Modeling of Rate and Perceptual Quality of Compressed Video as Functions of Frame Rate and Quantization Stepsize and Its Applications

Zhan Ma, *Student Member, IEEE,* Meng Xu, Yen-Fu Ou, and Yao Wang, *Fellow, IEEE*

***Abstract*—This paper first investigates the impact of frame rate and quantization on the bit rate and perceptual quality of compressed video. We propose a rate model and a quality model, both in terms of the quantization stepsize and frame rate. Both models are expressed as the product of separate functions of quantization stepsize and frame rate. The proposed models are analytically tractable, each requiring only a few content-dependent parameters. The rate model is validated over videos coded using both scalable and nonscalable encoders, under a variety of encoder settings. The quality model is validated only for a scalable video, although it is expected to be applicable to a single-layer video as well. We further investigate how to predict the model parameters using the content features extracted from original videos. Results show accurate bit rate and quality prediction (average Pearson correlation >0.99) can be achieved with model parameters predicted using three features. Finally, we apply rate and quality models for rate-constrained scalable bitstream adaptation and frame rate adaptive rate control. Simulations show that our model-based solutions produce better video quality compared with conventional video adaptation and rate control.**

***Index Terms*—Content feature, H.264/AVC, perceptual quality model, rate control, rate model, scalable video adaptation, scalable video coding (SVC).**

I. Introduction

A FUNDAMENTAL and challenging problem in video encoding is, given a target bit rate, how to determine at which spatial resolution (i.e., frame size), temporal resolution (i.e., frame rate), and amplitude (i.e., SNR) resolution [usually controlled by the quantization stepsize (QS) or consequently quantization parameter (QP)], to code the video. One may code the video at a high frame rate, large frame size, but high QS, yielding noticeable coding artifacts in each coded frame. Or one may use a low frame rate, small frame size, but small QS, producing high quality frames. These and other combinations can lead to very different perceptual quality. In traditional encoder rate-control algorithms, the spatial and temporal resolutions are prefixed based on some empirical rules, and the encoder varies the QS to reach a target bit rate. Selection of QS is typically based on models of rate versus QS. When varying the QS alone cannot meet the target bit rate, frames are skipped as necessary. Joint decision of QS and frame skip has also been considered, but often governed by heuristic rules, or using the mean square error (MSE) [1] as a quality measure. Ideally, the encoder should choose the spatial, temporal, and amplitude resolution (STAR) that leads to the best perceptual quality, while meeting the target bit rate. Optimal rate control solution requires accurate rate and perceptual quality prediction at any STAR combination.

In video streaming, the same video is often requested by receivers with diverse sustainable receiving rates. To address this diversity, a video may be coded into a scalable stream with many STAR combinations. Given a particular user's sustainable rate, either the server or proxy needs to extract from the original bitstream a certain layer corresponding to a particular STAR combination to meet the rate constraint. This problem is generally known as the bitstream adaptation. Different combinations are likely to yield different perceptual quality. Here again the challenging problem is to determine at which STAR to extract the bitstream, to maximize the perceptual quality. The latest H.264 scalable video coding (SVC) standard [2] enables lightweight bitstream manipulation [3] and also can provide the state-of-the-art coding performance [4], by its network-friendly interface design and efficient compression schemes inherited from the H.264/AVC [5]. However, before SVC video can be widely deployed for practical applications, efficient mechanisms for SVC stream adaptation to meet different user constraints need to be developed. Optimal adaptation requires accurate prediction of the perceived quality as well as the total rate at any STAR combination.

Although much work has been done in perceptual quality modeling and in rate modeling for video at a fixed spatial and temporal resolution, the impact of spatial and temporal resolutions on the perceptual quality and rate has not been studied extensively. To the best of our knowledge, no prior works have attempted to predict the rates corresponding to different STAR combinations, and none of the prior works has deployed rate and perceptual quality models to choose the best

Manuscript received March 13, 2011; revised July 20, 2011; accepted September 10, 2011. Date of publication November 22, 2011; date of current version May 1, 2012. This work was supported in part by the National Science Foundation, under Grant 0430145. This paper was recommended by Associate Editor O. C. Au.

Z. Ma was with the Polytechnic Institute of New York University, Brooklyn, NY 11201 USA. He is now with the Dallas Technology Laboratory, Samsung Telecommunications America, Richardson, TX 75082 USA (e-mail: zhan.ma@ieee.org).

M. Xu, Y.-F. Ou, and Y. Wang are with the Polytechnic Institute of New York University, Brooklyn, NY 11201 USA (e-mail: mxu02@students.poly.edu; you01@students.poly.edu; yao@poly.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2011.2177143

STAR combination for either video adaptation or encoder rate control. For encoder rate control, Liu and Kuo [1] attempted to jointly consider quantization (for spatial quality) and frame rate (for temporal quality); however, the quality is evaluated using MSE and the calculation of optimal quantization and frame rate requires intensive computation. Having accurate rate model and quality model would enable one to determine the optimal STAR combination for a given rate constraint efficiently and effectively, for both encoder rate control and adaptation of precoded scalable video.

In this paper, we focus on modeling the impact of temporal and amplitude resolutions (in terms of frame rate and quantization stepsize, respectively) on both rate and quality. We defer the spatial resolution for future study. We have developed the analytical rate and quality models based on scalable video bitstreams. The same models can be used for the single-layer video (coded using the H.264/AVC) as well, but with different values of model parameters in comparison to the scalable video. We further apply these models to do frame rate adaptive rate control for single-layer video encoding, and also solve the rate-constrained SVC adaptation problem.

Our rate model predicts the rate from quantization stepsize and frame rate. It uses the product of a metric that describes how the rate changes with the quantization stepsize when the video is coded at the highest frame rate, and a temporal correction factor for rate, which corrects the predicted rate by the first metric based on the actual frame rate. It fits the measured rates of decoded scalable video from different temporal and amplitude layers very accurately. We further validate that the rate model works for the single-layer video with high prediction accuracy as well. The Pearson correlation (PC) for all coding scenarios exceeds 0.99 over seven test sequences.

Our quality model relates the perceptual quality with the quantization stepsize and frame rate. It is derived based on our prior work [6], which uses the product of a metric that assesses the quality of a quantized video at the highest frame rate, based on the PSNR of decoded frames, and a temporal correction factor for quality (TCFQ), which reduces the quality assigned by the first metric according to the actual frame rate. In the quality model proposed here, we replace the first term by a metric that relates the quality of the highest frame rate video with the quantization stepsize. Each term has a single parameter, and the overall model is shown to fit very well with the subjective ratings, with an average PC of 0.97 over seven test sequences. Although the quality model is validated only for scalable video, we believe that the model works for the single-layer video as well.

Both rate and quality model parameters are content dependent. Hence, we also investigate how to predict the parameters accurately using features that can be computed from the original video. We develop a generalized linear predictor that can predict all five model parameters from three content features. Accurate bit rate and quality prediction can be achieved, with average PC over 0.99 for seven test sequences, using the predicted parameters. These content features can be calculated using a lightweight preprocessor, conducting macroblock based integer motion estimation on the original video.

In the remainder of this paper, we present the proposed rate model in Section II, and the quality model in Section III. Section IV details the parameter prediction using content features. Using these two developed models, we address the problem of rate-constrained bit stream adaptation for scalable video and frame rate adaptive rate control for single-layer video in Section V. Section VI concludes this paper and discusses future research directions.

II. Rate Model

In this section, we develop a rate model $R(q,t)$, which relates the rate R with the quantization stepsize q and frame rate t. Several prior works have considered rate modeling under a fixed frame rate, and have proposed models that relate the average bit rate versus quantization stepsize q. Ding and Liu reported the following model [7]:

$$R = \frac{\theta}{q^{\gamma}} \tag{1}$$

where θ and γ are model parameters, with $0 \le \gamma \le 2$. Chiang and Zhang [8] suggested the following model:

$$R = \frac{A_1}{q} + \frac{A_2}{q^2}. \tag{2}$$

This so-called quadratic rate model has been used for rate-control in MPEG-4 reference encoder [9]. We note that by choosing A_1 and A_2 appropriately, the model in (2) can realize the inverse power model of (1) with any $\gamma \in (1,2)$. Only the quadratic term was included in the model by Ribas-Corbera and Lei [10], that is

$$R = \frac{A}{q^2}. \tag{3}$$

Furthermore, a simplified and yet efficient linear rate-quantization model is developed in [11], that is

$$R = \mathrm{MAD} \cdot \frac{X_1}{\mathrm{QP}} + X_2 \tag{4}$$

where MAD is the mean absolute difference after signal prediction. X_1 and X_2 are model parameters, and they are updated with linear regression after encoding each frame. This linear rate-quantization model is used in the H.264/AVC reference software for rate control. He [12] proposed the ρ-model

$$R(\mathrm{QP}) = \theta\left(1 - \rho(\mathrm{QP})\right) \tag{5}$$

with ρ denoting the percentage of zero quantized transform coefficients with a given quantization parameter. This model has been shown to have high accuracy for rate prediction. A problem with the ρ-model is that it does not provide explicit relation between QP and ρ. Therefore, it does not lend itself to theoretical understanding of the impact of QP on the rate.

