAL QUALITY CONTROL ALGORITHM

This paper provides a computationally efficient VQM which measures the visual quality of video compression on the one hand. On the other hand, to control visual quality during encoding process is mostly expected. The general video encodings are under the assumption that the channel bandwidth and buffer resources are limited, so the rate control functionality should also be provided.

Rate control should guarantee a small gap between the target bit rate and the actual coded bit rate, and achieve good R-D performance. Meanwhile, the consistent visual quality would be mostly expected, since it gives the spectators temporally consistent and thus comfortable visual experience. Additionally, the buffer constraint of output bitstream is compelling for the real-time video communications, such as live video encoding and video streaming. From [26], the R-D relation usually was modeled as a LOG function

$$R = \log_2 \frac{\sigma}{D}, \tag{9}$$

and the theoretical distortion model was given by

$$D = \frac{Q^2}{12}. \tag{10}$$

They were both based on the assumption of the Laplacian distribution of DCT coefficients. "σ" represents the standard deviation of input signal, which is replaced by Mean Absolute Difference (MAD) in practice. Eq. (10) was usually approximated by a linear or quadratic R-D mode [27][28]. The linear R-D model was written as

$$Q = \frac{a \times MAD}{R}, \tag{11}$$

where a is a model parameter, Q represents QP, and R indicates bits quote of a coding basis unit, e.g., a MB, a frame or a window [23].

A. Proposed VQM-based R-D model

The window model [23] proved that the best tradeoff between visual quality consistency and buffer constraint can be achieved in theory. To use window model, an R-D model should be provided at window-level, and then a corresponding window-level rate control can be operated at window-level R-D model. In this paper, the liner R-D model (10) is expanded to window-level as

$$Q_f = \frac{a \sum_{j=f}^{L-1} MAD_j}{T - \sum_{j=0}^{f-1} R_j}, \tag{12}$$

where Q_f is the QP of the f-th frame, MAD_f and R_f are the MAD and bits usage of the f-th frame respectively. T is the total bits allocated to a window. Eq. (12) can also be used at MB level rate control. From (12), QP is computed for each frame of the window. In addition, the remaining bits quotes and MADs are updated after encoding a frame. Such a method could provide a frame-level consistent visual quality profile. In addition, the bits control is accurate because (12) is implemented progressively. Furthermore, there is no bit allocation operation in (12), which usually results in the bad bits usage and bad visual quality as in the traditional rate control [27][28].

As shown in (10), the distortion is measured by MSE in video coding. Theoretically, all coding basic units should have the same MSE if a constant QP at frame level is employed according to (10). From Section II, a new VQM is defined in (6) to measure the visual quality of a MB. Importing (10) into (6) will generate a new theoretical model as

$$VQM_m = \frac{w_m}{k_m} \times Q_m^2. \tag{13}$$

With the parameters $\{w_m\}$ and $\{k_m\}$ for MBs or frames, the consistent visual quality at MB or frame level can be achieved. Eq. (13) actually provides a visual quality map $\{p_m\}$ for each coding units (MBs or frames). Firstly, (12) calculates a QP for the m-th coding unit. Secondly, the QP of each coding unit is revised accordingly by

$$\begin{cases} Q_m = Q_m \times p_m \\ p_m = \dfrac{w_m}{k_m} \times Q_f{}^2 \end{cases}, \tag{14}$$

where Q_f is a frame-level QP if (14) is used at MB level. p_m is adaptive to both video content and bit rate, so it is more suitable for video compression. If (14) is used at frame level, all the parameters in (14) are changed into the corresponding forms at frame level. Then, we can come up with a new form of (14) as

$$\begin{cases} Q_f = Q \times P_f \\ P_f = \sum_{m=0}^{M-1} p_m \end{cases}, \tag{15}$$

TABLE I
THE SYMBOLS USED IN THE FOLLOWING ALGORITHM

symbol	meaning
r	Bit rate
f	Frame rate
L	Window size
B	Buffer size
T	Target bits for a window
$\{P_f\}$, $\{p_m\}$	Visual quality map at frame level and MB level
Q_f, Q_m	QP of frame level and MB level for encoding
Q	Predicted QP for pre-analysis

TABLE II
THE PROPOSED VISUAL QUALITY CONTROL ALGORITHM

Step 1: Initializing window size $L=(B/r)\times f$;

Step 2: Getting the total bits of the current window $T=(r/f)\times L$;

