Quality Estimation for H.264/SVC Inter-layer Residual Prediction in Spatial Scalability

Ren-Jie Wang[1], Yan-Ting Jiang[1], Jiunn-Tsair Fang[2], and Pao-Chi Chang[1]

[1] Dept. of Communication Engineering, National Central Univ., Jhongli, Taiwan
[2] Dept. of Electronic Engineering, Ming Chuan Univ., Taoyuan, Taiwan
{rjwang,ytjiang}@vaplab.ce.ncu.edu.tw, fang@mail.mcu.edu.tw,
pcchang@ce.ncu.edu.tw

Abstract. Scalable Video Coding (SVC) provides an efficient compression for the video bitstream equipped with various scalable configurations. H.264 scalable extension (H.264/SVC) is the most recent scalable coding standard. It involves the state-of-the-art inter-layer prediction to provide higher coding efficiency than previous standards. Moreover, the requirements for the video quality on distinct situations like link conditions or video contents are usually different. Therefore, it is very desirable to be able to construct a model so that the target quality can be estimated in advance. This work proposes a Quantization-Distortion (Q-D) model for H.264/SVC spatial scalability, and then we can estimate video quality before the actual encoding is performed. In particular, we further decompose the residual from the inter-layer residual prediction into the previous distortion and Prior-Residual so that the residual can be estimated. In simulations, based on the proposed model, we estimate the actual Q-D curves, and its average accuracy is 88.79%.

Keywords: H.264, Scalable Video Coding, Spatial Scalability, Quality Estimation, Quantization-Distortion Model.

1 Introduction

The fundamental principle of Scalable Video Coding (SVC) is to generate a single compressed bit stream that can adapt to the varying bit rates, display resolutions, and computational resource constraints of various receivers rapidly and easily. There are three kinds of scalability, including temporal, spatial, and quality (SNR) scalability. The spatial scalability that provides various resolutions is suitable for display devices with different sizes nowadays available. In order to remove redundancy between layers, the enhancement layer can be coded using the inter-layer prediction which includes the motion, texture and residual information from the base layer. In H.264/SVC, there exist three kinds of inter-layer prediction tools. There are Inter-Layer Motion Prediction (ILMP), Inter-Layer Intra Prediction (ILIP), and Inter-Layer Residual Prediction (ILRP) [1][2]. ILMP up-samples motion vectors as a motion predictor. ILIP up-samples the reconstructed blocks for the prediction of luminance. Moreover, ILRP up-samples the residual for the residual compensation.

Y.-S. Ho (Ed.): PSIVT 2011, Part II, LNCS 7088, pp. 252–261, 2011.

The requirement for the video quality on distinct situations like link conditions or video content is usually different. Therefore, it is very desirable to be able to construct a Quantization Distortion (Q-D) model so that the target quality can be achieved by selecting a proper encoder Quantization Parameter (QP). Most of the proposed Q-D models were for a single layer video coding [3-7]. In particular, their models were based on the assumption of residual distributions [3-5]. That is, the distortion can be modeled as a function of QP and the variance of the residual distribution. Recently, two Q-D models for SVC spatial scalability and temporal scalability were proposed to perform the optimal rate allocation [8][9]. For the real time application, their parameters of the model are estimated during the encoding procedure.

In this work, we propose a Q-D model for H.264/SVC spatial scalability to estimate video quality. However, the model parameter and the quality score have to be obtained before the entire coding procedure starts. We introduce a residual decomposition technique for ILRP, in which the residual can be decomposed to the coding error and the displacement difference (Prior-Residual). Then the distortion can be modeled as a function of quantization step and Prior-Residual that can be estimated before encoding.

In the remaining of this paper, the analysis of the distortion in the transform domain and related works on Q-D model are discussed in Section 2. The proposed Q-D model and quality estimation for ILRP are described in Section 3. The results for validating the accuracy of proposed model and specifying the model parameters are depicted in Section 4. Finally we summarize our proposed method and results in conclusion.

