A Multi-Objective Optimization Perspective for Joint Consideration of Video Coding Quality

Wei Gao*†
*School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, Shenzhen, China
†Peng Cheng Laboratory, Shenzhen, China
E-mail: gaowei262@126.com

Abstract—**Traditional efforts for video coding usually focus on either rate-distortion (R-D) performance or quality smoothness (QS), which may not effectively achieve the desirable visual quality of experience for reconstructed videos. Single-objective optimization cannot guarantee the performance of the other one. In this paper, we would like to introduce a new perspective for the better joint consideration of video coding quality, where both R-D and QS are simultaneously evaluated, and then the traditional video coding optimization problem is converted from single objective to multiple objectives. We provide the details to demonstrate the establishment method of the joint video coding quality (JVCQ), which is given as the comprehensive analytical model to jointly considers spatial and temporal artifacts. The FixedQP results and consumed bits are referred to construct the comparable JVCQ results. Finally, by collecting JVCQ results for four different target bit rates and five different weighting strategies, we can obtain the rate-averaged JVCQ (RJVCQ), weighting-averaged JVCQ (WJVCQ) and rate-weighting-averaged JVCQ (RWJVCQ), respectively. By comparing the different rate control (RC) algorithms, experiments validate the consistency and applicability of the proposed JVCQ with the original two separate evaluation metrics, and the proposed RC algorithms can have better performances. Besides the video coding quality evaluation, the group of JVCQs can be used to guide the optimization process of video coding to achieve gains on both evaluation metrics simultaneously.**

I. Introduction

The developments of video coding standards, including H.264/AVC [1] and H.265/HEVC [2], are promoting the vast popularity of diversified video applications into our daily life. More and more video data are generating increased heavy overloads to the multimedia information systems [3]. Therefore, the reservation and delivery tasks for these huge videos still puzzle researchers and engineers on many sides. One significant problem is the enhancement of perceived visual experience with the given communication bandwidth constraint. Accompanied by the rapidly growing video applications, the vast optimization endeavors for video encoders will constantly play an indispensable role in improving the overall systematic performances for the multimedia information processing applications.

For high efficiency video coding (H.265/HEVC) [4], lots of optimization efforts have been extensively carried out with goals of optimizing either rate-distortion (R-D) performance [5]-[9] or quality smoothness [10]-[11] by allocating coding bits and quantization parameters (QPs) in rate control (RC). Unified rate-quantization (URQ) model [5] and R-λ model [6] based RC algorithms were proposed to exploit coding efficiency, while other algorithms used statistical modeling and quality dependency [7], game theoretical bit allocation [8]-[9], machine learning based R-D model prediction [9] and coding parameter prediction [22], etc. Moreover, for the consistent quality oriented video coding optimization, the advanced models and reference distortions were employed [10]-[11]. These schemes all focused on the single objective optimization. However, the R-D performance and quality smoothness (QS) are both the key factors to determine the perceived quality in rate-constrained video coding.

Single objective optimization was the mainstream manner in existing optimization endeavors. Nevertheless, for better visual quality optimization, a more comprehensive metric is desired urgently for coding quality evaluation. Traditionally, mean squared error (MSE) based peak signal noise ratio (PSNR) and structural similarity (SSIM) [8], [12] were adopted to measure signal and textural loss, although many researchers have devised various metrics for video quality assessment (VQA) [13]-[15] to better reflect the visual perception on distorted content. By introducing the data-driven and machine learning strategies, VQA metrics are owning the increased computation overloads, which actually cannot been applied into the video encoder optimization. There are two major reasons. The first is that the developed quality metrics can not be widely recognized due to limited tested samples. The second is that the computation overloads were too unacceptable in many practical systems, and the achieved benefits were undeserving of large computation investment. Thus, the video coding standardization still uses PSNR, and the research community may sometimes use low complexity SSIM to conduct the optimization efforts. Thus, we can see that the quality evaluation metric with much lower computation complexity is favored in video coding.

The rate-constrained coding quality evaluation is different from simple VQA procedure. We introduce use the simple calculation manner to evaluate and compare video coding quality, i.e. using PSNR and SSIM. Instead of the single objective optimization, we propose to adopt the multiple objective optimization. The new joint video coding quality (JVCQ) model is introduced to evaluate the video coding performance from the perspective of rate-constrained quality optimization. First, we can decompose the influential factors determining the human visual experience into two major factors, including the rate-constrained spatial quality, i.e. the R-D performance, and the temporal fluctuations. Second, we utilize the FixedQP results to make these two components comparable and combined to generate a comprehensive model. Third, we introduce the mean JVCQ over different rates (RJVCQ), the mean JVCQ over different weighting strategies (WJVCQ), and the mean JVCQ over different bit rates and weighting strategies (RWJVCQ) to generate comprehensive evaluation results for different coding schemes for the fairness on different bandwidths and practical requirements.

The remainder of this paper is organized as follows. In Section II, the motivations of evaluating spatial and temporal quality in video coding are discussed, including the rate-constrained coding optimization frameworks. In Section III, the proposed JVCQ metrics are presented according to the perspective of multiple objective optimization. In Section IV, experiments are conducted to validate the consistency of the proposed JVCQ metrics for the different state-of-the-art RC algorithms. Finally, the conclusions are drawn in Section V.

