MODELING OF RATE AND PERCEPTUAL QUALITY OF VIDEO AND ITS APPLICATION TO FRAME RATE ADAPTIVE RATE CONTROL

Zhan Ma[†]*, Meng Xu*[†]*, Kyeong Yang*[‡] *and Yao Wang*[†]

[†] Dept. of Electrical and Computer Engineering
Polytechnic Institute of NYU, Brooklyn, NY 11201
[‡] Video Technologies, Dialogic Inc., Parsippany, NJ 07054

ABSTRACT

In a prior work, we have developed both rate and perceptual quality models for temporal and amplitude (i.e., SNR) scalable video produced by the H.264/SVC encoder. In this paper, we validate from experimental data that the functional form of the rate model is applicable to H.264/AVC encoded video, which has the same temporal scalability but no SNR scalability, but the model parameter values differ. We further investigate how to predict both rate and quality model parameters using content features computed from the original video. Experimental data show that with proper feature combination, we can estimate the model parameters very accurately, and the estimated bit rate and quality using the predicted model parameters match with the measured bit rate and quality with high Pearson correlation (PC) and small root mean square error (RMSE). We have implemented a simple pre-processor in the H.264/AVC encoder to guide the frame rate adaptive rate control. Results show that our model-based frame rate adaptive rate control outperforms the default rate control algorithm with better quality.

Index Terms— Rate model, perceptual quality model, frame rate adaptive rate control, H.264/AVC

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

This paper focuses on the joint impact of quantization and frame rate. We first develop two analytical models to address the effects of frame rate and quantization on bit rate and perceptual quality respectively. These two analytical models are initially built for temporal- and SNR-scalable video [2]. We have validated that the same functional forms of the rate model can also be applied to the video with temporal scalability only, but with different values for model parameters. According to our extensive simulations, we have found the model parameters are content dependent. Hence, we propose a simple preprocessor in the encoder to extract the content features, such as displaced frame difference, video contrast, motion vectors, etc, to predict the parameters. Results show that with proper feature combination, we can predict the model parameter very accurately (with average PC $>$0.99). Together with the predicted parameters, we apply our proposed models to guide the frame rate adaptive rate control, i.e., given a bit rate budget, we can obtain the optimal combination of the frame rate and quantization stepsize (i.e., $t_{\texttt{opt}}$, $q_{\texttt{opt}}$) so as to produce the best video quality. Optimal frame rate and quantization stepsize (as initial quantization stepsize) are passed into the H.264/AVC encoder to do the rate control.

The rest of this paper is organized as follows: Sec. 2 introduces the analytical models for rate and perceptual quality, followed by the discussion on model parameter prediction using content features in Sec. 3. Frame rate adaptive rate control using proposed rate and quality models is explored in Sec. 4. Sec. 5 concludes our work and discusses the future direction.

Emails: zhan.ma@gmail.com, mxu02@students.poly.edu, kyeong.yang@dialogic.com, yao@poly.edu.

2. ANALYTICAL RATE AND QUALITY MODEL

In [2], we have proposed two analytical models regarding the rate and perceptual quality for scalable video coded using the H.264/SVC encoder with both temporal and SNR scalability. Specifically, rate model is the product of a power function of quantization stepsize q and a power function of frame rate t, i.e.,

$$R(q,t) = R_{\max}\left(\frac{q}{q_{\min}}\right)^{-a}\left(\frac{t}{t_{\max}}\right)^{b}, \qquad (1)$$

where $R_{\max} = R(q_{\min}, t_{\max})$, a and b are content dependent parameters, $q_{\min}$ and $t_{\max}$ are known constants for typical applications. Currently, we assume $q_{\min} = 16$ (corresponding to QP 28) and $t_{\max} = 30$ Hz. Recently, we have validated model (1) for videos coded at different frame rates and QP using the H.264/AVC compliant encoder. The different frame rates are realized using the dyadic hierarchical B structure. In our experiments, the GOP (group of picture) length is 16. Therefore, we can have five different frame rates, i.e., 1.875, 3.75, 7.5, 15 and 30 Hz. Results show that our rate model still works very well for single layer video (i.e., without SNR scalability), with small relative root mean square error (RMSE) and high Pearson correlation (PC) as shown in Table 1. Figure 1 shows our proposed rate model can predict the rate very accurately. Although we only show results for CIF resolution video here, we have validated that the model is very accurate for videos at QCIF, 4CIF, WVGA and 720p resolutions.

