Brief Papers

Quantization-Distortion Models for Interlayer Predictions in H.264/SVC Spatial Scalability

Ren-Jie Wang, Jiunn-Tsair Fang, Yan-Ting Jiang, and Pao-Chi Chang

***Abstract*—H.264 scalable extension (H.264/SVC) is the current state-of-the-art standard of the scalable video coding. Its interlayer prediction provides higher coding efficiency than previous standards. Since the standard was proposed, several attempts have been made to improve the performance based on its coding structure. Quantization-distortion (Q-D) modeling is a fundamental issue in video coding; therefore, this paper proposes new Q-D models for three interlayer predictions in 264/SVC spatial scalability, that is, interlayer motion prediction, intraprediction, and residual prediction. An existing single layer offline Q-D model is extended to H.264/SVC spatial scalable coding. In the proposed method, the residual power from the interlayer prediction is decomposed into the coding distortion and the prediction distortion. The prediction distortion is the mean square error (MSE) between two original signals that can be obtained by preprocessing with low complexity. Therefore, the coding distortion can be estimated based on both the quantization parameter (QP) and a precalculated prediction distortion before the encoding process. Consequently, the estimated quality based on the proposed models achieved a high accuracy of over 90% for the three interlayer predictions in average.**

***Index Terms*—H.264, quality estimation, quantization-distortion model, scalable video coding, spatial scalability.**

I. Introduction

THE PRINCIPLE of scalable video coding is to generate a compressed bitstream that can be adapted to various bit rates, display resolutions, and computational resource constraints of the platform in the receiver. H.264 scalable extension (H.264/SVC) provides temporal, spatial, and quality (SNR) scalabilities. In particular, spatial scalability, which provides various resolutions suitable for display devices with different sizes, is widely used and receives the most attention. The main coding structure of spatial scalability is the interlayer prediction, in which the enhancement layer is encoded by using the motion, texture, and residual information from the base layer. H.264/SVC provides three interlayer prediction methods: interlayer motion prediction (ILMP), interlayer intra prediction (ILIP), and interlayer residual prediction (ILRP), which uses the information of the motion vector, the reconstructed block, and the residual block from the base layer, respectively [1], [2].

Since H.264/SVC was proposed, several studies have been made on its coding structure, such as traffic and quality evaluation [3], joint rate allocation in video broadcast [4], or quality metric for fully scalable SVC content [5]. Quantization-distortion (Q-D) modeling is an essential research topic in the study of rate control with rate-distortion (R-D) optimization. Most current Q-D models are based on single-layer video coding, in which the distortion is modeled as a function of a quantization step size and the variances of residual coefficients [6]–[11]. Zhang and Comer proposed theoretical models for sub-band and pyramid interlayer prediction structures to analyze R-D performance [12]. The model could be applied to H.264/SVC; however, the comparison of the Q-D curve from the proposed model and that from the real H.264/SVC coding was not shown. Recently, two Q-D models for H.264/SVC spatial scalability and temporal scalability have been proposed to perform optimal rate allocation [13], [14]. However, the parameters of these proposed models, by fitting the previous encoded data, cannot be estimated in advance.

This study proposes Q-D models for the three interlayer predictions in H.264/SVC spatial scalability. The residual power from the interlayer prediction is decomposed into the coding distortion and the prediction distortion. The prediction distortion is defined as the mean square error (MSE) between the original current frame and the original previous frame that can be obtained easily by preprocessing with low complexity. Therefore, both the quantization parameter (QP) and a precalculated prediction distortion can be used to estimate video quality before the encoding process. Based on these proposed models which are the relationship functions between the distortion and the quantization step, the required MSE quality can be achieved by selecting a suitable quantization step.

The remainder of this paper is organized as follows: Section II presents the analysis of the distortion in the transform domain and related works on the Q-D modeling; Section III presents Q-D models for quality estimation in three interlayer predictions; Section IV provides the simulation results for specifying the model parameters and validating the accuracy of the proposed model. Lastly, Section V offers a conclusion.

Manuscript received January 12, 2013; revised November 27, 2013; accepted January 17, 2014. Date of publication March 11, 2014; date of current version June 4, 2014.

R.-J. Wang, Y.-T. Jiang, and P.-C. Chang are with the Department of Communication Engineering in National Cerntral University, Jhong-Li, Taiwan 320. (e-mail: rjwang@vaplab.ce.ncu.edu.tw; ytjiang@vaplab.ce.ncu.edu.tw; pcchang@ce.ncu.edu.tw).

J.-T. Fang is with the Department of Electronic Engineering in Ming Chuan University, Tao-yuan, Taiwan 333, (e-mail: fang@mail.mcu.edu.tw).

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/TBC.2014.2307486

II. Distortion Analysis and Q-D Modeling

Q-D modeling is used to estimate the relationship between quantization and distortion. Frame distortion is generally defined as the MSE between the original frame and the reconstructed frame. However, the distortion is often calculated in the transform domain [6]–[11] because the residual of each frequency sample in the transform domain can be modeled by a specific probability distribution and the quantization procedure is also operated in the transform domain in the most adopted structure of video coding standards.