II. MOTIVATION

To illustrate the motivation of the proposed JVCQ model for the joint considerations to evaluate video coding quality, in the following paragraphs we will discuss visual degradation effects from both spatial and temporal dimensions, and provide a novel multiple objective optimization perspective in the context of rate-constrained video coding optimization.

A. *Visual Degradation Effect from Spatial Distortion*

Irrespective of intra or inter frame and block level coding, the objectives of optimizing them are to reduce the signal loss during video encoding process, particularly in residual quantization after prediction and transformation [7]. Thus, the main goal of improving video coding is the elimination of average signal-level distortions which can be deemed as the average frame-level spatial coding distortions.

In fact, many existing endeavors focused on reducing the average spatial distortions with rate constraint, which is known as rate-distortion optimization (RDO) [16]. Similarly, when the average spatial distortions are fixed, video coding optimization tries to minimize consumed bandwidth. RDO process is to minimize the following RD cost function J,

$$J = D + \lambda \cdot R, \tag{1}$$

where λ is the Lagrange multiplier relating to the specific quantization parameter (QP) for each R-D pair. Hence, the formulation for minimizing average spatial distortions with rate constraints can be expressed as

$$\min_{\{QP_i\}} \sum_{i=1}^{N} D_i, \ s.t. \sum_{i=1}^{N} R_i \le R_T, \tag{2}$$

where R_i and D_i are consumed bits and distortion of the i-th coding frame or coding tree unit (CTU), i=1, 2, ..., N, and the total amount of bits should be less than the target R_T.

Undoubtedly, with a fixed rate constraint, less spatial coding distortions can bring better visual experience. We denote the visual degradation effect as VDE and the average spatial distortion as SD, while other factors that influence VDE are denoted as OF_1, then

$$VDE = f_1(SD, OF_1), \tag{3}$$

where f_1 depicts the relationship between VDE and $\{SD, OF_1\}$. Since average spatial quality is the major optimization goal, thus minimizing VDE is equivalent to minimizing SD,

$$\min_{\{QP_i\}}\{VDE\} \Leftrightarrow \min_{\{QP_i\}}\{SD\}, s.t. \sum_{i=1}^{N} R_i \le R_T, \tag{4}$$

where a list of frame-level QPs ($\{QP_i$, i=1, 2, ..., $N\}$) are obtained with the criterion of optimizing the R-D objective.

B. *Visual Degradation Effect from Temporal Distortion*

Visual experience is influenced by not only average spatial distortions, but also temporal variations of distortions. Some researchers have made efforts [10]-[11] to obtain the frame-level quality smoothness by optimizing the RC process. Therefore, it is widely recognized that the perceived visual experience can be significantly influenced by the variations of frame-level distortions, namely temporal distortions. The minimization of frame-level fluctuations with constraint of bit rates can be formulated as

$$\min_{\{QP_i\}} F(\{D_i, i = 1, 2, \ ..., N\}), \ s.t. \sum_{i=1}^{N} R_i \le R_T, \tag{5}$$

where the total amount of frames is N, and F is to measure the fluctuations of frame-level quality, which can be defined as the standard variance of frame-level distortions.

Similarly, we denote the temporal distortion as TD, while other factors that influence VDE are denoted as OF_2, then

$$VDE = f_2(TD, OF_2), \tag{6}$$

where f_2 depicts the relationship between VDE and $\{TD, OF_2\}$. Approximately, temporal smooth quality is the major optimization goal, therefore minimizing VDE is equivalent to minimizing TD,

$$\min_{\{QP_i\}}\{VDE\} \Leftrightarrow \min_{\{QP_i\}}\{TD\}, s.t. \sum_{i=1}^{N} R_i \le R_T, \tag{7}$$

where a list of frame-level QPs ($\{QP_i$, i=1, 2, ..., $N\}$) are obtained with the criterion of optimizing the frame-level quality smoothness objective.

C. *From Rate-Constrained Spatial and Temporal Quality Analyses to Multi-Objective Optimization*

As aforementioned, we can see that with certain bandwidth constraint, the spatial and temporal quality can be improved separately by using optimization techniques. Nevertheless, the separate single objective optimization cannot guarantee that the two influential aspects can be enhanced together. In fact, almost all existing optimization efforts in video coding focused on either spatial quality or quality smoothness enhancements, and no works have noticed the importance of jointly optimizing both of them to achieve better visual experience. After taking both spatial and temporal factors, the formulation of VDE can be updated as

$$VDE = f_3(SD, TD, OF_3), \tag{8}$$

where f_3 depicts the relationship between VDE and $\{SD, TD, OF_3\}$, OF_3 means the set of other factors. Since SD and TD play the most key roles in evaluating visual experience, we can neglect the other implicit factors to have the clear and simple relationship modeling,

$$VDE = f(SD, TD) = W \cdot SD + (1 - W) \cdot TD, 0 \le W \le 1, \tag{9}$$

where W is a tradeoff parameter between the spatial and temporal distortion evaluations, while the function f models the relationship between VDE and $\{SD, TD\}$. Hence, the traditional separate single objective optimization for R-D performance or quality smoothness can be reformulated as multiple objective optimization. Based on this formulation, the spatial and temporal quality can be jointly optimized.