In our work on rate modeling, we focus on the impact of frame rate t on the bit rate R, under the same quantization stepsize q; while using prior models to characterize the impact of q on the rate, when the video is coded at a fixed frame rate. Toward this goal, we recognize that $R(q,t)$ can be written as

$$R(q,t) = R_{\max} R_q(q; t_{\max}) R_t(t; q) \tag{6}$$

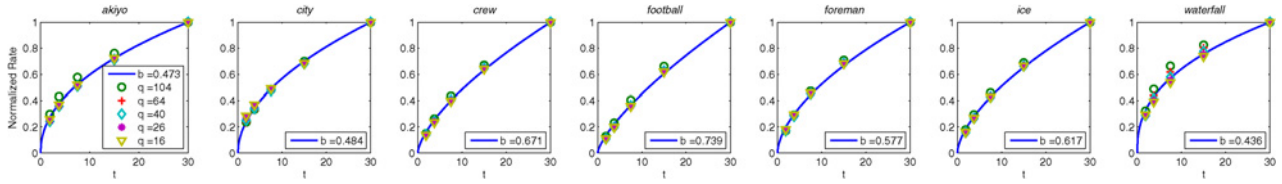


Fig. 1. NRT using different q. Points are measured rates, curves are predicted rates using (7).

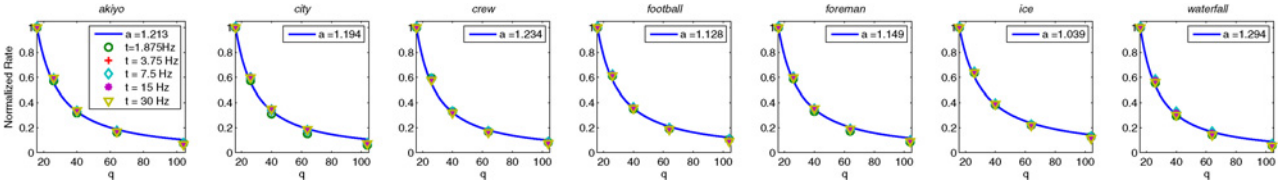


Fig. 2. NRQ using different frame rates t. Points are measured rates, curves are predicted rates using (8) at $t_{\max}$.

where $R_{\max} = R(q_{\min}, t_{\max})$ is the maximum bit rate obtained with a chosen minimal quantization stepsize $q_{\min}$ and a chosen maximum frame rate $t_{\max}$; $R_q(q; t_{\max}) = R(q, t_{\max})/R(q_{\min}, t_{\max})$ is the normalized rate versus quantization stepsize (NRQ) under the maximum frame rate $t_{\max}$, and $R_t(t; q) = R(q, t)/R(q, t_{\max})$ is the normalized rate versus temporal resolution (NRT) under a given quantization stepsize q. Note that the NRQ function $R_q(q; t_{\max})$ describes how the rate decreases when the quantization stepsize q increases beyond $q_{\min}$, under the frame rate $t_{\max}$; while the NRT function $R_t(t; q)$ characterizes how the rate reduces when the frame rate decreases from $t_{\max}$, under a given quantization stepsize q. As will be shown later by experimental data, the impact of q and t on the bit rate is actually separable, so that $R_t(t; q)$ can be represented by a function of t only, denoted by $R_t(t)$, and $R_q(q; t)$ by a function of q only, denoted by $R_q(q)$.

To see how quantization and frame rate, respectively, influence the bit rate, we encoded several test videos using the SVC reference software JSVM919 [13] and measured the actual bit rates corresponding to different q and t. Specifically, seven video sequences, *Akiyo*, *City*, *Crew*, *Football*, *Foreman*, *Ice*, and *Waterfall*, all in CIF (352×288) resolution, are encoded into five temporal layers using dyadic hierarchical prediction structure, with frame rates 1.875, 3.75, 7.5, 15, and 30 Hz, respectively, and each temporal layer contains five amplitude layers [1] obtained with QP of 44, 40, 36, 32, 28.[2] Using the H.264/AVC mapping between q and QP, i.e., $q = 2^{(\mathrm{QP}-4)/6}$, the corresponding quantization stepsizes are 104, 64, 40, 26, 16, respectively. The bit rates of all layers are collected and normalized by the rate at the highest frame rate $t_{\max}$, to find NRT points $R_t(t; q) = R(q, t)/R(q, t_{\max})$, for all t and q considered, which are plotted in Fig. 1. As shown in Fig. 1, the NRT data obtained with different qs overlap with each other, and can be captured by a single curve quite well. Similarly, the NRQ curves $R_q(q; t) = R(q, t)/R(q_{\min}, t)$ for different frame rates t are also almost invariant with the frame rate t, as shown in Fig. 2. These observations suggest that the effects of quantization q and frame rate t on the bit rate are separable, i.e., the quantization-induced rate variation is independent of the frame rate and vice versa. Therefore, the overall rate modeling problem is divided into two parts, one is to devise an appropriate functional form for $R_t(t)$, so that it can model the measured NRT points for all q in Fig. 1 accurately; the other is to derive an appropriate functional form for $R_q(q)$ that can accurately model the measured NRQ points in Fig. 2 for $t = t_{\max}$.

A. Model for Normalized Rate Versus Temporal Resolution

As explained earlier, $R_t(t)$ is used to describe the reduction of the normalized bit rate as the frame rate reduces. Therefore, the desired property for the $R_t(t)$ function is that it should be 1 at $t = t_{\max}$ and monotonically reduces to 0 at $t = 0$. Based on the measurement data in Fig. 1, we choose a power function

$$R_t(t) = \left(\frac{t}{t_{\max}}\right)^b. \tag{7}$$

Fig. 1 shows the model curve using this function along with the measured data. The parameter b is obtained by minimizing the squared error between the model predicted and measured rates. It can be seen that the model fits the measured data points very well. We also tried some other functional forms, including logarithmic and inverse falling exponential. We found that the power function yields the least fitting error.

B. Model for Normalized Rate Versus Quantization

Analogous to the $R_t(t)$ function, $R_q(q)$ is used to describe the reduction of the normalized bit rate as the quantization stepsize increases at a fixed frame rate. The desired property for the $R_q(q)$ function is that it should be 1 at $q = q_{\min}$ and monotonically reduces to 0 as q goes to infinity. Based on the measurement data in Fig. 2, we choose an inverse power function, that is

$$R_q(q) = \left(\frac{q}{q_{\min}}\right)^{-a}. \tag{8}$$

Fig. 2 shows the model curve using this function along with the measured data. It can be seen that the model fits the measured data points very well. The parameter a characterizes how fast the bit rate reduces when q increases. We also tried

[1]CGS or MGS can be used to provide amplitude scalability in SVC. We use constrained MGS to provide multiple amplitude resolutions, where motion estimation and compensation are constrained at current layer.

[2]Different from the JSVM default configuration utilizing different QPs for different temporal layers (i.e., QP cascading), the same QP is applied to all temporal layers at each amplitude layer.

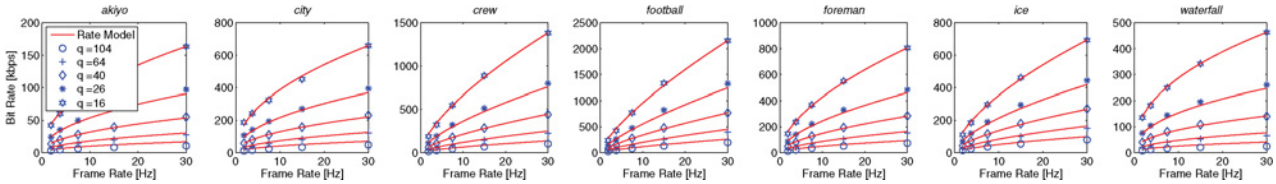


Fig. 3. Experimental rate points and predicted rates using (9) for scalable video, where bQP = 44, dQP = 4, and gopL = 16.

some other functional forms, including falling exponential. We found that the inverse power function yields the least fitting error. We note that the model in (8) is consistent with the model proposed by Ding and Liu [7], i.e., (1), for nonscalable video, where they have found that the parameter a is in the range of 0–2.