Step 3: Performing pre-analysis of only 16x16 inter prediction for P and B frames and 16x16 intra prediction for I frames in current window;

Step 4: Computing the MAD for each frame in the current window;

Step 5: Computing edge strength and motion activity from (2) and (4) for each MB, and computing visual quality map $\{P_f\}$ and $\{p_m\}$ according to (14) and (15);

Step 6: Computing a QP from (12) for the j-th frame;

Step 7: Computing a new QP Q_f by using (15) at frame level as $Q_f = Q_f \times P_f$ for the purpose of consistent visual quality;

Step 8: Encoding the j-th frame using Q_f from Step 7; meanwhile, Q_f is revised by using (14) at MB level as $Q_m = Q_f \times p_m$;

Step 9: If the sequence ends, terminates procedure; else go to Step 1.

which revises frame-level QPs in the sense of consistent visual quality at frame level. Thus, (15) cooperated with (12) will construct the proposed VQM-based R-D model. In addition, the proposed R-D model can be used in both frame-level and window-level rate controls.

B. Proposed visual quality control algorithm

The proposed VQM-based R-D model, along with the proposed window-level rate control algorithm [23] could realize both the temporally and spatially consistent visual quality. Firstly, the QP of each frame is obtained by using [23]. Then, it is revised by (15) considering video content difference P_f in temporal domain. Furthermore, the QP of each MB can be further modified according to (14) considering the video content difference in spatial domain.

To employ window-level rate control algorithm [23], the window size should be firstly decided before encoding a window. In real-time video communication, the window size is updated dynamically along with the change of video content as discussed in [29]. And the change of window size would provide as smooth as possible visual quality for natural videos under the limited buffer and bit rate resources. To apply the proposed window model to the real rate control procedure, the pre-analysis process is needed to provide the MADs as shown in (12). Thus, the visual quality maps $\{p_m\}$ and $\{P_f\}$ in (14) and (15) respectively could be derived without extra computations. The window-level rate control algorithm can be used without considering video content variation for simplicity [30]. At such a situation, the window size is computed only from the given buffer resources. The proposed visual quality control algorithm is shortly summarized in Table II as follows. The symbols used in the proposed algorithm are listed in Table I.

IV. EXPERIMENTS AND DISCUSSIONS

In this section, we firstly evaluate the performance of the proposed VQM on a published video quality database, named LIVE Video Quality Database [25]. Secondly, the proposed visual quality control algorithm based on the proposed VQM-based R-D model is competed with the traditional rate control algorithm with respect to visual quality consistency and R-D efficiency. We implement the proposed algorithm on JM14.0 of H.264/AVC under the conditions: *Profile/Level: 100/40, Reference frames: 2, Full search,* Search range: *32, RDO on* and *CABAC, IPPP* encoding structure.

A. Performance evaluation of the proposed VQM on LIVE database

The performance of a VQM can be evaluated by depicting the relationship of the obtained VQM values and the provided subjective ratings, specifically the DMOS value of each distorted video. The DMOS value is obtained by subjective viewing tests where many observers participated and provided their opinions on the visual quality of each distorted video. Therefore, it can be regarded as the ground truth for evaluating the metric performances. As suggested by video quality experts group (VQEG) HDTV test [1] and that in [31], we follow their evaluation procedure to evaluate the performance of the proposed metric. Let x_j represent the visual quality index of the i-th distorted image obtained from the corresponding VQA. The five parameter $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$ monotonic logistic function is employed to map χ_j and V_j

$$V_j = \beta_1 \times (0.5 - 1/(1+e^{\beta_2\times(x_j-\beta_3)})) + \beta_4 \times x_j + \beta_5 . \qquad (16)$$

The corresponding five parameters are determined by minimizing the sum of squared differences between the mapped objecting score and the subjective DMOS value. Generally, three statistical measurements, LCC, SROCC and RMSE are employed to evaluate the performance of a VQM. LCC measures the prediction accuracy. SROCC provides an evaluation of the prediction monotonicity. The RMSE is introduced for evaluating the error during the fitting process. According to the definitions, larger values of LCC and SROCC mean that the objective and subjective scores correlate better, that is to say, a better performance of the VQM. And the smaller RMSE values indicate smaller errors between the two scores, therefore a better performance.