Therefore, the minimization of VDE is equivalent to the joint minimization of SD and TD by implementing the multiple objective optimization,

$$\min_{\{QP_i\}}\{VDE\} \Leftrightarrow \min_{\{QP_i\}}\{W \cdot SD + (1-W)\cdot TD, 0 \le W \le 1\}, s.t. \sum_{i=1}^{N} R_i \le R_T \ , \tag{10}$$

where a list of frame-level QPs ($\{QP_i$, i=1, 2, ..., $N\}$) are obtained with the criterion of optimizing the joint rate-constrained multiple objectives.

III. Proposed Joint Video Coding Quality (JVCQ) Evaluation Metric

A. Normalization

Assume there are N frames, the collectable coding results after video encoding include frame-level bits $\{R_1, R_2, ..., R_N\}$ and generated distortions $\{D_1, D_2, ..., D_N\}$, then from the average bits, frame rate and distortions, we can calculate the average R and D to get the R-D performance evaluation for a particular coding scheme. Because different target bit rates will cause different bits and distortions for RC algorithms, the JVCQ results should be normalized to be comparable. Therefore, we propose to use the rate and distortion results of FixedQP as normalizers to JCVQ results. By considering rate-constrained video coding practice, the adopted strategy can make the optimization reasonable for visual experience enhancement and the obtained JVCQ evaluation results can be comparable among different coding schemes and target bandwidth conditions.

First, we propose to use the following SQR to evaluate the spatial quality with consideration on consumed bit rates,

$$SQR = \frac{D_F}{D_T}\cdot\frac{R_F}{R_T}\ , \tag{11}$$

$$SQR = \frac{Q_T}{Q_F}\cdot\frac{R_F}{R_T}\ , \tag{12}$$

where $\{R_F, D_F, Q_F\}$ and $\{R_T, D_T, Q_T\}$ are rate, distortion and quality of FixedQP and tested methods, respectively. The first items in (11) and (12) indicate the spatial quality evaluations, while the second items in (11) and (12) consider the rate-constrained conditions to achieve the related quality. The distortions can be evaluated by MSE or NSSIM=1-SSIM, and quality can be evaluated by PSNR or SSIM. SQR definitions in (11) and (12) can be named as distortion based and quality based SQR, respectively.

Second, for temporal quality, we use the standard variance of frame-level distortions or quality to evaluate the temporal influence on visual experience. For normalization, we can use the standard variance results of FixedQP which provides almost the smoothest coding results. The temporal quality is proposed to be measured by TQS,

$$TQS = \frac{STV(\{D_{F,i}, i=1,2,\ldots,N\})}{STV(\{D_{T,i}, i=1,2,\ldots,N\})}, \tag{13}$$

$$TQS = \frac{STV(\{Q_{F,i}, i=1,2,\ldots,N\})}{STV(\{Q_{T,i}, i=1,2,\ldots,N\})}, \tag{14}$$

where STV is the operator to obtain standard variance for a list of frame-level coding distortions and quality, $D_{F,i}$ and $D_{T,i}$ are coding distortions of the i-th frame by FixedQP and tested methods, respectively, while $Q_{F,i}$ and $Q_{T,i}$ are corresponding coding quality. (13) and (14) are distortion based and quality based TQS, respectively. Larger TQS indicates better quality smoothness.

B. JVCQ Metric for Comprehensive Evaluation

After defining SQR and TQS, we could propose the JVCQ metric to comprehensively evaluate the video coding quality by considering both spatial and temporal quality,

$$JVCQ = W \cdot SQR + (1\text{-}W)\cdot TQS, 0 \le W \le 1, \tag{15}$$

where SQR and TQS are normalized by FixedQP results, respectively. For calculation procedures based on distortions and quality, we can obtain the distortion based and quality based JVCQ evaluations, respectively, where the two items should be aligned with same evaluation approach, i.e. either distortions or quality, for uniformity. For combination, it is important to make these two items comparable. Obviously, the coding schemes with higher JVCQ scores are preferred. When FixedQP is tested, JVCQ is equal to 1.

Similar with BD-PSNR and BD-BR [17]-[18], four target bit rates $\{BR_1, BR_2, BR_3, BR_4\}$ obtained by FixedQP testings with fixed frame-level QPs {22, 27, 32, 37} are tested to fully compare different coding schemes. The mean JVCQ over different bit rates (RJVCQ) can be calculated as

$$RJVCQ = \frac{1}{I}\sum_{i=1}^{I} JVCQ_{BR_i}\ , \tag{16}$$

where I can be 4 to indicate the number of tested target bit rates or fixed QPs, $JVCQ_{BRi}$ is the JVCQ result with BR_i and a fixed weighting strategy for SQR and TQS. Additionally, to comprehensively weight the importance of spatial and temporal quality in visual experience evaluations, parameter W can be tested with different configurations ($0 \le W \le 1$).