C. Overall Rate Model

Combining (6), (7), and (8), we propose the following rate model:

$$R(q,t) = R_{\max}\left(\frac{q}{q_{\min}}\right)^{-a}\left(\frac{t}{t_{\max}}\right)^{b} \tag{9}$$

where $q_{\min}$ and $t_{\max}$ should be chosen based on the underlying application, $R_{\max}$ is the actual rate when coding a video at $q_{\min}$, and $t_{\max}$, and a and b are the model parameters. For streaming precoded scalable video, $R_{\max}$ can be easily obtained at network proxy from the actual bitstream, while for encoder optimization (such as rate control), we need to estimate $R_{\max}$ accurately. For simplicity, we treat $R_{\max}$ as another rate model parameter in addition to a and b, and discuss parameter prediction using content features in Section IV.

The actual rate data of all test sequences with different combinations of q and t, and the corresponding estimated rates via the proposed model (9) are illustrated in Fig. 3. We note that the model predictions fit very well with the experimental rate points. The model parameters, a and b, are obtained by minimizing the root mean squared errors (RMSE) between the measured and predicted rates corresponding to all q and t. Table I lists the parameter values and model accuracy in terms of relative RMSE (i.e., RRMSE = RMSE/$R_{\max}$), and the PC between measured and predicted rates, defined as

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}} \tag{10}$$

where x_i and y_i are the measured and predicted rates, and n is the total number of available samples. We see that the model is very accurate for all different sequences, with very small RRMSE and very high PC.

Note that parameter a characterizes how fast the bit rate reduces when q increases. A larger a indicates a faster drop rate. As seen, the *Waterfall* which has the rich details has the largest a. Parameter b indicates how fast the rate drops when the frame rate decreases, with a larger b indicating a faster drop. As expected, the *Football* sequence, which has higher motion, has a larger b.

TABLE I
PARAMETERS FOR THE RATE MODEL AND MODEL ACCURACY FOR SCALABLE VIDEO WHERE BQP = 44 (Q = 104), DQP = 4, AND GOPL = 16

	a	b	$R_{\max}$	RRMSE	PC
Akiyo	1.213	0.473	163	1.54%	0.9985
City	1.194	0.484	658	1.67%	0.9977
Crew	1.234	0.671	1382	1.25%	0.9989
Football	1.128	0.739	2154	1.54%	0.9983
Foreman	1.149	0.577	806	1.31%	0.9990
Ice	1.039	0.617	693	1.44%	0.9986
Waterfall	1.294	0.436	462	1.47%	0.9984
Average				1.46%	0.9985

TABLE II
PARAMETERS FOR THE RATE MODEL AND MODEL ACCURACY FOR SCALABLE VIDEO WHERE BQP = 36, DQP = 6, AND GOPL = 8

	a	b	$R_{\max}$	RRMSE	PC
Akiyo	1.304	0.349	448	0.77%	0.9994
City	1.400	0.462	2093	1.51%	0.9980
Crew	1.126	0.639	3820	0.78%	0.9993
Football	0.955	0.678	4932	0.82%	0.9992
Foreman	1.344	0.545	2498	0.89%	0.9993
Ice	0.964	0.549	1461	0.55%	0.9996
Waterfall	1.614	0.391	1904	1.19%	0.9988
Average				0.93%	0.9991

D. Rate Model Validation for Different Scalable Coding Structures

The rate data and model parameters presented so far are for scalable videos obtained with a particular encoder setting, with base layer QP at 44 (i.e., bQP = 44). We also validated our rate model under different encoder settings. Fig. 4 and Table II show the model accuracy and model parameters for another encoder setting. Validation results for other encoder settings can be found in [14]. We have found that our rate model is accurate for different base layer QP, different group of picture (GOP) length (gopL), and different interlayer delta QP (dQP), with average RRMSE less than 1% and average PC larger than 0.99, for all encoder settings examined. Note that the model parameters depend on the encoder setting, in addition to video content.

E. Rate Model Validation for Single-Layer Video

In this subsection, we validate the rate model (9) for video encoded using the JSVM single-layer mode [13], which is a H.264/AVC compliant single-layer encoder. We provide two sets of results: one is for video coded using dyadic hierarchical B picture and the other is for those using conventional IPPP

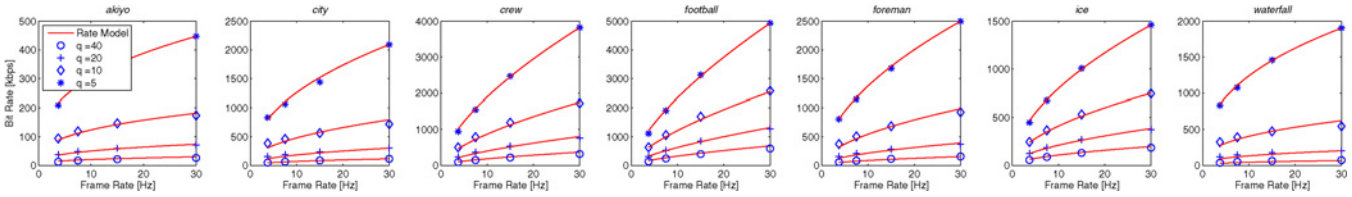


Fig. 4. Illustration of rate prediction using model (9) for scalable video, where bQP = 36 (q = 40), dQP = 6, and gopL = 8.

TABLE III
PARAMETERS FOR THE RATE MODEL AND MODEL ACCURACY FOR SINGLE-LAYER VIDEO USING HIERARCHICAL B STRUCTURE

	a	b	$R_{\max}$	RRMSE	PC
Akiyo	1.088	0.423	85	1.10%	0.9990
City	1.123	0.468	373	1.22%	0.9987
Crew	1.116	0.648	951	1.33%	0.9985
Football	0.982	0.708	1538	1.38%	0.9985
Foreman	1.082	0.562	478	1.13%	0.9988
Ice	0.837	0.595	420	1.17%	0.9988
Waterfall	1.199	0.434	274	0.64%	0.9996
Average				1.14%	0.9989

TABLE IV
PARAMETERS FOR THE RATE MODEL AND MODEL ACCURACY FOR SINGLE-LAYER VIDEO USING IPPP STRUCTURE

	a	b	$R_{\max}$	RRMSE	PC
Akiyo	1.272	0.490	108	1.17%	0.9988
City	1.464	0.599	540	1.66%	0.9983
Crew	1.187	0.699	1092	1.21%	0.9987
Football	1.020	0.739	1640	1.17%	0.9988
Foreman	1.353	0.639	624	1.47%	0.9984
Ice	0.933	0.605	443	1.21%	0.9986
Waterfall	1.492	0.621	459	1.07%	0.9994
Average				1.28%	0.9987

structure. In our experiments, the gopL is 16 for hierarchical B structure. Therefore, we can have five different frame rates, i.e., 1.875, 3.75, 7.5, 15, and 30 Hz. Similarly, we have encoded videos using the same five frame rates for IPPP structure. We code each video with different QPs ranging from 16 to 44.[3] Tables III, IV, Figs. 5, and 6 show that our rate model still works very well for single-layer video, with small relative RMSE and high PC. Our other experiments show that the proposed rate model works for the hierarchical P structure as well, and are also accurate at different spatial resolutions [14].

III. QUALITY MODEL

There are several published works examining the impact of either frame rate alone or both frame rate and quantization artifacts on the perceptual quality. Please see [6] for a review of prior work on this subject.

Like the rate model, we focus on examining the impact of frame rate on the quality, under the same quantization stepsize; while trying to use prior models to characterize the impact of quantization stepsize q on the quality, when the video is coded at a fixed frame rate. Our quality model is extended from our earlier work [6], [15]. The proposed model is motivated by the following decomposition, which has a general form of:

$$Q(q,t) = Q_{\max} Q_q(q;t_{\max}) Q_t(t;q) \tag{11}$$

where $Q_q(q;t_{\max}) = Q(q,t_{\max})/Q(q_{\min},t_{\max})$ is the normalized quality versus quantization stepsize (NQQ) under the maximum frame rate $t_{\max}$; $Q_t(t;q) = Q(q,t)/Q(q,t_{\max})$ is the normalized quality versus temporal resolution (NQT) under a given quantization stepsize q, and $Q_{\max} = Q(q_{\min},t_{\max})$. Note that $Q_q(q;t_{\max})$ models the impact of quantization on the quality when the video is coded at $t_{\max}$; while $Q_t(t;q)$ describes how the quality reduces when the frame rate reduces, under the same q. In other words, $Q_t(t;q)$ corrects the predicted quality by $Q_{\max} Q_q(q;t_{\max})$ based on the actual frame rate, and for this reason is also called TCFQ.