We evaluate the performance of the proposed metric on the LIVE database [25]. The LIVE database contains 150 distorted videos for ten uncompressed high-quality reference videos. There are 15 distorted videos for each reference video using four different distortion types, MPEG2 compression, H.264/AVC compression, simulated transmission of H.264 compressed bitstreams through error-prone IP networks and through error-prone wireless networks. Each video in the LIVE database was assessed by 38 human subjects, who scored the video quality on a continuous quality scale. The subjective ratings obtained from the subjective experiments, along with the reference and distorted videos, are provided in LIVE database.

The parameters, LCC, SROCC and RMSE are computed for measuring the correlation between the VQM values and subjective ratings. The computing results show that 0.74, 0.72 and 7.375 are for LCC, SROCC and RMSE respectively, which is better than PSRN, VSNR, VIF, MSSIM and SSIM metrics. It indicates that the proposed VQM can better assess the visual quality than those metrics which are without temporal information. We also compare the performances of the proposed VQM, PSNR, SSIM [14][15], MSSIM [16], VSNR [32], VIF [33], and MOVIE [24] . The computing results of LCC and SROCC results are illustrated in Table III and Table IV. The RMSEs of PSNR, SSIM, MSSIM, VSNR and VIF are 9.188, 8.267, 7.717, 9.777, 7.860 respectively, which are all larger than ours. As PSNR, SSIM, MSSIM, VSNR and VIF only provide frame-level quality scores, the final quality index of the video sequence is generated by averaging their outputs of each frame.

From Table III, it can be observed that PSNR performs poorly, because it is not related to the HVS perception. Also the VSNR performs badly, which can be attributed to two reasons. The first is that VSNR analyzes the HVS perception of the distortion in the wavelet domain. But the MPEG-2 and H.264 compression schemes introduce the distortions during the quantization process in DCT domain. The second one is that VSNR is an image quality metric designed to capture the spatial distortions. For video quality assessment, the temporal information is very important and needs to be accounted for. This is also the reason why SSIM, MSSIM and VIF perform

TABLE III: LINEAR CORRELATION COEFFICIENT

Algorithm	Wireless	IP	H.264	MPEG-2	All Data
PSNR	0.677	0.478	0.589	0.409	0.569
SSIM	0.473	0.537	0.611	0.582	0.503
MSSIM	0.684	0.684	0.692	0.632	0.676
VSNR	0.680	0.737	0.614	0.507	0.688
VIF	0.593	0.636	0.649	0.673	0.577
MOVIE	0.855	0.798	0.853	0.806	**0.829**
Proposed	0.762	0.736	0.709	0.556	**0.741**

TABLE IV: SPEARMAN RANK ORDER CORRELATION COEFFICIENT

Algorithm	Wireless	IP	H.264	MPEG-2	All Data
PSNR	0.671	0.430	0.477	0.394	0.553
SSIM	0.539	0.474	0.659	0.569	0.533
MSSIM	0.729	0.645	0.734	0.681	**0.735**
VSNR	0.694	0.693	0.641	0.587	0.672
VIF	0.538	0.553	0.638	0.635	0.558
MOVIE	0.834	0.735	0.837	0.739	**0.804**
Proposed	0.753	0.724	0.664	0.564	**0.721**

successfully in image quality evaluation, but not so well on the video quality assessment. From Table III, it can be observed that the performances of these metrics are not good enough, with SROCC values smaller than 0.6. The reason is that the temporal information is not accurately modeled. For video quality assessment, the temporal distortion is very important and needs to be considered for developing an effective video quality metric. Our proposed method outperforms PSNR, VSNR, SSIM, MSSIM and VIF. It means that the proposed metric can effectively depict the perceptual quality of the distorted videos. The scatter-plots of different VQMs over the LIVE video quality database are illustrated in Fig. 1. It can be observed that for our proposed method the sample points scatter more closely around the fitted line. It means that the values predicted by the proposed method correlate better with the subjective ratings, specifically the DMOS values, demonstrating a better performance.