There are some popular distributions to be modeled for the coefficients of the residual pixels, such as Laplacian distribution [6]–[8], Cauchy distribution [9], or Generalized Gaussian distributions (GGD) [10]–[11]. For the GGD distribution, it consists of many parameters, and its distribution is too complicated for derivation. For the Cauchy distribution, there is no empirical variance in the probability density function (pdf) that is unfavorable to the residual decomposition. We adopt the Laplacian distribution in this work because its pdf has a simple form of variance.

A. Distortion Analysis in the Transform Domain

A general encoding procedure of a hybrid encoder is shown in Fig. 1. The coding error of a block in kth frame is equal to the difference between the original block f_k and the reconstructed block $f^{'}_{k-1}(q)$ with a quantization step size q. The residual block $MC(\cdot)$ is defined as the original block minus the previous reconstructed block $f^{'}_{k-1}(q)$ with the motion compensation, denoted by the operator $MC(\cdot)$ Without loss of generality, we used only one prediction frame for ease of explanation. The coding error can be represented as

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

where r_k^{quan} is the kth residual block after quantization. Eq. (1) shows that the coding error can be represented by the difference between the quantized residual from the residual block. In addition, because the discrete cosine transform (DCT) is linear, we can rewrite (1) as

$$T(r_k(q) - r_k^{quan}(q)) = R_k(q) - R_k^{quan}(q) \tag{2}$$

where T represents the DCT operator, and the capitalized R indicates the block in the transform domain. Usually, we use MSE to represent the quality of video coding, and the equality of (1) and (2) still hold in the MSE measurement. The MSE of coding error can be represented by

$$\begin{aligned} &E[(f_k - f^{'}_k(q))^2] \\ &\quad = E[T(r_k(q) - r_k^{quan}(q))^2] = E[(R_k(q) - R_k^{quan}(q))^2] \end{aligned} \tag{3}$$

where operator E represents the average. Note that the first equality of (3) is applied by the Pareval theorem. Eq. (3) indicates that the MSE between the original and the reconstructed block is equal to the MSE between the original and the quantized residuals in the transform domain. However, from another point of view, if the residual coefficient is assumed to be a random variable with a specific distribution, the distortion can be calculated based on the variance of a specific distribution and a quantization step size.

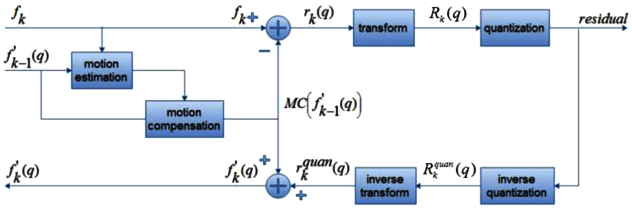

Fig. 1. Structure of a hybrid video encoder. k is frame index; f and r means frame and residual; q is step size; MC means motion compensation.

B. DCT Coefficient Distribution and the Distortion Model

The residual coefficients in the transform domain can be assumed as random variables with a specific distribution. Turaga *et al.* assume that the residual coefficients follow Laplacian distribution [6]. The pdf of a Laplacian distribution is

$$p(y) = \frac{1}{\sigma_y\sqrt{2}} e^{-\sqrt{2}\cdot|y|/\sigma_y} \tag{4}$$

where y is a random variable representing the transform coefficient; σ_y is the standard deviation of y. With these assumptions, a closed form of the coding distortion $D(q)$ can be derived based on the quantization and reconstruction procedure in H.264/AVC [15], as follows:

$$\begin{aligned} D(q) &= \sigma_y^2 - ((1-2\alpha)q + \sqrt{2}\sigma_y) \cdot \frac{q \cdot e^{-\sqrt{2}(1-\alpha)q/\sigma_y}}{1 - e^{-\sqrt{2}q/\sigma_y}} \\ &\triangleq g(\sigma_y^2, q) \end{aligned} \tag{5}$$

where α is the length of the dead zone.α is equal to 1/3 for I-frames or 1/6 for P-frames or B-frames in the reference software JM [16] for H.264/AVC. We use a function operator $g(\cdot)$ with parameters σ_y^2 and q in the right-hand side of (5) to represent this complicated equation.

III. Proposed Q-D Estimation Method

Section II presents a Q-D model for single-layer coding based on the residual coefficient distribution. This section presents a Q-D estimation method for the interlayer predictions in H.264/SVC spatial scalability. Interlayer predictions include motion prediction, residual prediction, and intra prediction. For the quality estimation in H.264/AVC, Guo *et al.* [17] decompose the residual coefficient into the displacement difference and the coding error to separate the distortion effect from the characteristics of video content and coding error separately. We further build a Q-D estimation framework for the three interlayer predictions based on the residual decomposition in this paper.