In our prior work [15], we conducted subjective tests to obtain mean opinion scores (MOS) for the same set of test sequences coded using joint temporal and amplitude scalability. The subjective tests were performed for decoded sequences, at frame rates of 30, 15, 7.5, 3.75 Hz, and QP equals to 28, 36, 40 and 44 (corresponding to quantization stepsize of 16, 40, 64, 104, respectively). We have found that the NQT data are quite independent of q [6]. As shown in Fig. 7, NQT can be modeled accurately using an inverse exponential function of the form

$$Q_t(t) = \frac{1 - e^{-d\frac{t}{t_{\max}}}}{1 - e^{-d}}. \tag{12}$$

In our prior work [15], the NQQ term is expressed in terms of the PSNR of encoded frames. In this paper, we try to model the NQQ term directly in terms of the quantization stepsize q, to enable optimization of both t and q for video encoding and adaptation. By observing the NQQ data shown in Fig. 8, we propose to use an exponential function

$$Q_q(q) = e^c e^{-c\frac{q}{q_{\min}}} \tag{13}$$

with c as the model parameter. Compared with the original two-parameter sigmoid function proposed in [15], the single-parameter exponential function is much simpler and easier to analyze. Comparing the measured and predicted quality in Fig. 8, we see that the model captures the quantization-induced quality variation very well at the highest frame rate.

Combing (11), (12) and (13), the normalized quality defined as $\tilde{Q}(q,t) = Q(q,t)/Q_{\max}$ can be modeled as

$$\tilde{Q}(q,t) = \frac{e^{-c\frac{q}{q_{\min}}}}{e^{-c}} \frac{1 - e^{-d\frac{t}{t_{\max}}}}{1 - e^{-d}}. \tag{14}$$

[3]Because of the space limitation, we only present the results for QP = 44, 40, 36, 32, 28.

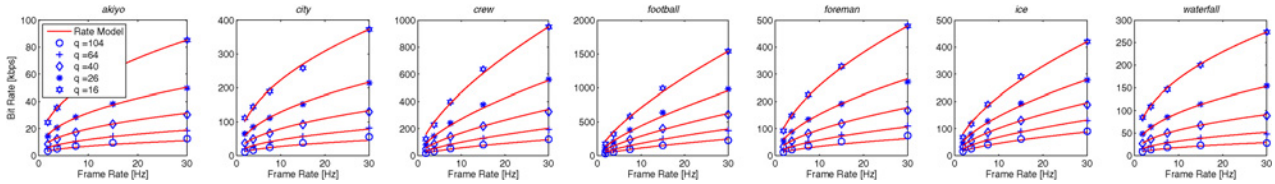


Fig. 5. Illustration of rate prediction using model (9) for single-layer video using hierarchical B structure.

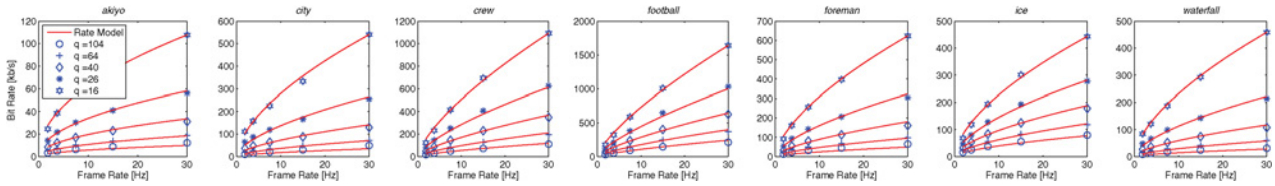


Fig. 6. Illustration of rate prediction using model (9) for single-layer video using IPPP structure.

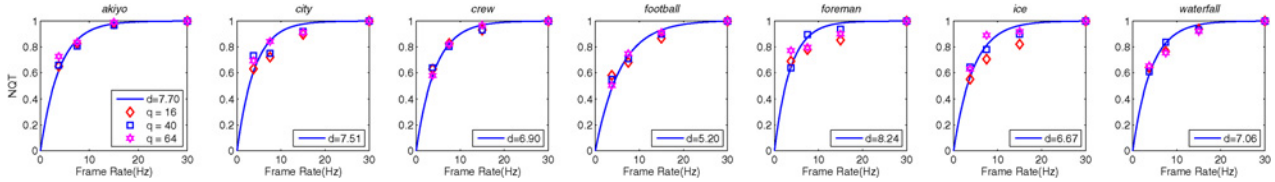


Fig. 7. Normalized quality versus temporal resolution (NQT), for different quantization stepsize q. Points are measured data, curves are predicted quality using (12).

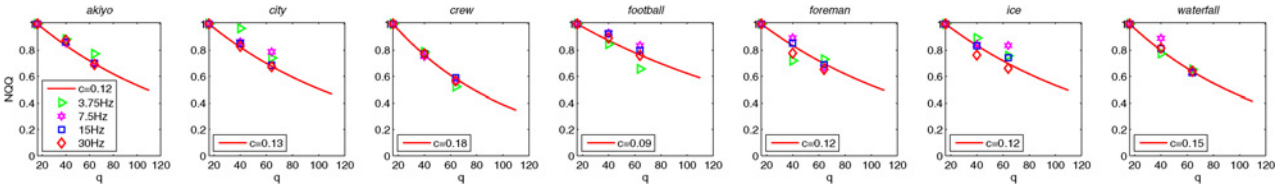


Fig. 8. Normalized quality versus the quantization stepsize (NQQ) for different frame rates t. Points are measured data and curves are predicted quality for t = 30 Hz, using (13).

Fig. 9 compares the measured and predicted quality ratings by the model in (14). The parameters c, d are obtained by least-squares error fitting to match the quality. We use NQT data at three different QPs for curving fitting in this paper, while NQT data at four different QPs are used in [6]. Table V summarizes the parameters and the model accuracy in terms of RRMSE (defined as RMSE/$Q_{\max}$) and PC values for the seven sequences. Overall, the proposed model, with only two content-dependent parameters, predicts the MOS very well, for sequences *Akiyo*, *Crew*, *Football*, and *Waterfall* with a very high PC (>0.98). The model is less accurate for *Ice*, *Foreman* and *City*, but still has a quite high PC. We would like to point out that the measured MOS data for these two sequences do not follow a consistent trend at some quantization levels, which may be due to the limited number of viewers participating the subjective tests.

Note that parameter c indicates how fast the quality drops with increasing q, with a larger c suggesting a faster drop. On the other hand, parameter d reveals how fast the quality reduces as the frame rate decreases, with a smaller d corresponding to a faster drop.

TABLE V
PARAMETERS FOR THE QUALITY MODEL AND MODEL ACCURACY

	c	d	RRMSE	PC
Akiyo	0.12	7.70	3.06%	0.9868
City	0.13	7.51	6.41%	0.9448
Crew	0.18	6.90	2.50%	0.9926
Football	0.09	5.20	4.54%	0.9801
Foreman	0.12	8.24	5.49%	0.9419
Ice	0.12	6.67	5.38%	0.9575
Waterfall	0.15	7.06	3.65%	0.9823
Average			4.40%	0.9694

IV. MODEL PARAMETER PREDICTION USING CONTENT FEATURES

As shown in previous sections, model parameters are highly content dependent. In this section, we investigate how to predict the parameters accurately using content features that can be computed from original video signals. We have five parameters in total for both rate and quality models, i.e., a, b, $R_{\max}$, c, and d. According to our simulations, we

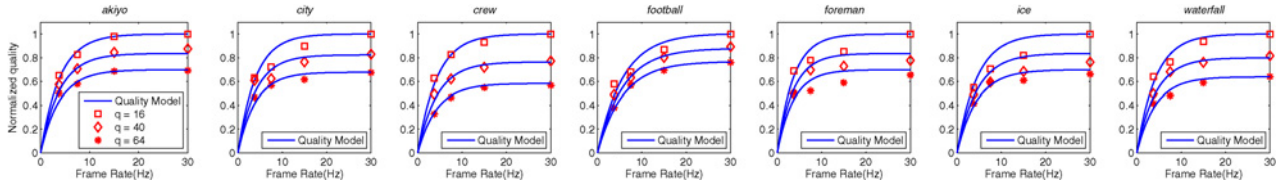


Fig. 9. Normalized quality versus quantization stepsize and frame rate. Points are normalized MOS data; curves are predicted quality using (14).

TABLE VI
LIST OF CONTENT FEATURES IN CONSIDERATION

Input source	Feature	
Original features		
Residual	FD/DFD	μ_{FD}, σ_{FD}, μ_{DFD}, σ_{DFD}
Motion	MVM/MDA	μ_{MVM}, σ_{MVM}, σ_{MDA}
Original	VFC	σ_{org}
Internormalized features		
$\eta(\mu_{\mathrm{FD}}, \sigma_{\mathrm{org}}) = \mu_{\mathrm{FD}}/\sigma_{\mathrm{org}}$, $\eta(\mu_{\mathrm{DFD}}, \sigma_{\mathrm{org}}) = \mu_{\mathrm{DFD}}/\sigma_{\mathrm{org}}$ $\eta(\mu_{\mathrm{MVM}}, \sigma_{\mathrm{org}}) = \mu_{\mathrm{MVM}}/\sigma_{\mathrm{org}}$, $\eta(\mu_{\mathrm{MVM}}, \sigma_{\mathrm{MVM}}) = \mu_{\mathrm{MVM}}/\sigma_{\mathrm{MVM}}$ $\eta(\mu_{\mathrm{MVM}}, \sigma_{\mathrm{MDA}}) = \mu_{\mathrm{MVM}}/\sigma_{\mathrm{MDA}}$		

have found that these parameters are related to the *residual* (error) signal, such as frame difference (FD), displace frame difference (DFD), and so on; *motion* fields, such as motion vector magnitude (MVM), motion direction activity (MDA), and so on; as well as the video frame contrast (VFC) using the *original video signal*. Toward this goal, we have implemented a simple, lightweight preprocessor, which uses integer motion estimation over fixed size (16×16) blocks. A set of content features is derived from residual signal, motion fields and original signal, as detailed in Table VI. More details regarding the original feature definition can be found in [6]. Note that we choose not to consider those features used in our prior work [6] that require Gabor filtering, to reduce the feature computation complexity.