2 Distortion Analysis and Related Works

2.1 Distortion Analysis in the Transform Domain

Most literatures on Q-D modeling analyze the distortion, specifically the Mean Square Error (MSE) between the original and the reconstructed frames, in the transform domain [3][4]. Two major reasons are that transform coefficients have more similar characteristics than pixels in the spatial domain among various video contents, and the quantization in hybrid video coding, which is the basic structure for most current video coding standards, is performed in the transform domain.

From Fig. 1, we observe that the difference between original frame f_k and reconstructed frame $f_k^{'}(q)$ equals to the difference between residual $r_k(q)$ and quantized residual $r_k^{quan}(q)$ as shown in (1).

$$\begin{aligned} f_k - f_k^{'}(q) &= [f_k - MC\left(f_{k-1}^{'}(q)\right)] - [f_k^{'}(q) - MC\left(f_{k-1}^{'}(q)\right)] \\ &= r_k(q) - r_k^{quan}(q) \end{aligned} \tag{1}$$

Because DCT transform is linear, the equality holds in DCT domain.

$$F_k - F_k^{'}(q) = R_k(q) - R_k^{quan}(q) \tag{2}$$

By Parseval's theorem,

$$E[(f_k - f_k^{'}(q))^2] = E[(F_k - F_k^{'}(q))^2] = E[(R_k(q) - R_k^{quan}(q))^2] \tag{3}$$

we can derive that the MSE between the original and the reconstructed frames equals to the MSE between the original and the quantized residuals in the transform domain. Hence, with the assumption for residual distribution, the distortion is possible to be modeled as a function of QP and parameters of the distribution.

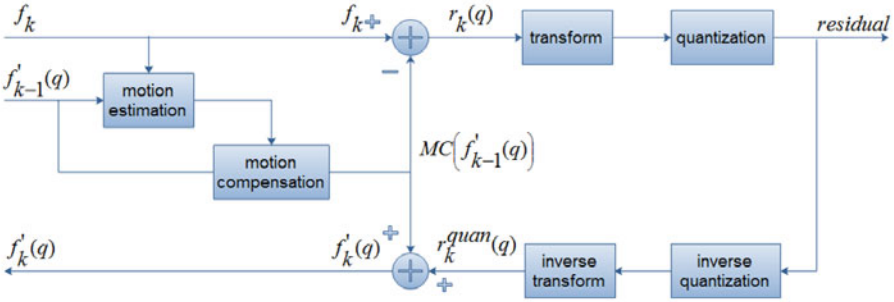


Fig. 1. DPCM based encoder structure

2.2 Laplacian and Cauchy Distributions for DCT Coefficients

The quantized residual can be regarded as a Laplacian-distributed random variable [3]. Then a closed-form expression of distortion is derived. Recently, Kamaci *et al.* [4] proposed Cauchy density function as the residual distribution. Its probability density function (pdf) is shown as

$$p(x) = \frac{1}{\pi}\frac{\mu}{\mu^2 + x^2} \tag{4}$$

where μ is the half-width at half-maximum of the pdf. It basically reflects the variance of the distribution, and can be denoted as a function of σ_x^2, *i.e.*, $\mu = h(\sigma_x^2)$.

The closed-form expression of the distortion is derived and approximated to a power function of q in [4] as

$$\begin{aligned} D(q) &= \sum_{i=-\infty}^{\infty} \int_{(i-\frac{1}{2})q}^{(i+\frac{1}{2})q} (x-iq)^2 p(x) dx \\ &= \sum_{i=-\infty}^{\infty} \int_{(i-\frac{1}{2})q}^{(i+\frac{1}{2})q} (x-iq)^2 \frac{1}{\pi}\frac{h(\sigma_x^2)}{(h(\sigma_x^2))^2 + x^2} dx \approx aq^b = f(\sigma_x^2, q) \end{aligned} \tag{5}$$

where a, b>0, and depend on σ_x^2.

It also demonstrates the Cauchy density is more accurate in estimating the distribution of the DCT coefficients than the traditional Laplacian density. Furthermore, it yields less estimation error for Q-D curve. Therefore, Cauchy distribution is assumed in our work.