B. *Performance of the proposed visual quality control algorithm*

The proposed visual quality control algorithm is performed on the conventional CBR encoding with limited bandwidth and buffer resources. The traditional rate control (JVT-H017r3) [28] is a benchmark employing PSNR measurement. The comparisons between the proposed visual quality control and [28] are summarized in Table V with

TABLE V: PERFORMANCE COMPARISON FOR ALL TESTING ALGORITHMS IN TERMS OF BIT CONTROL ERROR AND PSNR/VQM GAIN

Sequence	Target bit rate (kbps)	Traditional rate control				Proposed rate control			
		Bit rate (kbps)	PSNR	VQM	Error (%)	Bit rate (kbps)	PSNR	VQM	Error (%)
Foreman	*2000*	2002.60	42.25	0.981	0.09	1996.98	42.18	0.985	-0.15
	1000	1001.76	39.31	0.966	0.44	1000.56	39.31	0.971	0.06
	500	502.52	36.56	0.941	0.35	500.41	36.62	0.949	0.08
	300	302.88	34.48	0.916	1.03	300.97	34.7	0.925	0.32
News	*2000*	2002.25	47.48	0.993	0.09	1998.33	48.1	0.997	-0.08
	1000	1002.33	44.81	0.988	0.22	999.91	44.94	0.991	-0.01
	500	502.10	41.76	0.976	0.33	499.9	41.9	0.982	-0.02
	300	301.14	39.22	0.968	0.37	300.17	39.49	0.973	0.06
Silent	*2000*	2003.86	45.30	0.977	0.09	2000.39	45.92	0.990	0.02
	1000	1003.59	41.77	0.952	0.35	999.78	42.2	0.966	-0.02
	500	501.97	38.50	0.891	0.5	499.72	38.46	0.912	-0.06
	300	301.16	36.15	0.849	0.53	299.97	35.96	0.865	-0.01
Tennis	*2000*	2000.99	41.79	0.957	0.09	1999.83	42.16	0.983	-0.08
	1000	1002.09	38.31	0.941	0.22	1001.03	38.72	0.967	-0.01
	500	501.76	35.11	0.879	0.33	501.32	35.4	0.918	-0.02
	300	301.46	32.80	0.831	0.37	299.79	33.31	0.875	0.06
Average			**39.73**	**0.937**	**0.33**		**39.96**	**0.953**	**0.03**
Night	*10000*	10006.65	39.19	0.974	0.07	9992.06	39.29	0.980	-0.08
	8000	8006.51	38.45	0.969	0.08	7994.46	38.61	0.975	-0.07
	5000	5007.32	36.93	0.955	0.15	4995.81	37.2	0.961	-0.08
	2000	2004.89	33.49	0.908	0.24	2000.69	34.01	0.913	0.03
Crew	*10000*	10010.06	41.53	0.958	0.1	9996.95	41.39	0.962	-0.03
	8000	8010.63	40.94	0.954	0.13	7995.43	40.84	0.958	-0.06
	5000	5007.76	39.74	0.944	0.16	4996.46	39.77	0.948	-0.07
	2000	2004.81	36.97	0.915	0.24	1994.48	37.3	0.917	-0.28
Harbour	*10000*	10005.67	37.41	0.965	0.06	9997.14	37.41	0.970	-0.03
	8000	8005.80	36.52	0.961	0.07	7997.97	36.53	0.966	-0.03
	5000	5005.04	34.68	0.950	0.1	4999.71	34.7	0.955	-0.01
	2000	2004.08	31.21	0.913	0.2	1996.56	31.28	0.917	-0.17
Average			**37.26**	**0.947**	**0.13**		**37.36**	**0.952**	**-0.07**

respect to both visual quality improvement and bit control accuracy. The visual quality is both measured by PSNR and the proposed VQM. From Table V, the proposed algorithm achieves a significant PSNR and VQM improvement than the benchmark. The PSNR improvement of the proposed algorithm is up to 0.23dB over the benchmark on CIF sequences. Table V also gives the visual quality comparison between the proposed algorithm and [28] for each sequence at each bit rate, where the same observation as PSNR can be concluded. In Table V, the visual quality values are normalized from 0 to 1. From Table V, the bit rate mismatch of the proposed algorithm is below 0.32%. It indicates that the rate control of conventional CBR encoding can also be fulfilled by the proposed visual quality control algorithm.

V. CONCLUSIONS

In this paper, a new VQM is firstly proposed for the visual quality control of the conventional video encodings which are usually under the constraints of channel bandwidth and buffer resource. The proposed VQM is computationally efficient, and highly related to quantization operation of video compression. Secondly, a VQM-based rate control algorithm is proposed to control visual quality consistency during encoding process. The further investigations will be conducted on the subjective assessment of the proposed algorithm, consistent visual quality assessment and comprehensive analyses of the advantage of the proposed algorithm.

References

[1] "Final report from the video quality experts group on the validation of objective models of video quality assessment," Tech. Rep., Video Quality Expert Group, March 2000, Available [Online]: http://www.vqeg.org.