For a fixed target BR_i or QP, if different weighting strategies are applied, the same coding algorithm will also generate different JVCQ results, thus we propose to use the mean JVCQ over different weighting strategies (WJVCQ) to evaluate the robustness of different algorithms to differently weighted JVCQ metrics, and WJVCQ is calculated as

$$WJVCQ = \frac{1}{J}\sum_{j=1}^{J} JVCQ_{W_j}\ , \tag{17}$$

where J can be 5 to indicate the number of different weighting strategies. Typical values of W can be {0.00, 0.25, 0.50, 0.75, 1.00}, which means: (1) all optimization efforts are put into temporal quality improvements; (2) spatial and temporal quality are jointly optimized with more efforts on temporal quality; (3) spatial and temporal quality are jointly optimized with equal efforts; (4) spatial and temporal quality are jointly optimized with more efforts on spatial quality; (5) all optimization efforts are put into spatial quality improvements, respectively.

It can be seen that RJVCQ and WJVCQ can cover different bandwidths and different weighting strategies for different requirements. Then, we can further propose the mean JVCQ over different rates and weighting strategies (RWJVCQ) which be expressed as the following equation,

$$RWJVCQ = \frac{1}{I \times J}\sum_{i=1}^{I}\sum_{j=1}^{J} JVCQ_{BR_i,W_j}, \quad (18)$$

which can comprehensively compare different video coding schemes. Obviously, the JVCQ improvements can indicate the advantages of the tested coding scheme than the other. We advocate the use of the proposed RWJVCQ results for complete comparisons among different coding schemes with both different target bit rates and weighting strategies in diverse video content and application requirements. Similar with R-D curves, RJVCQ-W curves and WJVCQ-R curves can be easily plotted to compare different coding schemes.

IV. Experimental Results

To demonstrate the consistency and applicability of the JVCQ model, we would like to conduct experiments to compare different RC schemes by using JVCQ analytics, and illustrate its usage in evaluating video coding quality.

In [18], the latest HEVC reference software HM-16.19 adopted the R-λ model based RC algorithm [6] which derived the λ for QP determination [19]-[20]. In [10], the adjacent reference frame-level distortions were used to set target distortions, and then QP was determined for consistent quality by using ρ-domain RC optimization [21]. In [11], similar with the R-λ model [6], the D-λ model was used for the consistent quality oriented QP determination. In [9], we have proposed the joint machine learning and game theory (MLGT) based RC optimization method for R-D optimization. In [22], we have proposed the learning-based initial QP (LIQP) method to also enhance the R-D performance. In this paper, we only test LIQP using MSE based RD optimum and present the NSSIM distortion and SSIM quality JVCQ results. Due to the monotonic relationship of optimal initial QP and target bit rates in [22], intra frame QP increment is beneficial to or give optimization priority to the relatively low bandwidth coding, while inter frame bit ratios were adjusted for smooth frame-level quality. Therefore, to test the nature of the joint MLGT framework, in the experimental comparisons we remove these two additional parts. Therefore, the MLGT and LIQP methods gave the inter and intra frame RC optimization, respectively.

These different coding schemes were tested on different previous versions of HM reference software, and may use different coding parameters and bandwidth constraints. In this paper, for fairness, we will implement all of them on the HM-16.19 using the target bit rates generated by FixedQP using frame-level QPs {22, 27, 32, 37} to maintain the same test conditions. Since the relative performance differences (on the R-D performance and quality smoothness) of these schemes have been given in the previous works [18], [10]-[11], [9] and [22], for simplicity we will not list these coding results redundantly, and these results can be used as the reference for the JVCQ comparisons. The video coding schemes in [18], [10], [11], [9] and [22] are denoted as "HM-16.19", "TIP13-Seo", "TIP16-Wang", "TIP17-Gao" and "TBC19-Gao", respectively. Additionally, the FixedQP method using fixed frame-level QPs is also tested to calculate JVCQ.

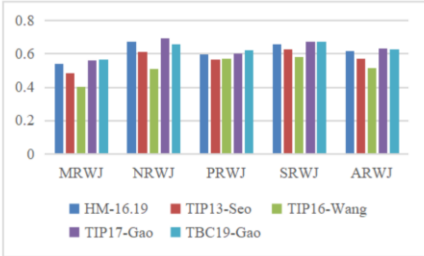

Fig. 1. Comparisons of different types of JVCQ scores for different video coding algorithms

From Table I and Fig. 1, the MSE-based RWJVCQ (MRWJ), NSSIM-based RWJVCQ (NRWJ), PSNR-based RWJVCQ (PRWJ), SSIM-based RWJVCQ (SRWJ), and the average RWJVCQ (ARWJ) are all listed and compared, where our proposed TIP17-Gao [9] and TBC19-Gao [22] methods can outperform the other methods significantly. These results are generally consistent with the tested results for R-D performance and quality smoothness in [9] and [22]. The intrinsic reason is that JVCQ only introduces the linear combination and weighting strategies for evaluations, which does not destroy the general consistency of performances.

In Table II and III, the RJVCQ with different weighting strategies and the WJVCQ with different target bit rates are compared, respectively, from which we can also see that the MLGT and LIQP proposed methods can be better than the other methods. Fig. 2 provides the RJVCQ-W and WJVCQ-R curves where we can also find the advantages of the proposed schemes, as well as the consistency and applicability of the introduced JVCQ models in evaluating video coding quality.