2.3 Residual Decomposition for Single Layer

For residual decomposition, Guo *et al.*[10] proposed a quality estimation method for single layer coding. The residual can be decomposed into the displacement difference and the coding error as shown in (6).

$$\begin{aligned} R_k(q) &= F_k - F'_{k-1}(q) \\ &= (F_k - F'_{k-1}) + \left(F'_{k-1} - F'_{k-1}(q)\right) \\ &= I_k + E_{k-1}(q) \end{aligned} \tag{6}$$

The residual $R_k(q)$ is the difference between the original frame F_k and the predicted frame $F'_{k-1}(q)$ that is compensated for by the previous reconstructed frame. On the other hand, F'_{k-1} is the predicted frame that is compensated for by the previous original frame. The residual can be decomposed to the displacement difference I_k , and the coding distortion of the previous frame $E_{k-1}(q)$. Furthermore, with the assumption that both I_k and $E_{k-1}(q)$ have zero mean and are uncorrelated, the variance is also decomposable as (7) shows. That is, the variance $\sigma^2_{R_k}(q)$ is equal to the sum of $\sigma^2_{I_k}(q)$ and $\sigma^2_{E_{k-1}}(q)$.

$$\sigma^2_{R_k}(q) = \sigma^2_{I_k} + \sigma^2_{E_{k-1}}(q) \tag{7}$$

3 Proposed Q-D Estimation Method

In this section, we describe the proposed Q-D estimation method for H.264/SVC inter-layer residual prediction in spatial scalability in detail. With the power form Q-D model and the residual decomposition as basis, we can build up the quality estimation mechanism. The Q-D model for single layer coding that prediction data only come from its own layer is described first. The model can be applied to the base layer or enhancement layers without inter-layer prediction in SVC. Moreover, for SVC inter-layer prediction, the enhancement layer quality or the residual will vary with the similarity between two layers. A Q-D mode for enhancement layers with inter-layer residual prediction is then proposed.

3.1 Q-D Model for Single Layer Coding

As mentioned in [10], with the assumption that a video sequence is a locally temporal stationary process, the corresponding variables in successive frames have the same variance. Thus

$$\begin{aligned}\sigma_R^2(q) &= \sigma_{I_k}^2 + \sigma_{E_{k-1}}^2(q) \\ &= \sigma_I^2 + \sigma_E^2(q) \\ &\overset{\Delta}{=} PR + D(q)\end{aligned} \tag{8}$$

where Prior-Residual, defined as $PR = \sigma_I^2$, is the variance of the displacement difference.

Then, we can put (8) into (5) to obtain (9)

$$\begin{aligned}D(q) &= f\left(\sigma_R^2(q), q\right) \\ &= f(PR + D(q), q)\end{aligned} \tag{9}$$

Because $D(q)$ is a function of PR, and the DCT coefficients are modeled by Cauchy distribution in [4], (9) can be further simplified as (10).

$$D(q) = f^{'}(PR, q) \approx aq^{b} = aq^{cPR^{d}} \tag{10}$$

Where a, c, and d are constants. The specific relationship between b and PR can be built up by empirical tests. We will observe that PR can accurately predict the distortion curve as a good parameter to identify the sequence characteristic.

3.2 Q-D Model for Inter-layer Residual Prediction

The inter-layer residual prediction in SVC is depicted as Fig. 2. Because the high correlation of residual signals between the current and the reference layer, the difference $R_{RP_k}(q)$ between the residuals of two layers instead of residual signal itself is encoded as the enhancement information to improve the coding efficiency.