A. Q-D Model for Interlayer Motion Prediction

Because of the high correlation of motion vectors between the enhancement layer and the reference layer, the up-sampled motion vector from the base layer can be applied to the motion prediction in the enhancement layer. According to the H.264/SVC, the motion vector from the base layer can be selected as a motion predictor or replacement for the motion vector in the enhancement layer. This study focus on the replacement mode, that is, motion copy mode, because the motion

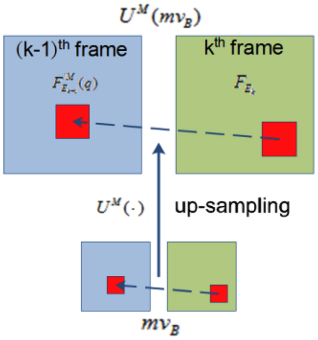

Fig. 2. Inter-layer motion prediction structure in SVC spatial scalability. mv_B: motion vector in base layer; U^M: motion vector upsampling.

copy mode has higher Q-D variation than the predictor mode, compared with the case in the single layer. In the case of predictor mode, since the predictor mode and the single layer coding result in similar motion vectors, the Q-D behaviors of the predictor mode and the single layer coding are similar. Thus, the conventional Q-D model can be used for predictor mode.

The interlayer motion prediction is shown in Fig. 2. The up-sampled motion vector from the base layer is used for the motion prediction in the enhancement layer, where $U^M(\cdot)$ is an operator of up-sampling and mv_B is the motion vector from the base layer. Without loss of generality, we use only one enhancement layer for the presentation purpose. In the interlayer motion prediction, the residual block $R_{M_k}(q)$ in the kth frame is a subtraction of the predicted block $F'^M_{E_{k-1}}(q)$ in the previous reconstructed frame with the up-sampled motion vector (MV) from the corresponding block F_{E_k} in the original frame. The capital letters indicate that subtraction is performed in the transform domain.

The residual block in the kth frame caused by the motion prediction can be expressed by

$$\begin{aligned} R_{M_k}(q) &= F_{E_k} - F'^M_{E_{k-1}}(q) \\ &= (F_{E_k} - F^M_{E_{k-1}}) + \left(F^M_{E_{k-1}} - F'^M_{E_{k-1}}(q)\right) \\ &\triangleq I^M_{E_k} + E^M_{E_{k-1}}(q) \end{aligned} \tag{6}$$

where $F^M_{E_{k-1}}$ is the corresponding block in the previous frame with the motion compensation, but without quantization; $I^M_{E_k}$ is the difference between the original frame and the previous frame with motion compensation, but without quantization, that is the displacement difference; and $E^M_{E_{k-1}}$ is the coding error of the block in the previous frame. The last equation shows that the residual block caused by the interlayer motion prediction can be decomposed into the displacement difference between two consecutive frames and the coding error of the previous frame.

We assume that both $I^M_{E_k}$ and $E^M_{E_{k-1}}(q)$ are zero means and uncorrelated; therefore, the two correspondent variances can be added directly, that is, the variance of residual becomes

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

Furthermore, a video sequence is assumed as a locally temporal stationary process [17], that is, the variance is independent of the frame number k, we have

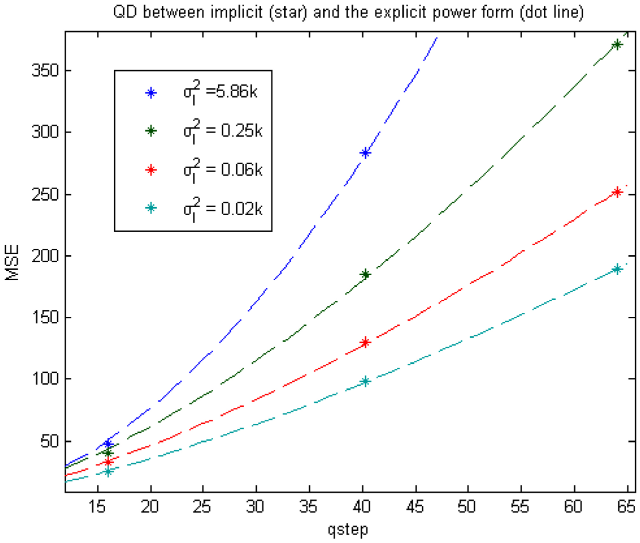

Fig. 3. the Q-D relationship in implicit form (star) and explicit power form (dot line) for various residual variations.

$$\begin{aligned} \sigma^2_R(q) &= \sigma^2_I + \sigma^2_E(q) \\ &\triangleq PR_M + D(q) \end{aligned} \tag{8}$$

We define σ^2_I as the team *Prior-Residual* (PR_M) for interlayer motion prediction because it can predict the characteristics of the residual signals that determine the relationship between quantization and distortion before the encoding procedures, including rate distortion optimization (RDO), transform, and quantization procedures. Finally, (8) is inserted into (5) to obtain the distortion

$$D(q) = g(PR_M + D(q), q) \tag{9}$$

Subsequently, we build a distortion function based on the sequence characteristic and quantization parameter. However, this is an implicit function, which is difficult and time consuming to obtain the solution. Therefore, we use a heuristic method to simulate their relationship between MSE and step size, which is shown in Fig. 3. By plotting samples of MSE and step size according to the implicit function (9) under various prior residuals, we observe that the relationship can be simplified to a power function, that is,

$$D(q) = \hat{g}(PR_M, q) \approx aq^b \tag{10}$$

where a and b depend on PR_M.