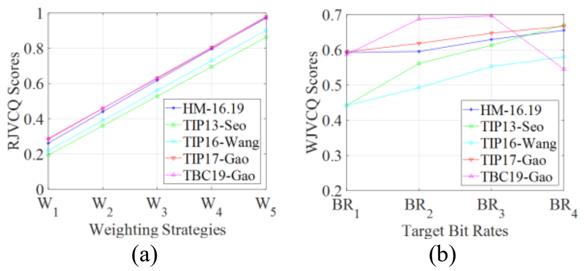

Fig. 2. Comparisons of different video coding algorithms on: (a) RJVCQ-W and (b) WJVCQ-R curves

By referring to the R-D performance and the quality smoothness results of the compared methods, we can see that RWJVCQ can consider both spatial and temporal influences on visual experience, and also consider both different bit rates and different weighting strategies required in practice. Thus, the introduced JVCQ models can be capable to provide the comprehensive comparisons. Another distinct feature is that the JVCQ models are easy to use and fast for computations, which can benefit its wide adoptions. The two items in the JVCQ are originally not comparable, but after normalization, the scores become not monotonically related bit rates, which can be deemed as another contribution of this paper. Thus, it should be noted that the JVCQs are not used for video quality assessment, but mainly for the evaluations and comparisons of rate-constrained coding schemes, as well as for guiding the coding optimization process for the achievement of gains on R-D performance, quality smoothness and visual experience.