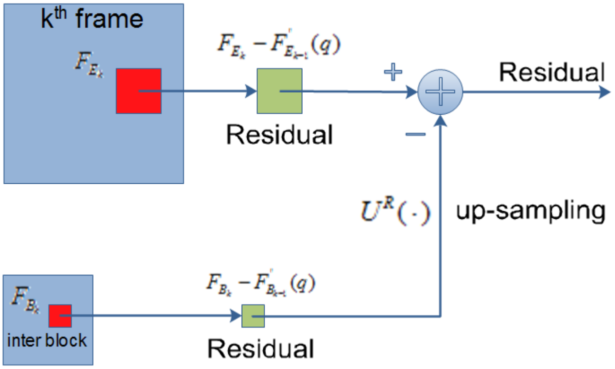


Fig. 2. Inter-layer residual prediction structure in SVC spatial scalability

We also employ the residual decomposition to this structure. By involving the predicted frame by non-distorted data in the enhancement layer $F'_{E_{k-1}}$ and that for the base layer $F'_{B_{k-1}}$, the residual can also be decomposed to the distortion from imperfect prediction and quantization error as in (11).

$$\begin{aligned} R_{RP_k}(q) &= F_{E_k} - F'_{E_{k-1}}(q) - U^R\left(F_{B_k} - F'_{B_{k-1}}(q)\right) \\ &= (F_{E_k} - F'_{E_{k-1}}) + \left(F'_{E_{k-1}} - F'_{E_{k-1}}(q)\right) \\ &\quad - U^R\left((F_{B_k} - F'_{B_{k-1}}) + \left(F'_{B_{k-1}} - F'_{B_{k-1}}(q)\right)\right) \\ &= I_{E_k} + E_{E_{k-1}}(q) - U^R\left(I_{B_k} + E_{B_{k-1}}(q)\right) \\ &= [I_{E_k} - U^R(I_{B_k})] + [E_{E_{k-1}}(q) - U^R\left(E_{B_{k-1}}(q)\right)] \end{aligned} \tag{11}$$

where $U^R(\cdot)$ means the upsampling procedure, which can be implemented by simple bi-linear interpolation or any more sophisticate interpolation operations.

We assume that both $I_{E_k} - U^R(I_{B_k})$ and $E_{E_{k-1}}(q) - U^R\left(E_{B_{k-1}}(q)\right)$ have zero mean and are approximately uncorrelated. In addition, a video sequence is a locally temporal stationary process, *i.e.*, the corresponding variables in consecutive frames have the same variance. (11) can be derived as (12).

$$\begin{aligned} \sigma^2_{R_{RP}}(q) &= \mathrm{var}\left([I_{E_k} - U^R(I_{B_k})] + [E_{E_{k-1}}(q) - U^R\left(E_{B_{k-1}}(q)\right)]\right) \\ &= \mathrm{var}\left([I_{E_k} - U^R(I_{B_k})]\right) + \mathrm{var}\left(E_{E_{k-1}}(q)\right) + \mathrm{var}\left(U^R\left(E_{B_{k-1}}(q)\right)\right) \\ &\quad - 2\rho\sqrt{\mathrm{var}\left(E_{E_{k-1}}(q)\right)}\sqrt{\mathrm{var}\left(U^R\left(E_{B_{k-1}}(q)\right)\right)} \\ &= PR_{RP} + (1 + \beta - 2\rho\sqrt{\beta}) \cdot D_{RP}(q) \end{aligned} \tag{12}$$

where Prior-Residual for Residual Prediction is denoted as $PR_{RP} = \mathrm{var}(I_{E_k} - U^R(I_{B_k}))$. Because of high dependency between two layer distortions, the base layer distortion can be predicted by the enhancement one, which is $\mathrm{var}\left(U^R\left(E_{B_{k-1}}(q)\right)\right) = \beta\,\mathrm{var}\left(E_{E_{k-1}}(q)\right) = \beta \cdot D_{RP}(q)$. For most video sequences, $\beta > 1$ since the higher residual variance exist in the downscaled frame. β and ρ, which is the correlation coefficient between $E_{E_{k-1}}(q)$ and $U^R\left(E_{B_{k-1}}(q)\right)$, vary only slightly with different video contents and hence can be consider as constants. The distortion term $\mathrm{var}\left(E_{E_{k-1}}(q) - U^R\left(E_{B_{k-1}}(q)\right)\right)$ can be expressed as a constant times of $D_{RP}(q)$. Therefore, it is possible to use PR_{RP} to predict the real residual before encoding procedure including Rate Distortion Optimization (RDO), transform, and quantization procedure.