In our prior work for quality modeling [6], we use a generalized linear predictor to predict each parameter from a few features. The features to be included and the predictor coefficients for different parameters are determined separately. In this paper, we try to find a minimal set of features that can predict all the parameters for all test sequences accurately. Let $p_{m,j}$ denote the jth parameter of the mth sequence, $m = 1, \ldots, M$, and $f_{m,k}$ the value of the kth feature, $k = 1, 2, \ldots, K$, then $p_{m,j}$ is predicted using a generalized linear predictor $h_{j,0} + \sum_{k=1}^{K} h_{j,k} f_{m,k}$. These predictors can be described using a vector form $\mathbf{P}_m = \mathbf{H}\mathbf{F}_m$, where $\mathbf{F}_m = [1, f_1, \ldots, f_K]^T$, $\mathbf{H}$ is a $5 \times (K+1)$ matrix containing the coefficients $h_{j,k}$. Let $\mathbf{P}_m = [a_m, b_m, c_m, d_m, R_{\max,m}]^T$ denote the vector containing the original parameters of the mth sequence. We try to find both the feature set and $\mathbf{H}$ that will minimize a prediction error. In order for the solution to be generalizable to other sequences outside our test sequences, we use the leave-one-out cross-validation error (CVE) criteria [6]. For a particular set of chosen features, we arbitrarily set one sequence as the test sequence (i.e., γ_t) and the remaining $(M-1)$ sequences as the training sequences (i.e., Γ). We determine the weights $h_{j,k}$ to minimize the mean square fitting error for the $(M-1)$ training sequences, defined as $\sum_{m \in \Gamma} \|\mathbf{P}_m - \mathbf{H}\mathbf{F}_m\|^2$. We then evaluate the fitting error for the test sequence, i.e., $\|\mathbf{P}_m - \mathbf{H}\mathbf{F}_m\|^2$, $m = \gamma_t$. We repeat this process, each time using a different sequence as the test sequence. The average of the fitting errors for all the test sequences is the CVE associated with this feature set. For a given K, the features that lead to the least CVE are chosen. We evaluate the CVE starting with $K = 1$. We increase K until the minimal CVE does not reduce significantly. The resulting K features are the final set of features chosen. We then recompute the weighting coefficients to minimize the average fitting error over all test sequences, i.e., $\sum_{m=1}^{M} \|\mathbf{P}_m - \mathbf{H}\mathbf{F}_m\|^2$.

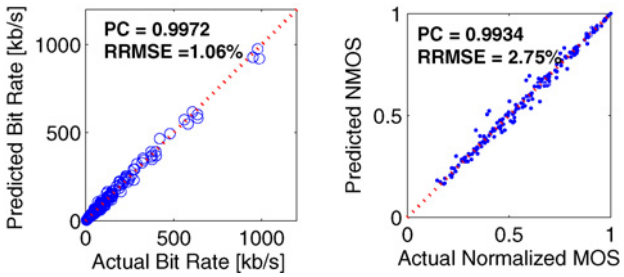


Fig. 10. Rate and quality model accuracy where the parameters are predicted using content features for single-layer video coded via hierarchical B structure at CIF resolution.

We have found that three features, i.e., μ_{FD}, μ_{MVM}, $\eta(\mu_{\mathrm{MVM}}, \sigma_{\mathrm{MDA}})$ are sufficient to provide good model accuracy as shown in Fig. 10, where we apply the predicted model parameters and verify the model accuracy between predicted and actual measured data points (for both bit rate and quality). The rate data are those obtained with single-layer encoding using hierarchical B structure. And the optimal predictor matrix is

$$\mathbf{H} = \begin{bmatrix} 1.1406 & -0.0330 & -0.0611 & 0.1408 \\ 0.4462 & 0.0112 & 0.0680 & -0.0667 \\ 0.1416 & -0.0008 & -0.0001 & -0.0036 \\ 8.9757 & -0.5728 & -0.8516 & 2.0528 \\ 67.73 & 49.45 & 281.7 & -245.6 \end{bmatrix}. \quad (15)$$

For other coding structures, the same three features provide high model accuracy, but the matrix $\mathbf{H}$ differs. For rate data obtained using all coding structures, results show that the three-feature prediction provides PC > 0.99, and RRMSE $< 3\%$. Note that the rows in $\mathbf{H}$ corresponding to the quality model parameters stay the same since we assume the same quality model for both single layer and scalable video.

We note that the quality model parameters very much depend on the underlying viewers. In this paper, the model is derived based on MOS obtained from a relatively large group of viewers, and hence is meant to characterize an "average viewer." Such models are useful when one designs

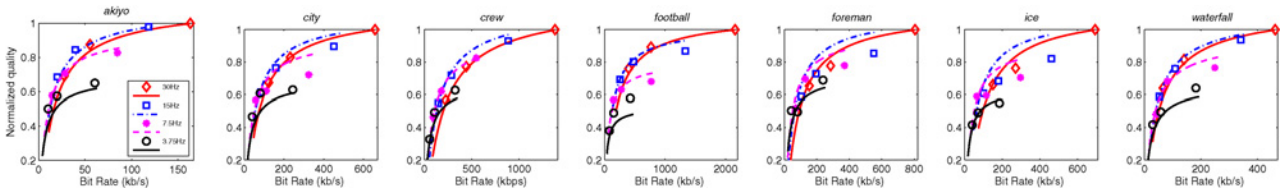


Fig. 11. Normalized quality versus rate at different frame rates. Points are measured data, curves are based on the rate model in (9) and the quality model in (14).

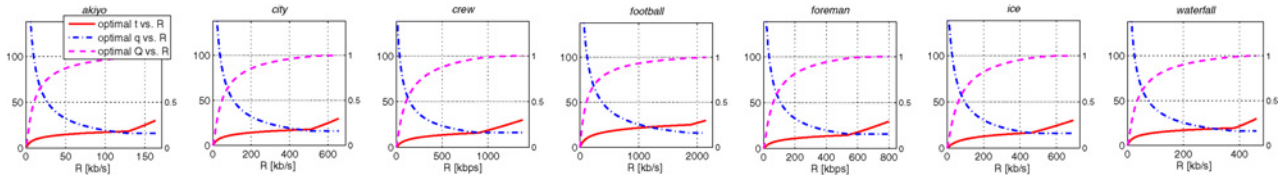


Fig. 12. Optimal quantization stepsize q_{opt}, frame rate t_{opt}, and the corresponding normalized quality $\tilde{Q}_{\mathrm{opt}}$ versus the bit rate R by assuming q and t can take on any continuous values within their respective ranges. (Left y-axis: t, q, right y-axis: normalized quality).

a video system to optimize the perceptual quality for all potential viewers. For any particular viewer, parameters c and d are likely dependent on the viewer's sensitivities to quantization artifacts and motion jerkiness, respectively. In order to optimize for individual user's perceptual quality, one can obtain the MOS data for different viewer categories, and find model parameters by fitting the model form to the MOS data from different viewer categories. To summarize, the rows of **H** corresponding to the rate model depend on the encoder setting, whereas rows of **H** for the quality model parameters depend on the viewer category.

V. Applications

Using our developed rate and quality models, we propose to solve the rate-constrained scalable bitstream adaptation at a streaming server or proxy, and conduct the model driven frame rate adaptive rate control at a video encoder for single-layer video. Although they are quite different applications, the essence of the problem is similar, i.e., given the bit rate budget, we need to provide the optimal combination of frame rate and quantization stepsize, so as to yield the best video quality. In scalable video streaming, the optimal combination is used to extract the temporal and amplitude layers from the full-resolution bitstream at network proxy. In rate control, the optimal frame rate and quantization stepsize (or consequently QP) are configured as the encoding frame rate and QP for single-layer video encoding. With our developed models, we have made the rate control problem analytically tractable, without requiring intensive computation [1].