Table I
Comparisons of RWJVCQ scores for different video coding algorithms

Class	Sequence	HM-16.19					TIP13-Seo					TIP16-Wang					TIP17-Gao					TBC19-Gao				
		MRWJ	NRWJ	PRWJ	SRWJ	ARWJ	MRWJ	NRWJ	PRWJ	SRWJ	ARWJ	MRWJ	NRWJ	PRWJ	SRWJ	ARWJ	MRWJ	NRWJ	PRWJ	SRWJ	ARWJ	MRWJ	NRWJ	PRWJ	SRWJ	ARWJ
A	PeopleOnStreet	0.4263	0.5964	0.5114	0.5492	**0.5208**	0.4123	0.6088	0.5115	0.5871	**0.5299**	0.1742	0.3281	0.4839	0.4931	**0.3698**	0.4728	0.6550	0.5185	0.5915	**0.5595**	0.5008	0.6300	0.5292	0.6083	**0.5671**
	Traffic	0.4233	0.5431	0.5222	0.5535	**0.5105**	0.4838	0.6024	0.5313	0.5791	**0.5492**	0.2699	0.3933	0.5019	0.5129	**0.4195**	0.4699	0.5858	0.5276	0.5756	**0.5397**	0.5414	0.6223	0.5689	0.6210	**0.5884**
	Average	**0.4248**	**0.5697**	**0.5168**	**0.5513**	**0.5157**	**0.4480**	**0.6056**	**0.5214**	**0.5831**	**0.5396**	**0.2220**	**0.3607**	**0.4929**	**0.5030**	**0.3947**	**0.4713**	**0.6204**	**0.5231**	**0.5836**	**0.5496**	**0.5211**	**0.6261**	**0.5491**	**0.6146**	**0.5777**
B	BasketballDrive	0.5756	0.5929	0.6525	0.6258	**0.6117**	0.4733	0.4995	0.6127	0.6001	**0.5464**	0.4996	0.4965	0.6742	0.6034	**0.5684**	0.6361	0.6549	0.6776	0.6762	**0.6612**	0.7185	0.7380	0.7416	0.7449	**0.7357**
	BQTerrace	0.6331	0.8314	0.6685	0.7754	**0.7271**	0.6160	0.8265	0.6675	0.7617	**0.7179**	0.4841	0.6559	0.6259	0.6503	**0.6041**	0.6421	0.8369	0.6705	0.7819	**0.7329**	0.6231	0.7788	0.7062	0.7817	**0.7225**
	Cactus	0.5156	0.5995	0.5659	0.5975	**0.5696**	0.5416	0.6558	0.5883	0.6632	**0.6122**	0.4360	0.5094	0.5455	0.5398	**0.5077**	0.5395	0.6255	0.5760	0.6217	**0.5907**	0.5844	0.6582	0.6151	0.6692	**0.6317**
	Kimono	0.6590	0.8878	0.7167	0.8836	**0.7868**	0.6717	0.8325	0.7116	0.8379	**0.7634**	0.5033	0.8186	0.6795	0.8555	**0.7142**	0.7188	0.8176	0.7397	0.8264	**0.7756**	0.8675	0.8723	0.8750	0.8848	**0.8749**
	ParkScene	0.5052	0.6710	0.5436	0.6402	**0.5900**	0.4839	0.6340	0.5325	0.6107	**0.5653**	0.4232	0.5387	0.5305	0.5616	**0.5135**	0.5299	0.6940	0.5490	0.6551	**0.6070**	0.5581	0.6917	0.5822	0.6939	**0.6315**
	Average	**0.5777**	**0.7165**	**0.6294**	**0.7045**	**0.6570**	**0.5573**	**0.6897**	**0.6225**	**0.6947**	**0.6411**	**0.4692**	**0.6038**	**0.6111**	**0.6421**	**0.5816**	**0.6133**	**0.7258**	**0.6426**	**0.7123**	**0.6735**	**0.6703**	**0.7478**	**0.7040**	**0.7549**	**0.7193**
C	BasketballDrill	0.5097	0.5285	0.5632	0.5498	**0.5378**	0.3881	0.4161	0.5227	0.5275	**0.4636**	0.4321	0.4428	0.5507	0.5266	**0.4881**	0.5277	0.5529	0.5649	0.5625	**0.5520**	0.5952	0.6379	0.6079	0.6364	**0.6193**
	BQMall	0.4938	0.6893	0.6303	0.6643	**0.6194**	0.4894	0.6965	0.6353	0.7003	**0.6304**	0.4631	0.6146	0.6264	0.6257	**0.5824**	0.5104	0.7155	0.6342	0.6876	**0.6369**	0.5482	0.7136	0.6591	0.7348	**0.6639**
	RaceHorsesC	0.5908	0.6480	0.6014	0.6145	**0.6137**	0.5213	0.5951	0.5990	0.6291	**0.5861**	0.4398	0.4750	0.5755	0.5434	**0.5085**	0.5968	0.6593	0.5996	0.6243	**0.6200**	0.5477	0.5936	0.5979	0.6233	**0.5906**
	PartyScene	0.6206	0.7144	0.6957	0.7669	**0.6994**	0.4922	0.5672	0.6507	0.6982	**0.6021**	0.4717	0.5270	0.6881	0.6604	**0.5868**	0.6222	0.7175	0.6937	0.7734	**0.7017**	0.6040	0.6781	0.7035	0.7454	**0.6828**
	Average	**0.5537**	**0.6451**	**0.6226**	**0.6489**	**0.6176**	**0.4728**	**0.5687**	**0.6019**	**0.6388**	**0.5706**	**0.4517**	**0.5149**	**0.6102**	**0.5891**	**0.5414**	**0.5643**	**0.6613**	**0.6231**	**0.6619**	**0.6277**	**0.5738**	**0.6558**	**0.6421**	**0.6850**	**0.6392**
D	BasketballPass	0.5898	0.6907	0.6468	0.6880	**0.6538**	0.3892	0.4569	0.6058	0.6016	**0.5134**	0.4285	0.4830	0.6269	0.5917	**0.5325**	0.5813	0.6815	0.6480	0.6875	**0.6496**	0.6143	0.7099	0.6710	0.7157	**0.6777**
	BlowingBubbles	0.4810	0.5512	0.5691	0.5547	**0.5390**	0.4377	0.5020	0.5694	0.5550	**0.5160**	0.4648	0.5194	0.5601	0.5477	**0.5230**	0.4786	0.5528	0.5671	0.5575	**0.5390**	0.4673	0.5232	0.5713	0.5692	**0.5327**
	BQSquare	0.5852	0.8725	0.5675	0.7111	**0.6841**	0.4419	0.7321	0.5355	0.6346	**0.5860**	0.5536	0.7873	0.5649	0.6750	**0.6452**	0.5816	0.8701	0.5676	0.7112	**0.6826**	0.5192	0.7538	0.5717	0.7491	**0.6484**
	RaceHorses	0.5817	0.7774	0.6311	0.7981	**0.6971**	0.4943	0.6587	0.6134	0.7584	**0.6312**	0.3130	0.3888	0.5820	0.5752	**0.4647**	0.5836	0.7819	0.6322	0.8058	**0.7009**	0.6167	0.7652	0.6627	0.8066	**0.7128**
	Average	**0.5594**	**0.7230**	**0.6036**	**0.6880**	**0.6435**	**0.4408**	**0.5874**	**0.5810**	**0.6374**	**0.5617**	**0.4400**	**0.5446**	**0.5835**	**0.5974**	**0.5414**	**0.5563**	**0.7216**	**0.6037**	**0.6905**	**0.6430**	**0.5544**	**0.6880**	**0.6191**	**0.7101**	**0.6429**
E	FourPeople	0.5864	0.7465	0.6081	0.7011	**0.6605**	0.5662	0.6729	0.5300	0.5988	**0.5920**	0.4741	0.5634	0.5761	0.5948	**0.5521**	0.5917	0.7459	0.6115	0.7056	**0.6637**	0.5165	0.5758	0.5925	0.5961	**0.5702**
	Johnny	0.6397	0.7472	0.6643	0.7390	**0.6975**	0.5286	0.6197	0.5150	0.5899	**0.5633**	0.4392	0.4819	0.5691	0.5684	**0.5146**	0.6521	0.7605	0.6788	0.7542	**0.7114**	0.5198	0.5501	0.5917	0.5930	**0.5636**
	KristenAndSara	0.5395	0.6780	0.5911	0.6501	**0.6147**	0.4052	0.5452	0.4586	0.5337	**0.4857**	0.4208	0.5016	0.5536	0.5620	**0.5095**	0.5473	0.6896	0.5936	0.6612	**0.6229**	0.5058	0.5680	0.5880	0.6089	**0.5677**
	Average	**0.5885**	**0.7239**	**0.6211**	**0.6967**	**0.6576**	**0.5000**	**0.6126**	**0.5012**	**0.5741**	**0.5470**	**0.4447**	**0.5156**	**0.5663**	**0.5751**	**0.5254**	**0.5970**	**0.7320**	**0.6280**	**0.7070**	**0.6660**	**0.5140**	**0.5646**	**0.5907**	**0.5993**	**0.5672**
Total Average		**0.5408**	**0.6756**	**0.5987**	**0.6579**	**0.6183**	**0.4838**	**0.6128**	**0.5656**	**0.6256**	**0.5720**	**0.4055**	**0.5079**	**0.5728**	**0.5813**	**0.5169**	**0.5604**	**0.6922**	**0.6041**	**0.6710**	**0.6319**	**0.5667**	**0.6565**	**0.6210**	**0.6728**	**0.6293**