Then, we can put (12) into (5) to obtain (13).

$$\begin{aligned} D_{RP}(q) &= f(\sigma^2_{R_{RP}}(q), q) \\ &= f(PR_{RP} + (1-\beta-2\rho\sqrt{\beta})\cdot D_{RP}(q), q) \end{aligned} \tag{13}$$

Because $D_{RP}(q)$ is a function of PR_{RP}, and the DCT coefficients are modeled by Cauchy distribution, (13) can be further simplified as (14).

$$D_{RP}(q) = f^{'}(PR_{RP}, q) \approx aq^{b} = aq^{cPR_{RP}{}^{d}} \tag{14}$$

where b can also be represented by a power form with PR_{RP}. It will be verified with c and d in the experiment section with real video data. Note that, PR_{RP} means the Prior-Residual for Residual Prediction, which is different from PR for single layer, and it provides more accurate description for Q-D behavior.

Block diagram for obtaining PR_{RP} is shown in Fig.3. Based on the obtained PR, Q-D function for a certain video sequence is established We then can either predict the distortion according a given QP, or select a suitable QP to obtain the video with target visual quality.

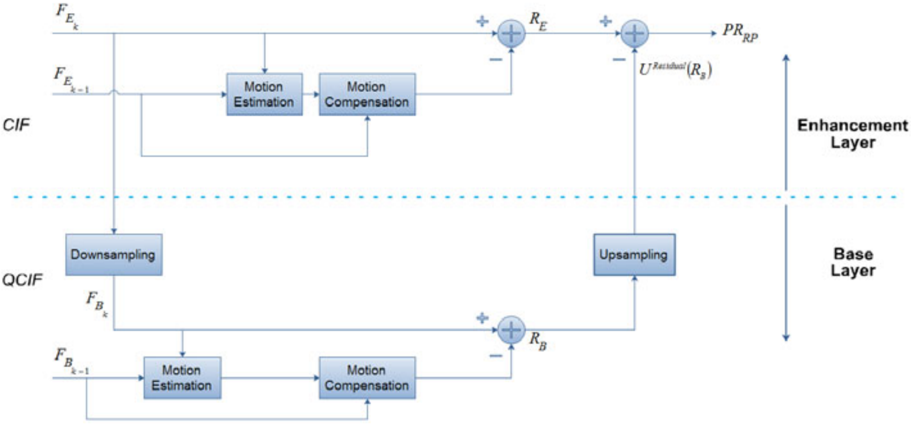


Fig. 3. Prior-Residual in inter-layer residual prediction

4 Experimental Results

In this section we construct an experiment to verify the proposed distortion model for ILRP. We will obtain the model parameters by fitting real coding results in the training phase, and then the performance with those sequences outside the training set will be demonstrated. Experiment setting is the following. Four training video sequences for two layers in CIF and QCIF formats at the frame rate of 30 frames/s, including Akiyo, Carphone, Harbour, Mobile are encoded by H.264/SVC reference software

JSVM 9.19.8. Six QPs (16, 20, 24, 28, 32, 36) are used in the encoding. The same QPs are used for both the base layer and the enhancement layer. We used 90 frames for training, and the first frame is an I-frame while the rest are P-frames. The inter-layer prediction flag was the inter-layer residual prediction (0,0,1). Five test video sequences including Bus, Foreman, Hall, Mother_daughter, and Soccer are encoded, and the rest experiments setting are the same with training process.

As Fig. 4 shows, black dots are the results after the SVC coding for four training sequences. The dotted lines are the approximated curves based on the power form Q-D relationship. We can obtain the specific value *b* that minimizes the estimate distortion for each sequence in the Table 1. Note that we pre-set *a* as a constant to simplified the model. As shown in Table 1, the numerical value of *b* can reflect the behavior that higher distortion or complicated content has a larger *b* at same QP among different video contents.