A. Rate-Constrained Bit Stream Adaptation

Combining the rate and quality models, we draw in Fig. 11, quality versus rate curves achievable at different frame rates for the SVC encoder setting given in Table I and Fig. 3. We also plot the measured MOS and rate data on the same figure. The model fits the measured data very well for sequences *Akiyo*, *Crew*, and *Waterfall*. But the model is not as accurate at some frame rates for *Football*, *City*, *Foreman*, and *Ice*. The inaccuracy is mainly due to the difference between the predicted quality and MOS for these sequences (see Fig. 9). It is clear from this figure that each frame rate is optimal only for a certain rate region.

For a given target rate R_0, the adaptation problem can be formulated as the following constrained optimization problem:

$$\begin{aligned} &\text{Determine } t, q \text{ to } \ \text{maximize } \ \tilde{Q}(q,t) \\ &\text{subject to } \ R(q,t) \le R_0, q \ge q_{\min}, 0 < t \le t_{\max}. \end{aligned} \tag{16}$$

In the following subsections, we employ proposed rate and quality models to solve this optimization problem, first assuming the frame rate can be any positive value in a continuous range, and then considering the discrete set of frame rates afforded by the dyadic temporal prediction structure. Finally, we consider a more practical case where both the frame rates and quantization stepsizes are discrete.

1) *Optimal Solution Under Continuous t and q:* We first solve the constrained optimization problem in (16) assuming both the frame rate t and quantization stepsize q can take on any value in the range of $t \in (0, t_{\max}]$, $q \in [q_{\min}, +\infty)$. To simplify the notation, let $\hat{Q} = \frac{1}{(1-e^{-d})e^{-c}}$, $\hat{t} = t/t_{\max}$, $\hat{q} = q/q_{\min}$, $\hat{R} = R/R_{\max}$, and $\hat{R}_0 = R_0/R_{\max}$, the rate and quality models in (9) and (14) become, respectively

$$\hat{R}\left(\hat{q}, \hat{t}\right) = \hat{q}^{-a}\hat{t}^{b} \tag{17}$$

$$\tilde{Q}\left(\hat{q}, \hat{t}\right) = \hat{Q}e^{-c\hat{q}}\left(1 - e^{-d\hat{t}}\right). \tag{18}$$

The range constraints for $\hat{q}$ and $\hat{t}$, respectively, become $\hat{q} \ge 1$, $\hat{t} \in (0, 1]$. By setting $\hat{R}\left(\hat{q}, \hat{t}\right) = \hat{R}_0$ in (17), we obtain

$$\hat{q} = \sqrt[a]{\left(\hat{t}^{b}/\hat{R}_0\right)} \tag{19}$$

which describes the feasible q for a given t, to satisfy the rate constraint R_0. Note that because $\hat{q} \ge 1$, the minimal normalized frame rate needed to reach a target rate $\hat{R}_0$ is $\hat{t}_{\min}(\hat{R}_0) = \sqrt[b]{\hat{R}_0}$.

Substituting (19) into (18) yields

$$\tilde{Q}(\hat{t}) = \hat{Q}e^{-\frac{c\cdot\hat{t}^{b}}{\sqrt[a]{\hat{R}_0}}}\left(1 - e^{-d\hat{t}}\right), \quad \hat{t} \in [\hat{t}_{\min}(\hat{R}_0), 1] \tag{20}$$

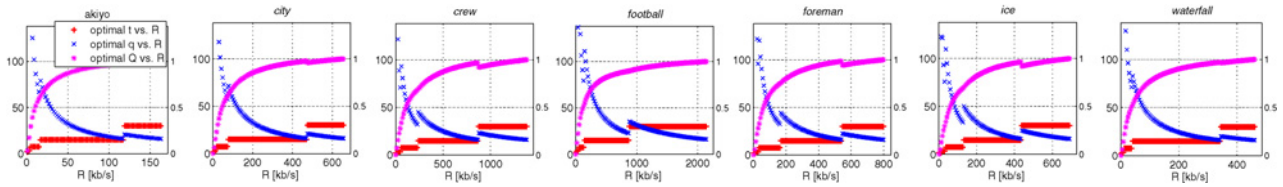


Fig. 13. Optimal operating points q_{opt}, t_{opt}, and $\tilde{Q}_{\text{opt}}$ versus R by assuming t can only take discrete values allowed by the dyadic prediction structure, whereas q can vary continuously. (Left y-axis: t, q, right y-axis: normalized quality).

where $\psi = b/a$. Equation (20) expresses the achievable quality with different frame rates under the rate constraint R_0. Clearly, this function has a unique maximum $\hat{t}_{\text{opt}}$, which can be derived by setting its derivative with respect to $\hat{t}$ to zero. This yields

$$\hat{R}_0 = \left(\frac{c\psi \hat{t}_{\text{opt}}^{\psi-1}(1 - e^{-d\hat{t}_{\text{opt}}})}{de^{-d\hat{t}_{\text{opt}}}} \right)^a. \tag{21}$$

Unfortunately, it is difficult to invert (21) to find an analytical relation of $\hat{t}_{\text{opt}}$ in terms of $\hat{R}_0$. However, for any given R_0 and hence $\hat{R}_0$, we can numerically solve for $\hat{t}_{\text{opt}}$ using (21). If the resulting $\hat{t}_{\text{opt}}$ is smaller than $\hat{t}_{\min}(\hat{R}_0)$, we will reset $\hat{t}_{\text{opt}}$ to $\hat{t}_{\min}(\hat{R}_0)$. With the so-determined t_{opt}, we can then determine the optimal quantization stepsize q_{opt}, and the corresponding normalized maximum quality $\tilde{Q}_{\text{opt}}$, using (19) and (20), respectively. Fig. 12 shows t_{opt}, q_{opt}, and $\tilde{Q}_{\text{opt}}$ as functions of the rate constraint R_0. As expected, as the rate increases, t_{opt} increases while q_{opt} reduces, and the achievable best quality continuously improves. Once the q_{opt} reaches the $q_{\min}$, q_{opt} stays at $q_{\min}$, while t_{opt} increases linearly with the normalized rate. Notice that t_{opt} increases more rapidly for the *Football* sequence than for the other sequences, because of its faster motion.

2) *Optimal Solution Under Dyadic Temporal Scalability Structure:* A popular way to implement temporal scalability is through the dyadic hierarchical B-picture prediction structure, by which the frame rate doubles with each more temporal layer. With five temporal layers, the corresponding frame rates are 1.875, 3.75, 7.5, 15, and 30 Hz. From a practical point of view, it will be interesting to see what are the optimal combinations of the frame rate and quantization stepsize for different bit rates under this structure. To obtain the optimal solution under this scenario, for each given rate $\hat{R}_0$, we determine the quality values corresponding to all possible frame rates in the range of $(\hat{t}_{\min}(\hat{R}_0), 1]$ using (20), and choose the frame rate [and its corresponding quantization stepsize using (19)] that leads to the highest quality. The results are shown in Fig. 13. Because the frame rate t can only increase in discrete stepsize, the optimal q does not decrease monotonically with the rate. Rather, whenever t_{opt} jumps to the next higher value (doubles), q_{opt} first increases to meet the rate constraint, and then decreases as the rate increases while t is held constant. Note that when q reaches $q_{\min}$ when t is still below $t_{\max}$ (e.g. *Crew* at around 800 kb/s), to reach higher target rate, t will jump to $t_{\max}$. To meet the target rate, q initially jumps to higher values, and then decreases until it reaches $q_{\min}$. There is a rate region (e.g., 800–1200 for *Crew*), for which the achievable quality is actually lower then that achieved at a lower rate (e.g., 800 kb/s for *Crew*). This means that with the $q_{\min}$ constraint, we should not operate in this rate region. Note for *Football*, this phenomenon does not occur, as q does not reach $q_{\min}$ at 15 Hz. In general, such problem can be avoided by choosing a sufficiently low $q_{\min}$.

TABLE VII
TARGET RATE ASSIGNMENT

Video	SVC Adaptation	Rate Control
Akiyo	16 32 64 176	16, 32, 48, 64
City	64 128 256 768	64, 128, 192, 256
Crew	128 256 512 1536	128, 256, 512, 768
Football	256 512 1024 1536	128, 256, 512, 1024
Foreman	64 128 256 880	64, 128, 256, 384
Ice	64 128 256 768	64, 128, 256, 384
Waterfall	32 64 256 512	32, 64, 128, 256

The results in Fig. 13 can be validated by cross checking with Fig. 11. For example, for *Football*, in the rate region between 38 and 64 kb/s, 3.75 Hz leads to the highest quality, in the rate range between 64 and 141 kb/s, 7.5 Hz gives the highest quality, between 141 and 906 kb/s, 15 Hz is the best, and beyond 906 kb/s, 30 Hz provides the highest quality. Connecting the top segments for each sequence in Fig. 11 will lead to the optimal $\tilde{Q}$ versus bit rate curve in Fig. 13.

3) *Performance Evaluation for Practical Adaptation System:* In practice, the SVC encoder with amplitude scalability does not allow the quantization stepsize to change continuously. Thus, both frame rate and quantization stepsize are discrete.