Table II
Comparisons of different types of RJVCQ scores with different weighting strategies for different video coding algorithms

RJVCQ	HM-16.19					TIP13-Seo					TIP16-Wang					TIP17-Gao					TBC19-Gao				
	W_1	W_2	W_3	W_4	W_5	W_1	W_2	W_3	W_4	W_5	W_1	W_2	W_3	W_4	W_5	W_1	W_2	W_3	W_4	W_5	W_1	W_2	W_3	W_4	W_5
MSE-based	0.1828	0.3618	0.5408	0.7199	0.8989	0.1069	0.2562	0.4055	0.5548	0.7042	0.1854	0.3346	0.4838	0.6330	0.7821	0.2043	0.3824	0.5604	0.7385	0.9166	0.2116	0.3892	0.5667	0.7443	0.9218
NSSIM-based	0.3270	0.5013	0.6756	0.8499	1.0242	0.1812	0.3446	0.5079	0.6713	0.8346	0.3126	0.4627	0.6128	0.7629	0.9130	0.3554	0.5238	0.6922	0.8606	1.0290	0.3479	0.5022	0.6565	0.8107	0.9650
PSNR-based	0.2093	0.4040	0.5987	0.7935	0.9882	0.1737	0.3732	0.5728	0.7724	0.9719	0.2015	0.3836	0.5656	0.7477	0.9297	0.2216	0.4128	0.6041	0.7954	0.9866	0.2465	0.4338	0.6210	0.8083	0.9955
SSIM-based	0.3270	0.4925	0.6579	0.8233	0.9887	0.3126	0.4691	0.6256	0.7821	0.9387	0.1812	0.3813	0.5813	0.7814	0.9814	0.3554	0.5132	0.6710	0.8289	0.9867	0.3479	0.5104	0.6728	0.8352	0.9977
Average	**0.2615**	**0.4399**	**0.6183**	**0.7967**	**0.9750**	**0.1936**	**0.3608**	**0.5280**	**0.6952**	**0.8624**	**0.2202**	**0.3906**	**0.5609**	**0.7313**	**0.9016**	**0.2842**	**0.4581**	**0.6319**	**0.8059**	**0.9797**	**0.2885**	**0.4589**	**0.6293**	**0.7996**	**0.9700**

Table III
Comparisons of different types of WJVCQ scores with different bit rates for different video coding algorithms

WJVCQ	HM-16.19				TIP13-Seo				TIP16-Wang				TIP17-Gao				TBC19-Gao			
	BR_1	BR_2	BR_3	BR_4	BR_1	BR_2	BR_3	BR_4	BR_1	BR_2	BR_3	BR_4	BR_1	BR_2	BR_3	BR_4	BR_1	BR_2	BR_3	BR_4
MSE-based	0.5172	0.5076	0.5476	0.5910	0.3097	0.4504	0.5418	0.6333	0.3002	0.3646	0.4525	0.5048	0.5164	0.5417	0.5745	0.6091	0.5282	0.6379	0.6468	0.4541
NSSIM-based	0.6603	0.6654	0.6863	0.6905	0.4710	0.6141	0.6563	0.7097	0.4116	0.4874	0.5564	0.5764	0.6642	0.6969	0.7066	0.7012	0.6120	0.7518	0.7443	0.5179
PSNR-based	0.5514	0.5780	0.6188	0.6467	0.4576	0.5550	0.6047	0.6453	0.5205	0.5561	0.5958	0.6188	0.5508	0.5848	0.6268	0.6539	0.5682	0.6366	0.6738	0.6055
SSIM-based	0.6404	0.6310	0.6659	0.6943	0.5324	0.6266	0.6490	0.6945	0.5355	0.5629	0.6069	0.6199	0.6457	0.6505	0.6826	0.7054	0.6401	0.7252	0.7224	0.6036
Average	**0.5923**	**0.5955**	**0.6297**	**0.6556**	**0.4427**	**0.5615**	**0.6130**	**0.6707**	**0.4420**	**0.4928**	**0.5529**	**0.5800**	**0.5943**	**0.6185**	**0.6476**	**0.6674**	**0.5871**	**0.6879**	**0.6968**	**0.5453**