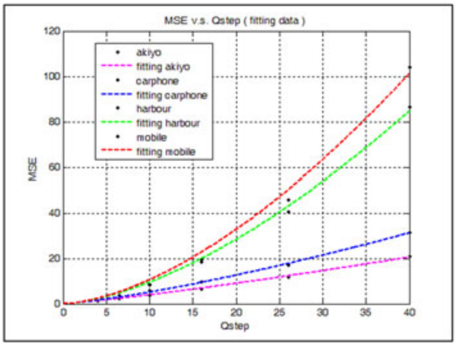
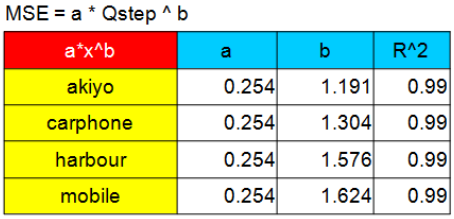


Fig. 4. The training Q-D curve in inter-layer residual prediction

Table 1. The Q-D model parameter in inter-layer residual prediction


MSE = a * Qstep ^ b

a*x^b	a	b	R^2
akiyo	0.254	1.191	0.99
carphone	0.254	1.304	0.99
harbour	0.254	1.576	0.99
mobile	0.254	1.624	0.99

From the experimental results, the Q-D model of inter-layer residual prediction can be precisely specified as the following

$$D(q) = MSE = 0.254 * Qstep^{b} \tag{15}$$

As we derived in the Section 3, the constant *b* is only related to the residual variance or PR. Hence, we observe the relationship between *b* and PR by Fig. 5. From the training data, represented by blue squares in the figure, we can observe that *b* can be modeled as a power function of PR, and *c* and *d* can be determined to be 1.03 and 0.08, respectively. The determination coefficient R^2 in the fitting process is up to 0.97, which implies an excellent fitting.

From the empirical data, the relationship between *b* and PR_{RP} in inter-layer residual prediction is shown as

$$b = 1.037 * PR_{RP}{}^{0.080} \tag{16}$$

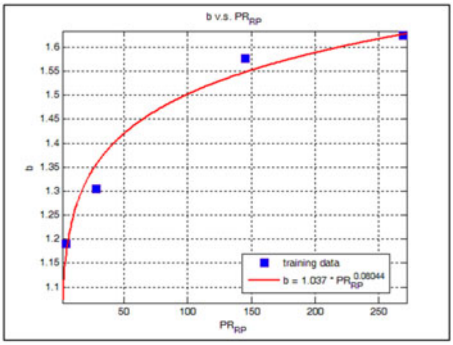


Fig. 5. The fitting curve about b and PR_{RP} in inter-layer residual prediction

Table 2. The relationship between b and PR_{RP} in inter-layer residual prediction

training	b	PR$_{RP}$
akiyo	1.191	4.18
carphone	1.304	28.16
harbour	1.576	145.50
mobile	1.624	269.51

fitting function	R^2
b=1.037*PR$_{RP}^{0.080}$	0.97

Fig.6 and Fig. 7 show the real and the estimated Q-D curves, respectively. It clearly shows that the estimate curves can fit the results obtaining from time-consuming SVC coding. Accuracy of the proposed Q-D model, defined as in (17), for various sequences and Qsteps are listed in Table 3. The average accuracies for all test sequences are more than 81.19%, and up to 93.54% in the sequence Hall. Translated to PSNR, the estimation error is no more than 0.74 dB.