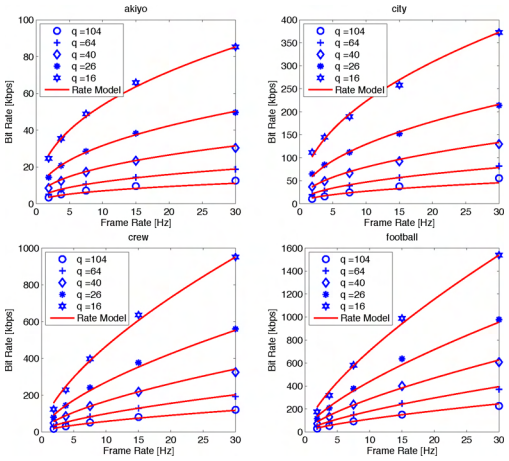

Fig. 1. Illustration of rate prediction using model (1) for single layer video.

Previously, we have also developed the perceptual quality model considering both frame rate and quantization artifacts [2, 3]. The overall quality model is the product of an inverted exponential function of frame rate and another exponential function of quantization stepsize, i.e.,

$$Q(q,t) = Q_{\max}\frac{e^{-c\frac{q}{q_{\min}}}}{e^{-c}}\frac{1-e^{-d\frac{t}{t_{\max}}}}{1-e^{-d}}, \qquad (2)$$

where c and d are content dependent parameters, $Q_{\max}$ is the quality of the video coded at $q_{\min}$ and $t_{\max}$ and is assumed to be a constant. We use a range of 0 to 100 for the quality, and we found that $Q_{\max}$ = 90 through our subjective tests. Although our quality model is derived based on subjective tests on videos coded using the scalable extension of H.264/AVC (SVC), we believe that the functional form of the model and model parameters are still applicable to the videos coded without SNR scalability. This is because the perceptual quality of a video coded directly at a frame rate t and quantization stepsize q, is expected to be the same as a video coded using temporal scalability and SNR scalability to reach the same t and q, although the bitstreams may be quite different and they may require different bit rates. Figure 2 plots the subjective quality prediction for different videos. Table 2 gives the model parameters and its accuracy.

Table 1. Parameters for the rate model and model accuracy

	akiyo	city	crew	football
a	1.088	1.123	1.116	0.982
b	0.423	0.468	0.648	0.708
$R_{\max}$	85	373	951	1538
RMSE/$R_{\max}$	1.10%	1.22%	1.33%	1.38%
PC	0.9990	0.9987	0.9985	0.9985

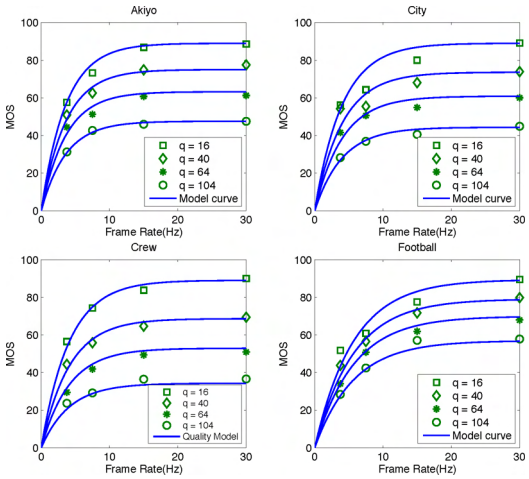

Fig. 2. Illustration of quality prediction using model (2).

3. MODEL PARAMETER PREDICTION USING CONTENT FEATURES

As shown in Section 2, model parameters are content dependent. In this section, we investigate how to predict the pa-

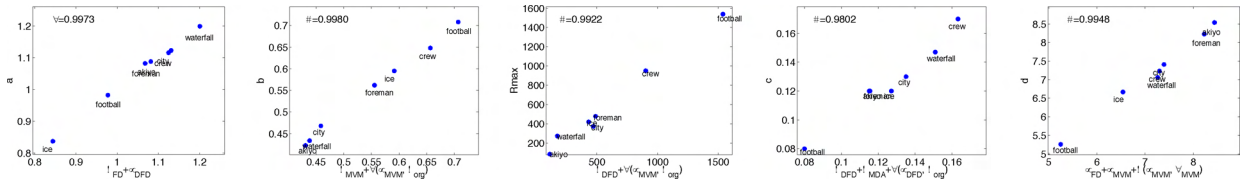

Fig. 3. Rate and quality model parameter prediction.

Table 2. Parameters for the quality model and model accuracy

	akiyo	city	crew	football
c	0.12	0.13	0.18	0.09
d	7.70	7.51	6.90	5.20
RMSE/$Q_{\max}$	3.06%	6.41%	2.50%	4.54%
PC	0.9868	0.9448	0.9926	0.9801

rameters accurately using content features which can be easily obtained from underlying video signals. According to our simulations, we have found that parameters are related to the residual (error) signal, such as frame difference (FD), displace frame difference (DFD), etc; motion fields, such as motion vector magnitude (MVM), motion direction activity (MDA), etc; as well as the video frame contrast (VFC, determined by the standard deviation of the gray level) using the original video signal. To reduce the complexity in computing the motion-based features, we apply the macroblock (i.e., 16x16) based integer motion estimation using the default fast algorithm in JSVM. Currently, we conduct the pre-processing for each frame, and assign a motion vector to every macroblock. A set of content features is computed from the residual signal, motion fields and original signal, as detailed in Table 3. In general, this set of content features consists of two subsets. One includes the original features derived from raw input source. The other contains the inter-normalized features using the prior subset.