[2] P. Corriveau and A. Webster, "Final report from the video quality experts group on the validation of objective models of video quality assessment, phase II," Tech. Rep., Video Quality Expert Group, July 2003.

[3] H. R. Sheikh, Z. Wang, L. Cormack, and A.C. Bovik, "Live image quality assessment database release 2," Available online: http://live.ece.utexas.edu/research/quality.

[4] N. Ponomarenko, M. Carli, V. Lukin, K. Egiazarian, J. Astola, and F. Battisti, "Color image database for evaluation of image quality metrics," in Proceedings of the IEEE Workshop on Multimedia Signal Processing, Cairns, AU, October 2008.

[5] F. De Simone, M. Naccari, M. Tagliasacchi, F. Dufaux, S. Tubaro, and T. Ebrahimi, "Subjective assessment of H.264/AVC video streaming with packet losses," Eurasip Journal on Image and Video Processing,

[6] Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it? - A new look at signal fidelity measures," IEEE Signal Processing Magazine, vol. 26, no. 1, pp. 98-117, Jan. 2009.

[7] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," IEEE Transactions on Image Processing, vol. 19, no. 2, pp. 335–350, Feb. 2010.

[8] C. van den Branden Lambrecht and O. Verscheure, "Perceptual quality measure using a spatio-temporal model of the human visual system," in Proc. Int. Soc. Opt. Eng. (SPIE), 1996, vol. 2668, pp. 450–461.

[9] S. Winkler. A perceptual distortion metric for digital color video. In SPIE Proceedings of Human Vision and Electronic Imaging, volume 3644, San Jose, CA, January 1999.

[10] J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," in Digital Images and Human Vision, A. B. Watson, Ed. The MIT Press, 1993, pp. 163–178.

[11] A. Watson, J. Hu, and J. McGowan, "Digital video quality metric based on human vision," J. Electron. Imaging, vol. 10, no. 1, pp. 20–29, Jan. 2001.

[12] M. Masry, S. S. Hemami, and Y. Sermadevi, "A scalable wavelet-based video distortion metric and applications," IEEE Trans. Circuits Syst. Video Technol., vol. 16, no. 2, pp. 260–273, 2006.

[13] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Considering temporal variations of spatial visual distortions in video quality assessment," IEEE J. Sel. Topics Signal Process., vol. 3, no. 2, pp. 253–265, 2009.

[14] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," Signal Proc.: Image Comm., vol. 19, pp. 121-132, 2004.

[15] Z. Wang, H. R. Sheikh, A. C. Bovik and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, pp. 600-612, 2004.

[16] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," Invited Paper, IEEE Asilomar Conference on Signals, Systems and Computers, Nov. 2003.

[17] Z. Wang and Q. Li, "Video quality assessment using a statistical model of human visual speed perception," Journal of the Optical Society of America A -Optics Image Science and Vision, vol. 24, no. 12, pp. B61-B69, Dec. 2007.

[18] K. Seshadrinathan and A. C. Bovik, "A structural similarity metric for video based on motion models," IEEE Inter. Conf. Acoustics, Speech, and Signal Processing, April, 2007.

[19] K. Seshadrinathan and A. C. Bovik, "An information theoretic video quality metric based on motion models," Third Inter. Workshop on Video Proc. and Quality Metrics for Consumer Electronics, Jan , 2007.

[20] A. K. Moorthy and A. C. Bovik, "Efficient video quality assessment along temporal trajectories," IEEE Transactions on Circuit system for Video Technology, vol.20, no.11, pp.1653-1753, Nov. 2010.

[21] A. Bhat, S. Kannangara, Y. F Zhao and I. Richardson, "A full reference quality metric for compressed video based on mean squared error and video content," IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 2, pp. 165-173, Feb. 2012.

[22] Y.-F. Ou, Z. Ma, and Y. Wang, "Perceptual quality assessment of video considering both frame rate and quantization artifacts," IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, no. 3, pp. 286-298, March 2010.

[23] L. Xu, D.B. Zhao, X. Y. Ji, L. Deng, S. Kwong, W. Gao, "Window level rate control for smooth visual quality and smooth buffer occupancy," IEEE Transactions on Imaging Processing, vol. 20, no. 3, pp. 723-734, August 2011.

[24] K. Seshadrinathan, R. Soundararajan, A. Bovik and L.Cormack, "Study of Subjective and Objective Quality Assessment of Video," IEEE Transactions on Image Processing, vol.19, no.6, pp.1427-1441, June 2010.