To illustrate how to apply the proposed models in a practical video adaptation system, we consider SVC video adaptation for the target rates given in the left part of Table VII. The reason we select different target rates for different videos is because some video requires much higher rates to achieve similar quality. We choose the maximum target rate based on the $R_{\max}$ corresponding to SVC in Table I. For each desired bit rate, we first obtain the optimal t_{opt} and q_{opt} according to the solution discussed in Section V-A2, where t_{opt} is discrete and q_{opt} is continuous. We then quantize the continuous q_{opt} to its nearest discrete level from above. We compare this model-based solution ("Model") to two alternative approaches. "Alternative 1" simply takes the pair of (q, t) that has a rate that is closest to the target bit rate from below. For "Alternative 2," we choose the frame rate based on the target bit rate, following the heuristic rule: $t = 3.75$ Hz when $R \leq 16$ kb/s, $t = 7.5$ when $16 < R \leq 64$, $t = 15$ when $64 < R < 256$, $t = 30$ Hz for $R \geq 256$ kb/s. The q is then chosen such that the rate is closest to the target rate from below.

Fig. 14 illustrates the performance comparison for scalable video adaptation between our model based solution and

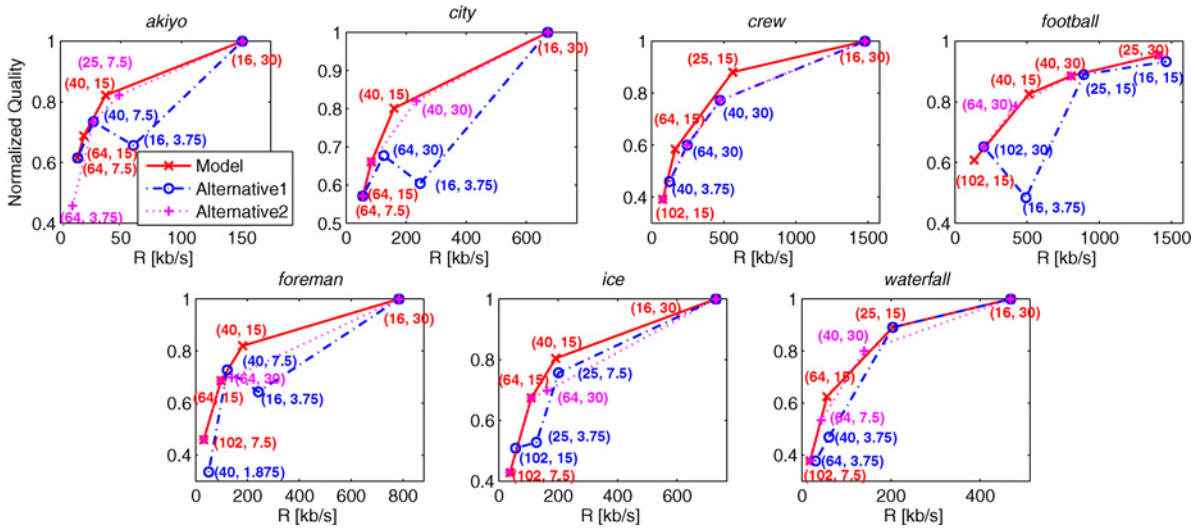


Fig. 14. Perceptual quality comparison for our model based scalable video adaptation and alternative scheme with corresponding (q, t) annotated for each point. Overlapped points are annotated with one (q, t).

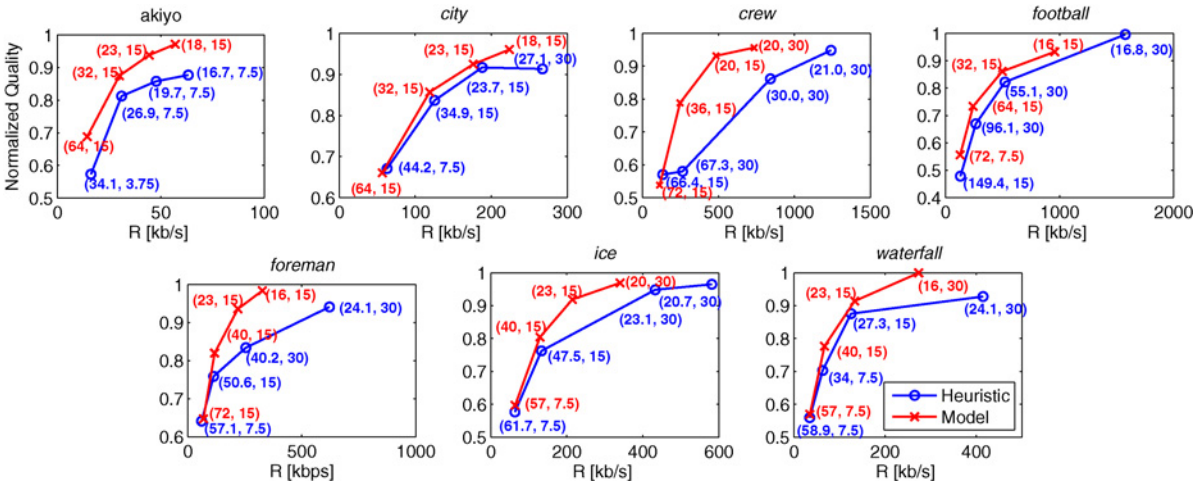


Fig. 15. Normalized quality comparison for proposed model based rate control and heuristic scheme, average q (over entire sequence) and frame rate (t) are annotated for each corresponding point.

two alternative schemes. Results show that our model based solution can yield significantly higher quality at many rate points than "Alternative 1." "Alternative 2" achieves quality levels that are quite close to our model-based solution, because the frame rates chosen by the heuristic rule are fairly close to the optimal frame rates for many target rates. When the frame rates differ, the model based solution can yield noticeable quality improvement.

To implement the model-based solution, the proxy needs to know the rate and quality model parameters. Because the proxy has the complete scalable stream available, it can easily derive the rate model parameters from the rates corresponding to different (t, q) combinations using least-squares fitting. On the other hand, the parameters for the quality model need to be determined based both on the video content and the viewer preference setting. We can embed content features $\mathbf{F}$ in the full-resolution bitstream, and construct model parameters via $\mathbf{P} = \mathbf{HF}$, as described in Section IV, where $\mathbf{H}$ contains only rows corresponding to the quality model parameters. For different coding structures, user preferences, and so on, we may predesign different $\mathbf{H}$. In a simpler implementation, the adaptor may ignore the user's preference setting, and use the $\mathbf{H}$ matrix tuned for "average" viewers.

B. Frame Rate Adaptive Rate Control

Together with the proposed model parameter predictor, our proposed models can be embedded in the H.264/AVC encoder to do frame rate adaptive rate control. A preprocessor is applied to collect the content features and compute the parameters for both rate and quality models as described in Section IV. Please note that we have the same best features for different encoder settings for all test videos. On the other hand, we can store the predictor $\mathbf{H}$ for different encoder settings for real-time encoding. Then these parameters are plugged into proposed models [i.e., (9) and (14)] to determine the optimal frame rate t_{opt}, quantization stepsize q_{opt} and consequently $\mathrm{QP}_{\mathrm{opt}}$ so as to yield the best video quality given the total bit rate budget R_0. In practice, we derive the discrete t_{opt} and q_{opt} (or $\mathrm{QP}_{\mathrm{opt}}$) as discussed in Section V-A3, but use the parameter predictor designed for single-layer encoder.

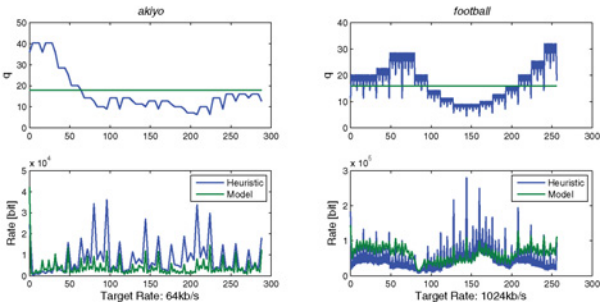


Fig. 16. Illustration of q and bits versus video display order.

We have implemented such model based frame rate adaptive rate control on top of the JSVM [13] single-layer encoding mode, which is compliant with the H.264/AVC standard. With our proposed rate and quality models, we make the determination of frame rate and QP analytically tractable. Because we have to determine the content features and consequently the model parameters before encoding, it is not easy to embed the feature computation in the encoding process and use the same motion estimation results. Hence we choose to perform motion estimation on original video frames in the feature extraction preprocessor, as described in Section IV.