Table 3. List of content features in consideration

input source	feature	
original features		
residual	FD/DFD	$\mu_{\mathtt{FD}}$, $\sigma_{\mathtt{FD}}$, $\mu_{\mathtt{DFD}}$, $\sigma_{\mathtt{DFD}}$
motion	MVM/MDA	$\mu_{\mathtt{MVM}}$, $\sigma_{\mathtt{MVM}}$, $\sigma_{\mathtt{MDA}}$
original	VFC	$\sigma_{\mathtt{org}}$
inter-normalized features		
$\eta(\mu_{\mathtt{FD}}, \sigma_{\mathtt{org}}) = \mu_{\mathtt{FD}}/\sigma_{\mathtt{org}}$, $\eta(\mu_{\mathtt{DFD}}, \sigma_{\mathtt{org}}) = \mu_{\mathtt{DFD}}/\sigma_{\mathtt{org}}$ $\eta(\mu_{\mathtt{MVM}}, \sigma_{\mathtt{org}}) = \mu_{\mathtt{MVM}}/\sigma_{\mathtt{org}}$, $\eta(\mu_{\mathtt{MVM}}, \sigma_{\mathtt{MVM}}) = \mu_{\mathtt{MVM}}/\sigma_{\mathtt{MVM}}$ $\eta(\mu_{\mathtt{MVM}}, \sigma_{\mathtt{MDA}}) = \mu_{\mathtt{MVM}}/\sigma_{\mathtt{MDA}}$		

According to our experimental data, we have found that a single feature does not estimate the parameter very well, thus several features are combined together and their weighted sum is used to predict the model parameter, i.e., $\sum_k \omega_k F_k + \omega_0, k = 1, 2, \ldots, K$. For a given K, we apply the leave-one-out cross-validation to choose the best features and the weighting coefficients, so that the solution is generalizable to sequences outside our test videos. Assume the total number of sequences is M. For a particular set of chosen features, we arbitrarily set one sequence as the test sequence and the remaining $M-1$ sequences as the training sequences. We determine the weights ω_k to minimize the mean square fitting error for the training sequences. We then evaluate the square fitting error for the test sequence. 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 cross-validation error (CVE) associated with this feature set. We compute the CVE for each possible combination of K features, and we use the K features that lead to the smallest CVE. The optimal weighting coefficients are finally computed using all sequences for chosen best features. Figure 3 shows the best feature set for rate and quality model parameters using two and three features, respectively. The actual predictors are given below:

$$a = 1.11 - 0.035\sigma_{\mathtt{FD}} + 0.049\mu_{\mathtt{DFD}}$$
$$b = 0.42 + 0.082\sigma_{\mathtt{MVM}} + 0.7\eta(\mu_{\mathtt{MVM}}, \sigma_{\mathtt{org}})$$
$$c = 0.10 - 0.026\sigma_{\mathtt{DFD}} + 0.077\sigma_{\mathtt{MDA}} + 0.26\eta(\mu_{\mathtt{DFD}}, \sigma_{\mathtt{org}})$$
$$d = 7.05 - 0.92\mu_{\mathtt{FD}} + 0.048\mu_{\mathtt{MVM}} + 0.43\eta(\mu_{\mathtt{MVM}}, \sigma_{\mathtt{MVM}})$$
$$R_{\max} = -25.35 - 162.2\sigma_{\mathtt{DFD}} + 214.8\eta(\mu_{\mathtt{MVM}}, \sigma_{\mathtt{org}})$$

Results show that with proper feature combination, we can estimate the model parameter accurately with average PC $>$ 0.99.

4. FRAME RATE ADAPTIVE RATE CONTROL

Together with the proposed model parameter estimation method, our proposed models can be embedded in the H.264/AVC encoder to do frame rate adaptive rate control. Figure 4 shows the systematic illustration for our model based frame rate adaptive rate control. At first, the pre-processor is applied to the original video to compute the necessary features and consequently the model parameters (including a, b, c, d, $R_{\max}$). Then these parameters are plugged into proposed models to do rate-constrained optimization analytically [2]. We then quantize the optimal analytical solution to