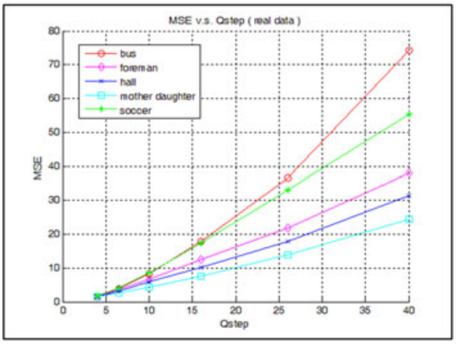


Fig. 6. The encoded Q-D curve in inter-layer residual prediction

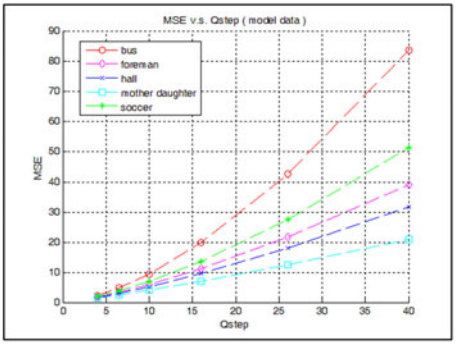


Fig. 7. The modeled Q-D curve in inter-layer residual prediction

Table 3. The accuracy of the Q-D model in inter-layer residual prediction

EL with ILRP cif	Qstep						Average Accuracy
	4	6.5	10	16	26	40	
bus	62.06	76.81	87.16	89.38	84.25	87.49	81.19
foreman	94.85	90.86	87.23	89.55	98.83	97.54	93.14
hall	91.98	91.67	86.09	93.23	98.99	99.31	93.54
mother_daughter	91.73	91.98	91.49	90.87	88.80	85.58	90.07
soccer	88.36	92.38	81.28	78.56	83.30	92.22	86.02

$$\text{Accuracy} = \left(1 - \frac{\left|\text{Actual MSE} - \text{Estimated MSE}\right|}{\text{Actual MSE}}\right) \times 100\% \tag{17}$$

5 Conclusion

We have proposed a Q-D model for inter-layer residual prediction in SVC. The distortion is modeled as a function of quantization step and Prior-Residual that can be efficiently estimated before encoding. Experimental results show that the proposed model can estimate the actual Q-D curves for inter-layer prediction, and the average accuracy of the model is 88.79% in MSE or the estimated error less than 0.74 dB in PSNR, which is suitable for practical use. In the future, we will extend the residual decomposition and Q-D modeling to all inter-layer prediction tools.

References

1. Schwarz, H., Marpe, D., Wiegand, T.: Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. IEEE Trans. Circuits Syst. Video Technol. 17(9), 1103–1120 (2007)
2. Segall, A., Sullivan, G.J.: Spatial Scalability Within the H.264/AVC Scalable Video Coding Extension. IEEE Transactions on Circuits and Systems for Video Technology 17(9), 1121–1135 (2007)
3. Turaga, D.S., Chen, Y., Caviedes, J.: No reference PSNR estimation for compressed pictures. Signal Process. Image Commun. 19, 173–184 (2004)
4. Kamaci, N., Altinbasak, Y., Mersereau, R.M.: Frame bit allocation for the H.264/AVC video coder via Cauchy density-based rate and distortion models. IEEE Trans. Circuits Syst. Video Technol. 15(8), 994–1006 (2005)
5. Berger, T.: Rate-Distortion Theory: A Mathematical Basis for Data Compression. Prentice-Hall, Englewood Cliffs (1971)
6. Takagi, K., Takishima, Y., Nakajima, Y.: A study on rate distortion optimization scheme for JVT coder. In: Proc. SPIE, vol. 5150, pp. 914–923 (2003)
7. Wang, H., Kwong, S.: A rate-distortion optimization algorithm for rate control in H.264. In: Proc. IEEE ICASSP 2007, pp. 1149–1152 (April 2007)
8. Liu, J., Cho, Y., Guo, Z., Kuo, C.C.: Bit Allocation for Spatial Scalability Coding of H.264/SVC With Dependent Rate-Distortion Analysis. IEEE Trans. Circuits Syst. Video Technol. 20(7), 967–981 (2010)
9. Hu, S.H., Wang, H., Kwong, S., Zhao, T., Kuo, C.C.: Rate Control Optimization for Temporal-Layer Scalable Video Coding. IEEE Trans. Circuits Syst. Video Technol. 21(8), 1152–1162 (2011)
10. Guo, L., Au, O.C., Ma, M., Liang, Z., Wong, P.H.W.: A Novel Analytic Quantization-Distortion Model for Hybrid Video Coding. IEEE Trans. Circuits Syst. Video Technol. 19(5), 627–641 (2009)