In (10), it shows that the relationship between quantization and distortion has a power form. This relation is consistent to the Q-D model derived from the single layer [9]. However, in [9], their parameters are fitted for the residual variance. In (10), the parameters are consistent of constants and a factor PR_M describing the sequences characteristics, and can be estimated before the entire encoding but fitting the residual and resulting RD data during encode.

The distortion caused by the interlayer motion prediction can be predicted by using (10) because PR_M can be calculated in advance. The relationship between a (or b) and PR_M based

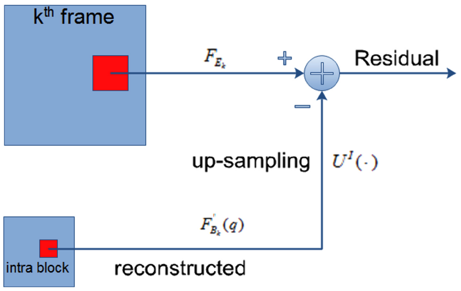

Fig. 4. Inter-layer intra prediction structure in SVC spatial scalability.

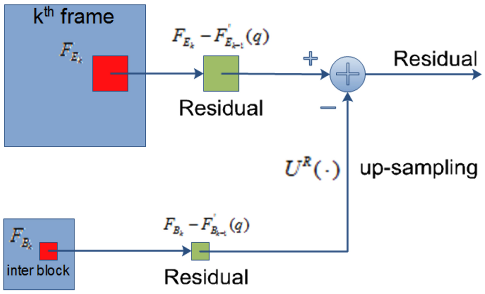

Fig. 5. Inter-layer residual prediction structure in SVC spatial scalability.

on empirical testing is presented in Section IV, and the final Q-D model is

$$D_M(q) = aq^{cPR_M^d} \tag{11}$$

where a, c, and d are constants. Further details are provided in Section IV, which shows that PR_M can accurately predict the distortion curve as a suitable parameter to identify the sequence characteristic.

B. Q-D model for Interlayer Intra Prediction

This sub-section analyzes another inter-layer prediction, that is, interlayer intra prediction. Because of the high correlation of the collocated block between the enhancement layer and the base layer, the reconstructed block in the base layer can be applied to the prediction of intra coding in the enhancement layer. For the kth frame, the residual block $R_{IP_k}(q)$ is the subtraction of the up-sampled prediction of the reconstructed block $F'_{B_k}(q)$ in the base layer from the corresponding block in the enhancement layer, as shown in Fig. 4.

We apply the residual decomposition for the interlayer intra prediction. Subsequently, the residual block $R_{IP_k}(q)$ can be decomposed by

$$\begin{aligned} R_{IP_k}(q) &= F_{E_k} - U^I\left(F'_{B_k}(q)\right) \\ &= F_{E_k} - U^I\left(F_{B_k}\right) + U^I\left(F_{B_k} - F'_{B_k}(q)\right) \\ &= [F_{E_k} - U^I(F_{B_k})] + U^I\left(E_{B_k}(q)\right) \end{aligned} \tag{12}$$

where $U^I(\cdot)$ denotes the up-sampling procedure that can be implemented by a simple bilinear interpolation or other sophisticated interpolation methods, and F_{B_k} is the predicted block from the original frame in the base layer. Eq. (12) shows that the residual caused by the intra prediction can be decomposed into the imperfect prediction from the base layer to the enhancement layer and the coding error from the base layer with an up-sampling operation.

We assume that both $F_{E_k} - U^I(F_{B_k})$ and $U^I(E_{B_k}(q))$ have zero means and are uncorrelated to each other, and that a video sequence is a locally temporal stationary process (12) can further be derived as

$$\begin{aligned} \sigma^2_{R_{IP}}(q) &= \text{var}\left([F_E - U^I(F_B)]\right) + \text{var}\left(U^I\left(E_B(q)\right)\right) \\ &\triangleq PR_{IP} + D_B(q) \end{aligned} \tag{13}$$

where var represents variance. We also define a term PR_{IP} as the prior-residual for interlayer intra prediction.

Because of the high dependency between two layers, the block distortion in the base layer can be predicted by the block distortion in the enhancement layer, that is, $\alpha D_E = D_B(q)$. Parameter α is generally greater than one because a downscaled frame has higher residual variance. According to our observation in the experiments, α varies only slightly with the video content, and can be considered as a constant. Subsequently, (13) can be written as

$$\sigma^2_{R_{IP}}(q) = PR_{IP} + \alpha D_E(q). \tag{14}$$

We inserted (14) into (5) to obtain the distortion of enhancement layer for inter-layer intra prediction

$$D_E(q) = g(\sigma^2_{I_{IP}} + \alpha \cdot D_E(q), q). \tag{15}$$

By using the similar derivations in III-A, we obtain a power form distortion function as

$$D_{IP}(q) = aq^b \tag{16}$$

where both a and b depend on PR_{IP}. The relationship between a (or b) and PR_{IP} can be constructed by empirical tests; the final Q-D model is

$$D_{IP}(q) = aq^{cPR_{IP}^d} \tag{17}$$

Note that, PR_{IP} differs from PR_M for motion prediction, and can provide a superior description for Q-D behavior in the interlayer intra prediction mode.