the discrete value, i.e., $t_{\mathtt{opt}}$ and $q_{\mathtt{opt}}$ for a whole sequence, so as to yield the best video quality given the total bit rate budget R_0. The $t_{\mathtt{opt}}$ is used to set the video frame rate, and $\mathrm{QP}_{\mathtt{opt}}$ is configured as the initial QP for the H.264/AVC rate control. We use the default QP search range (i.e., [9, 51]) in the JSVM software without change. We have implemented such model based frame rate adaptive rate control on top of the JSVM [4] single layer encoding mode (but with the hierarchical B-structure enabled), which is compliant with the H.264/AVC standard. Conventionally, we have to assign the video frame rate and initial QP for rate control manually, either through brute force iteration, or empirical estimation. With our proposed rate and quality models, we make the determination of the optimal frame rate and initial QP analytically tractable. Moreover, because the parameters are obtained through the simple preprocessor which takes underlying raw video as input, our model based frame rate adaptive rate control can be applied widely.

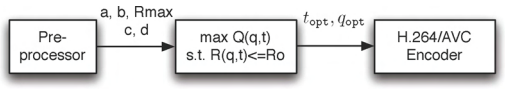

Fig. 4. Proposed frame rate adaptive rate control.

We have evaluated our algorithm and JSVM default rate control for four different bit rates, i.e., 64, 128, 256 and 512 kbps. In our proposed algorithm, the frame rate and initial QP are derived using proposed rate and quality models. For default JSVM encoder, we set 30 Hz frame rate for 256 and 512 kbps, 15 Hz for 128 kbps and 7.5 Hz for 64 kbps. Adaptive initial QP algorithm is enabled in th default rate control [4] to assign the initial QP. QP adjustment range is the same for our method and default scheme. All simulations are conducted using CIF resolution videos. Figure 5 summarizes experimental results for four different sequences. Perceptual quality (i.e., Q) is measured at sequence level. The quality value is in the range of [0,1] with "1" and "0" for the "best" and "worst" quality respectively. Overall, we can see that our proposed method provides better perceptual quality at some rate range, such as akiyo@64kbps, crew@256kbps, 512kbps, football@256kbps, etc.

5. CONCLUSION

In this paper, we have developed the rate and perceptual quality models for single layer video encoded by the H.264/AVC that specifically consider the impact of frame rate and quantization step- size on the bit rate and quality. This two analytical models have been extended from our previous work [2] for scalable video with different parameter values. We have found the parameters can be well predicted using the proper content feature combination. With the predicted parameters, we apply our models to do frame rate adaptive rate control

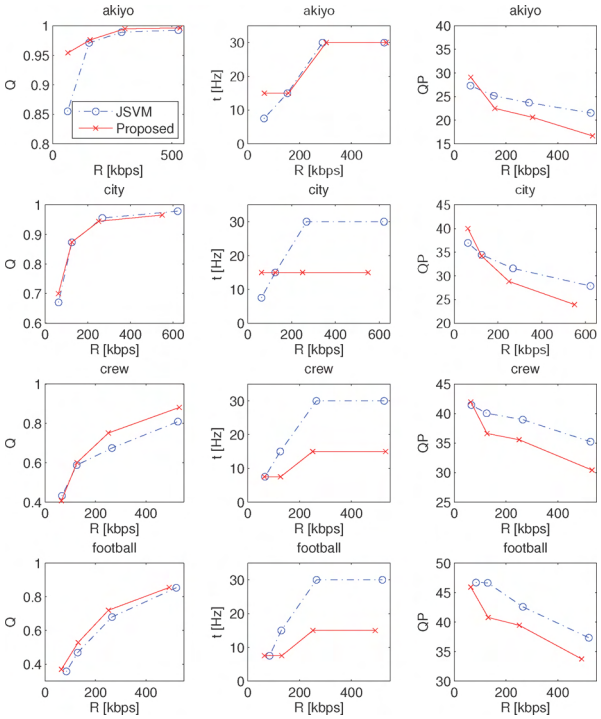

Fig. 5. Quality, frame rate and average QP with respect to the bit rates for different videos.

where optimal frame rate is derived from our proposed models under the bit rate constraint. Compared with the default rate control, where frame rate is configured empirically (usually not accurate), our proposed method provides better perceptual quality. As the future study, we will apply our models to more generic application scenario, where an efficient scene change detection is required, and the model parameters are predicted within separated scenes.

6. REFERENCES

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

[2] Y. Wang, Z. Ma, and Y.-F. Ou, "Modeling Rate and Perceptual Quality of Scalable Video as Functions of Quantization and Frame Rate and Its Application in Scalable Video Adaptation," in *Proc. of PacketVideo*, May 2009.

[3] Y.-F. Ou, Z. Ma, T. Liu, and Y. Wang, "Perceptual Quality Assessment of Video Considering both Frame Rate and Quantization Artifacts," *accepted by IEEE Trans. CSVT*, 2010.

[4] JSVM software, *Joint Scalable Video Model*, Joint Video Team, Doc. JVT-X203, July 2007.