[25] LIVE Video Quality Database. Available [Online]: http://live.ece.utexas.edu/research/quality/live_video.html.

[26] L. Xu, X.Y. Ji, W. Gao, D.B. Zhao, "Laplacian Distortion Model (LDM) for Rate Control in Video Coding," PCM 2007, pp.638-646, Dec. 2007.

[27] Z. G. Li, W. Gao, F. Pan, S. W. Ma, K. P. Lim, G. N. Feng, X. Lin, S. Rahardja, H. Q. Lu and Y. Lu, "Adaptive Rate Control for H.264", Journal of Visual Communication and Image Representation, vol. 17, no. 2, pp. 376-406, April, 2006.

[28] S.W. Ma, Z.G. Li, F. Wu, "Proposed Draft Adaptive Rate Control", Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, Doc. JVT-H017r3, 8th Meeting, Geneva, 20-26 May, 2003.

[29] L. Xu, S. Kwong, D. B Zhao and Y. Zhang, "A Encoder Framework for Window-level Rate Control Optimization," IEEE Transactions on Industiral Electronics, (Accepted), 2012.

[30] L. Xu, S. Kwong, H. L. Wang, D. B. Zhao, Y. Zhang, W. Gao, "A Universal Rate Control Scheme for Video Transcoding," IEEE

Transactions on Circuits Syst.Video Technol. vol. 22, no. 4, pp. 489-501, April 2012.

[31] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms”, *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440-3451, Nov. 2006.

[32] D. M. Chandler, and S. S. Hemami, “VSNR: a wavelet-based visual signal-to-noise ratio for natural images”, IEEE Trans. Image Process., vol. 16, no. 9, pp. 2284-2298, 2007.

[33] H. R. Sheikh, and A. C. Bovik, “Image Information and Visual Quality”, IEEE Trans. Image Process., vol: 15 no. 2, pp. 430-444, Feb. 2006.

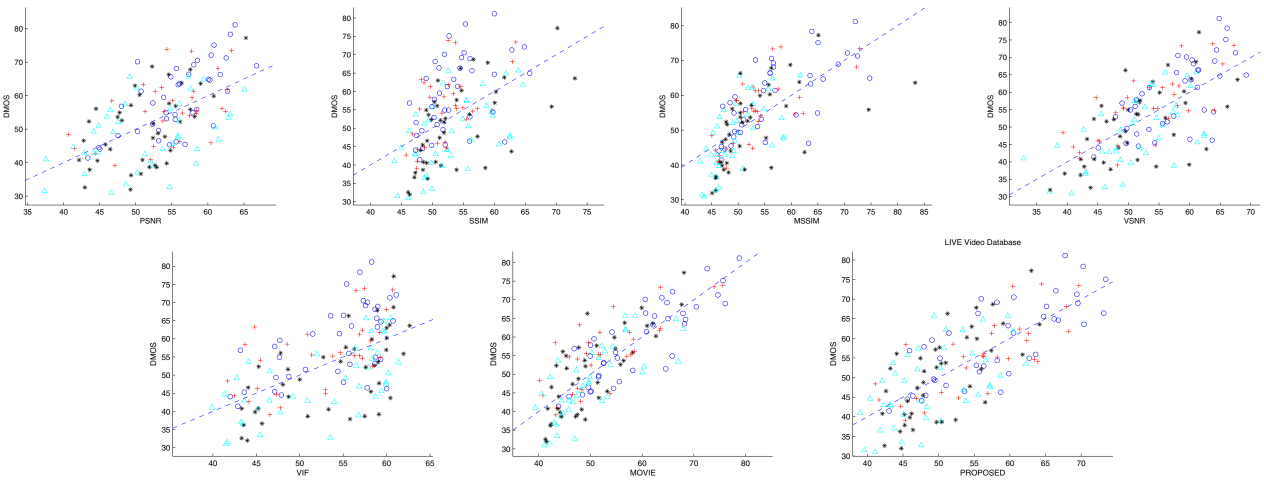

Fig. 1. Scatter plots of the DMOS values versus model predictions on the LIVE video quality database. Each sample point represents one test video. (The circle and “+” indicate the distorted video sequences under wireless network and IP network respectively. The star indicates H.264 encoded video sequence, while the triangle indicates the MPEG-2 compressed one.) First row from left to right: PSNR, SSIM, MSSIM, and VSNR; second row from left to right: VIF, MOVIE and the proposed method.