C. Q-D Model for Interlayer Residual Prediction

The last interlayer prediction analyzed in this study is the residual prediction which contributes most compression gain among three predictions [18]. Because of the high correlation of the residual block between the enhancement and reference layer, the encoding procedure in H.264/SVC subtracts the residual block in the reference layer from the corresponding residual block in the enhancement layer as the residual block for encoding in the enhancement layer. The interlayer residual prediction is shown in Fig. 5.

Similar to the residual decomposition of the two previous interlayer predictions, we apply the residual decomposition to each layer, and combine the displacement differences and coding errors from the enhancement layer and base layer,

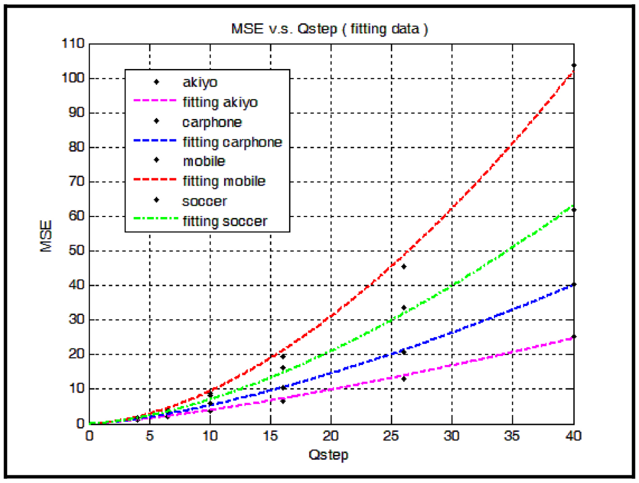

Fig. 6. The trained Q-D curves in inter-layer motion prediction.

respectively. Subsequently, the residual block $R_{RP_k}(q)$ can be decomposed as

$$\begin{aligned} R_{RP_k}(q) &= F_{E_k} - F'_{E_{k-1}}(q) - U^R\left(F_{B_k} - F'_{B_{k-1}}(q)\right) \\ &= (F_{E_k} - F'_{E_{k-1}}) + \left(F'_{E_{k-1}} - F'_{E_{k-1}}(q)\right) \\ &\quad -U^R\left((F_{B_k} - F'_{B_{k-1}}) + \left(F'_{B_{k-1}} - F'_{B_{k-1}}(q)\right)\right) \\ &= [I_{E_k} - U^R(I_{B_k})] + [E_{E_{k-1}}(q) - U^R\left(E_{B_{k-1}}(q)\right)] \end{aligned} \tag{18}$$

where I_{E_k} and I_{B_k} are the displacement difference from the enhancement layer and base layer, respectively. Eq. (18) shows that the residual block for interlayer residual prediction can be decomposed into the difference of imperfect prediction $I_{E_k} - U^R(I_{B_k})$ and the difference of coding error $E_{E_{k-1}}(q) - U^R(E_{B_{k-1}}(q))$ between the enhancement layer and base layer, respectively.

With the assumptions of zero means, uncorrelated, and temporal stationary as the same as previous two sections, the variance of $R_{RP_k}(q)$ can be expressed by

$$\begin{aligned} \sigma^2_{R_{RP}}(q) &= \text{var}\left(I_E - U^R(I_B)\right) + \text{var}\left(E_E(q) - U^R\left(E_B(q)\right)\right) \\ &\triangleq PR_{RP} + D_E + D_B - 2\rho\sqrt{D_E(q)}\sqrt{D_B(q)} \end{aligned} \tag{19}$$

where PR_{RP} is the prior-residual for interlayer residual prediction. Because of high dependency of the contents in the base layer and the enhancement layer, the distortion in the base layer can be predicted by distortion in the enhancement layer, that is, $D_B(q) = \beta D_E(q)$; ρ is the correlation coefficient between D_E and D_B. Based on our observation in the experiments, β and ρ exhibit a slight variance with the video content, and can be considered as constants. The variance of the residual block can be derived as

$$\sigma^2_{R_{RP}}(q) = PR_{RP} + (1 + \beta - 2\rho\sqrt{\beta}) \cdot D_E(q) \tag{20}$$

Subsequently, we inserted (20) into (5) to obtain

$$D_E(q) = g(PR_{RP} + (1 - \beta - 2\rho\sqrt{\beta}) \cdot D_E(q), q) \tag{21}$$

Similar to the derivation in III-A, we obtained a power-form distortion function as

$$D_E(q) = aq^b \tag{22}$$

where a and b depend on PR_{RP}. The specific relationship between a (or b) and PR_{RP} is subsequently constructed by empirical tests, as shown in Section IV, and the Q-D model for residual prediction is derived as

$$D_E(q) = aq^{cPR^d_{RP}} \tag{23}$$

Consequently, the Q-D models for the three interlayer predictions in H.264/SVC spatial scalability are constructed. With the pre-calculated PR for a video sequence in advance, the quality in MSE can subsequently be estimated based on a specified quantization step, or the required MSE quality can be achieved by selecting a suitable quantization step.