We have evaluated our model based algorithm ("Model") and a heuristic rate control scheme ("Heuristic") for four different target bit rates given in the right part of Table VII. These target rates are chosen based on the maximum rate for single-layer encoding given in Table III. With the "Heuristic" approach we choose the frame rate following the heuristic rule adopted for "Alternative 2" for video adaptation. Adaptive initial QP algorithm [13] is enabled in the heuristic scheme. QP adjustment range is [12, 51] (as implemented in JSVM) for the heuristic scheme. All simulations are conducted using hierarchical B structure (gopL = 8). Fig. 15 summarizes experimental results for seven different sequences. Overall, we can see that our proposed method provides better perceptual quality under the same rate than the heuristic method, and leads to rates that are quite close to the target rate. Surprisingly, the JSVM rate control scheme produces rates that are much higher than the target rates at the high rate range. Fig. 16 plots the q and bits consumption versus video display order for *Akiyo* and *Football* at 64 kb/s and 1024 kb/s, respectively. q varies largely for the heuristic approach which results in the large bit variation, while using a constant QP derived from our model leads to more stable bit consumption and the average rate matches the target rate more closely. Other sequences have the similar trend.

VI. Conclusion

In this paper, we examined the impact of frame rate t and quantization stepsize q on the rate and perceptual quality of scalable video. Both models are expressed as the product of a function of q and a function of t. Each requires only a few content-dependent parameters (i.e., two for quality model and three for rate model). The same rate model is validated for different coding structures including both scalable and nonscalable. The rate model fits the measured rates very accurately, with an average PC larger than 0.99, over seven video sequences. The quality model is validated for scalable video only, and also matches the MOS from subjective tests very well, with an average PC of 0.97. We expect that the same quality model form applies to videos coded using different coding structures, including single-layer video. In fact, we suspected that videos coded under the same frame rate and QP with different encoding structures will have similar perceptual quality, and hence the same quality model parameters may apply as well. These hypotheses need to be validated in future studies.

We also investigated how to predict the model parameters accurately using content features. Results show that three features are sufficient for providing accurate bit rate and MOS prediction (with average PC larger than 0.99). We found that best features are the same for different encoder settings, but with different predictor matrix $\mathbf{H}$.

We further applied these models for rate-constrained scalable bitstream adaptation and nonscalable encoder rate control. Results show that our model driven video adaptation and frame rate adaptive rate control produces better video quality compared with other alternative strategies.

Currently, our rate and quality models are developed assuming the video content is stationary over entire sequence. In reality, video content may change over the time and is not stationary. Adaptation of model parameters, either in sliding window approach [16], or based on scene change detection, would be necessary. This is also a direction for future work.

Acknowledgment

The authors would like to thank the reviewers for their valuable comments.

References

[1] S. Liu and C.-C. J. Kuo, "Joint temporal-spatial bit allocation for video coding with dependency," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 1, pp. 15–27, Jan. 2005.

[2] G. Sullivan, T. Wiegand, and H. Schwarz, *Text of ITU-T Rec. H.264 | ISO/IEC 14496-10:200X/DCOR1/AMD.3 Scalable Video Coding*, ISO/IEC JTC1/SC29/WG11, MPEG08/N9574, Antalya, TR, Jan. 2008.

[3] Y.-K. Wang, M. Hannuksela, S. Pateux, A. Eleftheriadis, and S. Wenger, "System and transport interface SVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1149–1163, Sep. 2007.

[4] M. Wien, H. Schwarz, and T. Oelbaum, "Performance analysis of SVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1194–1203, Sep. 2007.

[5] H.264/AVC, *Draft ITU-T Rec. and Final Draft Int. Std. of Joint Video Spec. (ITU-T Rec. H.264\ISO/IEC 14496-10 AVC)*, document JVT-G050, Joint Video Team, Mar. 2003.

[6] Y.-F. Ou, Z. Ma, and Y. Wang, "Perceptual quality assessment of video considering both frame rate and quantization artifacts," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 3, pp. 286–298, Mar. 2010.

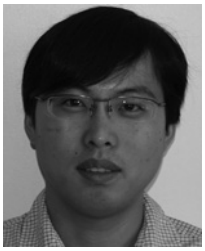
[7] W. Ding and B. Liu, "Rate control of MPEG video coding and recoding by rate-quantization modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 2, pp. 12–20, Feb. 1996.

[8] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 246–250, Feb. 1997.

[9] T. Chiang, H.-J. Lee, and H. Sun, "An overview of the encoding tools in the MPEG-4 reference software," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2000, pp. 295–298.

[10] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 2, pp. 172–185, Feb. 1999.

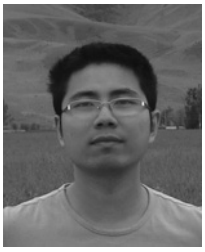
[11] Y. Liu, Z. G. Li, and Y. C. Soh, "A novel rate control scheme for low delay video communication of H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 1, pp. 68–78, Jan. 2007.
[12] Z. He and S. K. Mitra, "A novel linear source model and a unified rate control algorithm for H.264/MPEG-2/MPEG-4," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, May 2001, pp. 1777–1780.
[13] Joint Scalable Video Model (JSVM), *JSVM Software*, document JVT-X203, Joint Video Team, Geneva, Switzerland, Jun. 2007.
[14] Z. Ma, "Modeling of power, rate and perceptual quality of scalable video and its applications," Ph.D. dissertation, Dept. ECE, Polytechnic Instit. New York Univ., Brooklyn, NY, Jan. 2011.
[15] Y.-F. Ou, Z. Ma, and Y. Wang, "A novel quality metric for compressed video considering both frame rate and quantization artifacts," in *Proc. Int. Workshop Video Process. Quality Metrics Consumer*, Jan. 2009, pp. 1–6.
[16] J. Dong and N. Ling, "A context-adaptive prediction scheme for parameter estimation in H.264/AVC macroblock layer rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 8, pp. 1108–1117, Aug. 2009.



Zhan Ma (S'06) received the B.S. and M.S. degrees in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2004 and 2006, respectively, and the Ph.D. degree in electrical engineering from the Polytechnic Institute of New York University, Brooklyn, in 2011.

While pursuing the M.S. degree, he joined the National Digital Audio and Video Standardization (AVS) Workgroup to participate in standardizing the video coding standard in China. He interned with the Thomson Corporate Research Laboratory, Princeton, NJ, Texas Instruments, Dallas, TX, and Sharp Laboratories of America, Camas, WA, in 2008, 2009, and 2010, respectively. Since 2011, he has been with the Dallas Technology Laboratory, Samsung Telecommunications America, Richardson, TX, as a Senior Standards Researcher. His current research interests include the next-generation video coding standardization, video fingerprinting, and video signal modeling.

Dr. Ma received the 2006 Special Contribution Award from the AVS Workgroup, China, for his contribution in standardizing the AVS Part 7, and the 2010 Patent Incentive Award from Sharp.

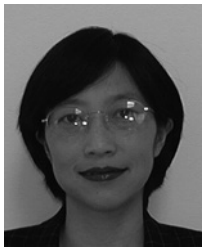


Meng Xu received the B.S. degree in physics from Nanjing University, Nanjing, China, in 2006, and the M.S. degree in electrical engineering from the Polytechnic Institute of New York University, Brooklyn, in 2009, where he is currently pursuing the Ph.D. degree in electrical engineering.

He interned with the Dialogic Media Laboratory, Eatontown, NJ, in 2010. His current research interests include video bit rate modeling, scalable video coding, and its applications.

Yen-Fu Ou received the B.S. and M.S. degrees in mechanical engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2000 and 2002, respectively, and the Masters degree in electrical and computer engineering from Columbia University, New York, in 2006. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, Polytechnic Institute of New York University, Brooklyn.

Since 2009, he has been participating in the Research Project with OoVoo, LCC, New York, and mainly focuses on quality-of-service (QoS) system design for video conferencing and transmission. His current research interests include perceptual video/image quality, video adaptation, and QoS systems on video streaming and image pattern recognition.



Yao Wang (M'90–SM'98–F'04) received the B.S. and M.S. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1983 and 1985, respectively, and the Ph.D. degree in electrical and computer engineering from the University of California, Santa Barbara, in 1990.

Since 1990, she has been with the Electrical and Computer Engineering Faculty, Polytechnic University, Brooklyn, NY (now Polytechnic Institute of New York University). She is the leading author of the textbook *Video Processing and Communications* (Prentice-Hall, 2001). Her current research interests include video coding and networked video applications, medical imaging, and pattern recognition.

Dr. Wang has served as an Associate Editor for the IEEE TRANSACTIONS ON MULTIMEDIA and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. She received the New York City Mayor Award for Excellence in Science and Technology in the Young Investigator Category in 2000. She was a co-winner of the IEEE Communications Society Leonard G. Abraham Prize Paper Award in the Field of Communications Systems in 2004. She received the Overseas Outstanding Young Investigator Award from the National Natural Science Foundation of China in 2005 and was named the Yangtze River Lecture Scholar by the Ministry of Education of China in 2007.