V. CONCLUSION

This paper has introduced a new strategy for the joint video coding quality (JVCQ) evaluation to comprehensively consider the spatial and temporal influences on visual experience, i.e. the R-D performance and quality smoothness in the rate-constrained scenario. First, we decompose the visual quality into spatial and temporal dimensions, and use the FixedQP based normalization method to formulate the multiple objective optimization framework. Second, by considering different conditions on target bit rates and weighting strategies, we introduce the RJVCQ, WJVCQ and RWJVCQ metrics for comprehensive comparisons. By adopting the simple linear combination and weighting strategies, JVCQ can exhibit the general consistency of performances, and the low complexity makes it easily applicable in practice. The two items in the JVCQ are originally not comparable, but after normalization, the scores become not monotonically related bit rates, which can be deemed as another contribution of this paper to make JVCQ comparable among different bit rates and weighting strategies. Experiments illustrate consistency and applicability of the JVCQ model in evaluating video coding quality. The key contribution can be attributed to the adoption of the new optimization perspective based on the multiple objectives in video coding, and the comprehensive evaluations over different bandwidth constraints and weighting preferences. Therefore, the introduced JVCQ is promising to be further improved by adaptively allocating weights according to the practical encoding requirements and preferences, and to be used to guide the optimization process of video coding to achieve better coding quality and visual experience.

ACKNOWLEDGMENT

This work was supported in part by Natural Science Foundation of China under Grant 61801303, and in part by the Start-up Funding of Peking University Shenzhen Graduate School under Grant 2390101081.

REFERENCES

[1] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560-576, Jul. 2003.

[2] G. J. Sullivan, J. R. Ohm, W.-J. Han, and T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[3] W. Zhu, P. Cui, Z. Wang and G. Hua, “Multimedia Big Data Computing,” *IEEE MultiMedia*, vol. 22, no. 3, pp. 96-c3, July-Sept. 2015.

[4] B. Bross, W.-J. Han, J. R. Ohm, G. J. Sullivan, Y.-K. Wang, and T. Wiegand, “High Efficiency Video Coding (HEVC) test specification draft 10”, JCTVC-L1003, 12th JCTVC Meeting, Geneva, CH, Jan. 2013.

[5] H. Choi, J. Yoo, J. Nam, J. Yoo, D. Sim, I. V. Bajić, “Pixel-wise unified Rate-Quantization model for multi-level rate control,” *IEEE J. Selected Topics Signal Process.*, vol. 7, no. 6, pp. 1112-1123, Dec. 2013.

[6] B. Li, H. Li, L. Li and J. Zhang, “λ Domain Rate Control Algorithm for High Efficiency Video Coding,” *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3841-3854, Sept. 2014.

[7] W. Gao, S. Kwong, H. Yuan and X. Wang, “DCT coefficient distribution modeling and quality dependency analysis based frame-level bit allocation for HEVC,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 139-153, Jan. 2016.

[8] W. Gao, S. Kwong, Y. Zhou and H. Yuan, “SSIM-based game theory approach for rate-distortion optimized intra frame CTU-level bit allocation,” *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 988-999, Jun. 2016.

[9] W. Gao, S. Kwong and Y. Jia, “Joint Machine Learning and Game Theory for Rate Control in High Efficiency Video Coding,” *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 6074-6089, Dec. 2017.

[10] C. Seo, J. Moon and J. Han, “Rate Control for Consistent Objective Quality in High Efficiency Video Coding,” *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2442-2454, June 2013.

[11] M. Wang, K. N. Ngan and H. Li, “Low-Delay Rate Control for Consistent Quality Using Distortion-Based Lagrange Multiplier,” *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 2943-2955, July 2016.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, April 2004.

[13] M. H. Pinson and S. Wolf, “A new standardized method for objectively measuring video quality,” *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312-322, Sept. 2004.

[14] K. Seshadrinathan and A. C. Bovik, “Motion Tuned Spatio-Temporal Quality Assessment of Natural Videos,” *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 335-350, Feb. 2010.

[15] A. Mittal, M. A. Saad and A. C. Bovik, “A Completely Blind Video Integrity Oracle,” *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 289-300, Jan. 2016.

[16] G. J. Sullivan and T. Wiegand, “Rate-distortion optimization for video compression,” *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 74-90, Nov. 1998.

[17] G. Bjøntegaard, Calculation of Average PSNR Differences Between RD Curves, document VCEG-M33, Austin, TX, USA, Apr. 2001.

[18] (June 2019) HEVC Reference Software HM-16.19. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.19.

[19] B. Li, H. Li, L. Li, J. Zhang, “Rate control by R-lambda model for HEVC,” JCTVC-K0103, 11th Meeting, Shanghai, CN, Oct. 2012.

[20] B. Li, D. Zhang, H. Li, J. Xu, “QP determination by lambda value,” JCTVC-I0426, 9th JCTVC meeting, Geneva, CH, May 2012.

[21] Z. He, S. K. Mitra, “Optimum bit allocation and accurate rate control for video coding via ρ-domain source modeling,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 10, pp. 840-849, Oct. 2002.

[22] W. Gao, S. Kwong, Q. Jiang, C. Fong, P. H. W. Wong and W. Y. F. Yuen, “Data-Driven Rate Control for Rate-Distortion Optimization in HEVC Based on Simplified Effective Initial QP Learning,” *IEEE Trans. Broadcast.*, vol. 65, no. 1, pp. 94-108, March 2019.