IV. Experimental Results

Experiments are performed to find the coefficients to fit the proposed Q-D relationship and subsequently estimate the performance for the three interlayer predictions. The video sequences for the experiments include four training video sequences (Akiyo, Carphone, Harbor, and Mobile), and five test video sequences (Bus, Foreman, Hall, Mother/Daughter, and Soccer). All sequences are encoded at 30 frames per second (FPS) by the reference software JSVM 9.19.8 of H.264/SVC [19]. For spatial scalability, the enhancement layer is in CIF format, and the base layer is in QCIF format. Six different QP values (16, 20, 24, 28, 32, and 36) are used for quantization, but with the same QP values for both layers. Finally, each sequence contains 90 frames, and the first frame is encoded as an I-frame, whereas the remaining frames are P-frames.

A. The Relationship Between the Model Parameter and Sequence Characteristic

For the interlayer motion prediction, the Q-D relationship for the training sequences is shown in Fig. 6, in which the black dots represent the simulation results, and dotted lines represent the fitting curves based on the power-form model derived in Section III.

TABLE I
The Q-D Model Parameters in Inter-Layer Motion Prediction. The Fitness is Represented in R^2

	a	b	R^2
Akiyo	0.185	1.326	0.996
Carphone	0.185	1.458	0.998
Mobile	0.185	1.711	0.998
Soccer	0.185	1.581	0.997

To reduce the fitting complexity, we assume that variable a is constant because the variable a in different sequences is very close to each other. Subsequently, we determine the variable b that minimizes the estimated distortion for each training sequence. The results are listed in Table I, a is obtained by taking the average of the values obtained from the regression process with the two parameters a and b in advance. A larger value of b indicates a higher distortion for complex content, such as Mobile sequence, while a smaller value of b indicates a lower distortion for smooth content, such as Akiyo sequence.

Subsequently, the Q-D model of interlayer motion prediction can be expressed by

$$D(q) = 0.185 \cdot q^b \tag{24}$$

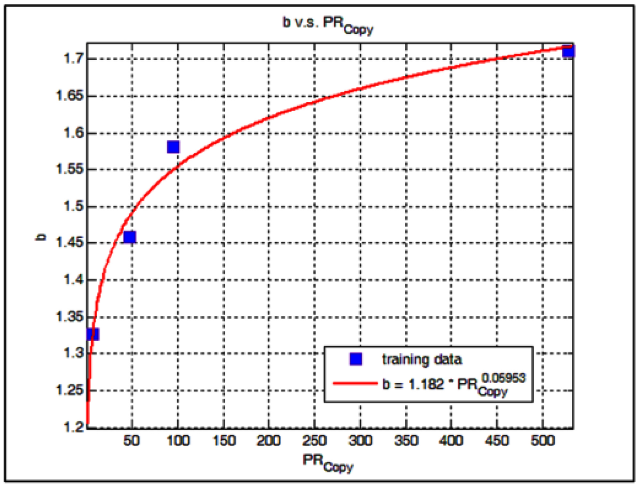

Fig. 7. The fitting curve of b and PR_M in inter-layer motion prediction.

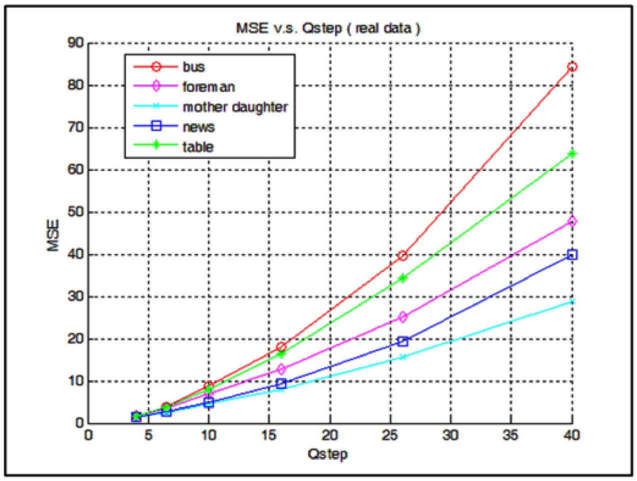

Fig. 8. The encoded Q-D curve in inter-layer motion prediction.

TABLE II
THE Q-D MODELS OF INTER-LAYER MOTION, INTRA, AND RESIDUAL PREDICTION

Inter-layer motion prediction	Inter-layer intra prediction	Inter-layer residual prediction
$D(q) = 0.185 \cdot q^b$	$D(q) = 0.244 \cdot q^b$	$D(q) = 0.254 \cdot q^b$
$b = 1.182 \cdot PR_M^{0.060}$	$b = 0.925 \cdot PR_{IP}^{0.089}$	$b = 1.037 \cdot PR_{RP}^{0.080}$

The relationship between b and PR_M is represented by the squares in Fig. 7. We use a power-form function as (25) to fit the results, in which the model parameters are determined as 1.182 and 0.060, respectively. The determination coefficient R^2 in the fitting process is up to 0.97.

$$b = 1.182 \cdot PR_M^{0.060} \tag{25}$$

For a new sequence, the PR_M is first estimated according to its definition in III-A, which is the displacement difference with the up-sampled motion vector from base layer. Then (25) and (24) are applied to obtain the relationship between the step size and the distortion.

TABLE III
THE ACCURACY OF THE Q-D MODEL IN INTER-LAYER MOTION PREDICTION (ILMP) WITH DIFFERENT QSTEPS

EL ILMP	Qstep						Accuracy
	4	6.5	10	16	26	40	
Bus	88.37	93.02	99.18	94.98	92.61	96.63	94.13
Foreman	91.06	84.47	82.4	91.4	95.21	95.58	90.02
Mother	81.22	88.48	90.97	96.12	93.82	91.35	90.33
News	91.11	93.6	97.17	99.73	95.98	86.28	93.98
Table	98.15	96.73	87.63	90.77	94.12	99.97	94.56

TABLE IV
THE ACCURACY OF THE Q-D MODEL IN INTER-LAYER INTRA PREDICTION (ILIP) WITH DIFFERENT QSTEPS

EL ILIP	Qstep						Accuracy
	4	6.5	10	16	26	40	
Bus	71.97	83.84	98.39	98.97	97.46	97.62	91.38
Foreman	90.43	96.38	94.11	94.34	96.5	98.12	94.98
Mother	92.93	96.37	88.66	94.27	99.63	97.07	94.82
News	87.49	90.66	96.16	92.24	91.32	93.24	91.85
Table	81.54	94.79	89.66	88.85	93.98	97.06	90.98

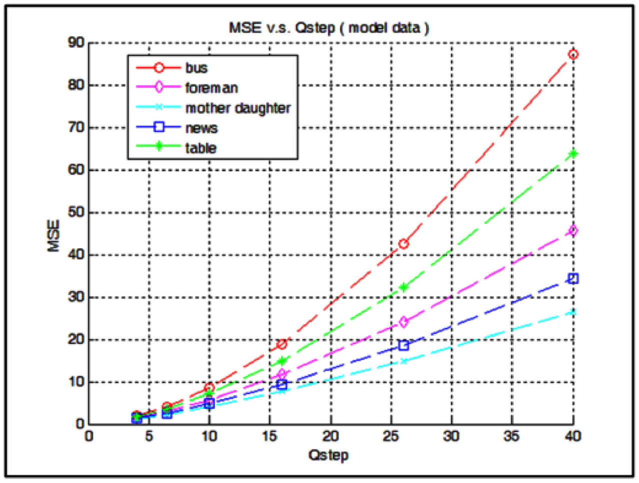

Fig. 9. The modeled Q-D curve in inter-layer motion prediction.

We apply the same procedures for the interlayer intra and residual predictions. Similar to previous analysis, a power-form curve can be applied to fit the simulation data of the Q-D relationship for each training sequence. The final QD models for the three interlayer predictions with sequences characteristic and model parameters are listed in Table II.

B. Accuracy Between Modeled and Actual QD Curves

We arrange a set of test sequences out of training sequences to evaluate the performance of QD estimation. The actual and the estimated Q-D curves are shown in Fig. 8 and 9 respectively. We define the accuracy as

$$\text{Accuracy} = \left(1 - \frac{|\text{Actual MSE} - \text{Estimated MSE}|}{\text{Actual MSE}}\right) \times 100\% \tag{26}$$

TABLE V
THE ACCURACY OF THE Q-D MODEL IN INTER-LAYER RESIDUAL PREDICTION (ILRP) WITH DIFFERENT QSTEPS

EL ILRP	Qstep						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
Mother	91.98	91.67	86.09	93.23	98.99	99.31	93.54
News	91.73	91.98	91.49	90.87	88.8	85.58	90.08
Table	88.37	92.38	81.28	78.56	83.3	92.22	86.02

TABLE VI
COMPUTATION OVERHEAD FOR OBTAINING PRIOR-RESIDUAL WITH N PIXELS IN ONE FRAME. THE OVERHEAD IS REPRESENTED IN TERM OF THE NUMBER OF MULTIPLICATION(*), ADDITION(+), AND SUBTRACTION(−)

Inter-layer motion prediction	Inter-layer intra prediction	Inter-layer residual prediction
MSE: n-1(+), n(-), n(*)	MSE: n-1(+), n(-), n(*)	MSE: n-1(+), n(-), n(*)
Motion vector Upsampling: n/16^2(*)	Bilinear interpolation: 3n(+), 4n(*)	Bilinear interpolation: 3n(+), 4n(*)
		Motion estimation (diamond search, SAD): (n/16^2)*SS SADs 16^2(-), 15^2-1 (+) / SAD

The accuracy for motion prediction under various step sizes q is listed in Table III. It shows that the average accuracy for each sequence is more than 90%, and the best-fit sequence can reach up to 94.56%. The accuracy for the other two prediction methods is listed in Table IV and V respectively. In Table V, it shows that the average accuracy for each sequence is more than 81.19%, and the best-fitting sequence, Hall, can be up to 93.54%. The estimation error is not higher than 0.74 dB under the measurement of Peak Signal-to-Noise Ratio (PSNR).

C. Computation Overhead for Pre-Processing

The complexity of the proposed model is also estimated. The computation overheads for obtaining prior-residual with n pixels in three types of interlayer predictions are listed Table VI. The operations include the calculation of MSE of two frames, motion vector or frame up-sampling, and motion estimation with a block size of 16x16 and a reasonable number of search step (SS) for the residual prediction. Compared with H.264/SVC encoding procedure [1], the low-complexity overhead can be achieved by using the proposed model.

V. CONCLUSION

This paper proposes Q-D models for three interlayer predictions in H.264/SVC spatial scalability. The distortion is modeled as a function of quantization step and prior-residual that can be estimated efficiently before encoding. The experimental results show that a high accuracy of over 90%, in average**,** can be achieved for the three predictions with low-computation overhead for preprocessing, compared with the Q-D curves by real encoding. Combining three models to get an overall distortion model could be a future work. Briefly, the proposed Q-D models can be applied according to the prediction type. However, since the off-line model is applied before the decision of the prediction type, it needs to pre-estimate the type in advance. One of the solutions is to pre-estimate the type from the previous frame because of the high correlation between the previous frame and the current frame.

REFERENCES

[1] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the H.264/AVC standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[2] A. Segall and G. J. Sullivan, “Spatial scalability within the H.264/AVC scalable video coding extension,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1121–1135, Sep. 2007.

[3] G. Van der Auwera, P. T. David, and M. Reisslein, “Traffic and quality characterization of single-layer video streams encoded with the H.264/MPEG-4 advanced video coding standard and scalable video coding extension,” *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 698–718, Sep. 2008.

[4] W. Yao, L.-P. Chau, and S. Rahardja, “Joint rate allocation for statistical multiplexing in video broadcast applications,” *IEEE Trans. Broadcast.*, vol. 58, no. 3, pp. 417–427, Sep. 2012.

[5] H. Sohn, H. Yoo, W. De Neve, C. S. Kim, and Y. M. Ro, “Full-reference video quality metric for fully scalable and mobile SVC content,” *IEEE Trans. Broadcast.*, vol. 56, no. 3, pp. 269–280, Sep. 2010.

[6] D. S. Turaga, Y. Chen, and J. Caviedes, “No reference PSNR estimation for compressed pictures,” *Signal Process. Image Commun.*, vol. 19, no. 2, pp. 173–184, Feb. 2004.

[7] E. Y. Lam and J. W. Goodman, “A Mathematical analysis of the DCT coefficient distributions for images,” *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000.

[8] X. Li, N. Oertel, A. Hutter, and A. Kaup, “Laplace distribution based lagrangian rate distortion optimization for hybrid video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 2, pp. 193–205, Feb. 2009.

[9] N. Kamaci, Y. Altinbasak, and R. M. Mersereau, “Frame bit allocation for the H.264/AVC video coder via Cauchy density-based rate and distortion models,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 994–1006, Aug. 2005.

[10] J. Sun, W. Gao, D. Zhao, and Q. Huang, “Statistical model, analysis and approximation of rate-distortion function in MPEG-4 FGS videos,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 4, pp. 535–539, Apr. 2006.

[11] F. Müller, “Distribution shape of two-dimensional DCT coefficients of natural images,” *Electron. Lett.,* vol. 29, no. 22, pp. 1935–1936, Oct. 1993.

[12] R. Zhang and M. L. Comer, “Rate distortion analysis for spatially scalable video coding,” *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2947–2957, Nov. 2010.

[13] J. Liu, Y. Cho, Z. Guo, and C. C. Kuo, “Bit allocation for spatial scalability coding of H.264/SVC with dependent rate-distortion analysis,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 7, pp. 967–981, Jul. 2010.

[14] S. H. Hu, H. Wang, S. Kwong, T. Zhao, and C. C. Kuo, “Rate control optimization for temporal-layer scalable video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 8, pp. 1152–1162, Aug. 2011.

[15] ISO/IEC ITU-T Rec. H264: Advanced Video Coding for Generic Audiovisual Services, Joint Video Team (JVT) of ISO-IEC MPEG & ITU-T VCEG, Int. Standard, May 2003.

[16] *H.264/AVC Reference Software: Joint Model (JM)*, JVT of ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/16 [Online]. Available: http://iphome.hhi.de/suehring/tml/index.htm

[17] L. Guo, O. C. Au, M. Ma, Z. Liang, and P. H. W. Wong, “A Novel Analytic Quantization-Distortion Model for Hybrid Video Coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 5, pp. 627–641, May 2009.

[18] X. Li, P. Amon, A. Hutter, and A. Kaup, “Performance analysis of inter-layer prediction in scalable video coding extension of H.264/AVC,” *IEEE Trans. Broadcast.*, vol. 57, no. 1, pp. 66–74, Mar. 2011.

[19] Joint Video Team JSVM Reference Software, JVT of ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/16 [Online]. Available: http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.htm