One-Sided ρ-GGD Source Modeling and Rate-Distortion Optimization in Scalable Wavelet Video Coder

Chia-Yang Tsai, *Member, IEEE,* and Hsueh-Ming Hang, *Fellow, IEEE*

***Abstract*—We develop an accurate source model, one-sided ρ-generalized Gaussian distribution (GGD), for approximating the residual signals in scalable wavelet video coding. An efficient piecewise linear expression is suggested to estimate the shape parameter of the one-sided ρ-GGD. We also improve the model accuracy in matching the real data by modifying the ρ parameter estimation formula. Continuing our previous work on developing the motion information gain metric to measure the motion information efficiency, we now incorporate the one-sided ρ-GGD model in the cost function, which is used for deciding the motion vectors and motion estimation mode in scalable wavelet video coding. Compared with the conventional Lagrangian optimization, our simulation results show that the new mode decision method generally improves the peak signal-to-noise ratio performance in the combined signal-to-noise ratio and temporal scalability cases.**

***Index Terms*—Interframe wavelet video, motion information gain (MIG), one-sided ρ-GGD, scalable wavelet video.**

I. Introduction

DUE TO THE growing popularity of broadband network, multimedia communication systems become an important class of network applications. The video content transmits over the wired/wireless networks, such as 3 G/4 G cellular system and WiMAX, may suffer from the bandwidth fluctuation or the receiver capability variation problems. Therefore, a video coder is expected to support several kinds of scalabilities, e.g., transmission bitrate, image resolution, and frame rate scalabilities; they are, in codec terminology, the so-called signal-to-noise ratio (SNR), spatial, and temporal scalabilities, respectively. The current standard scalable video coder, H.264 scalable video coding extension, is a representative of the discrete cosine transform (DCT)-based solution and was standardized in 2007 [1]. On the contrary, the wavelet-based coding scheme also shows its potential and advantages [2] during the MPEG scalable codec standardization competition. The most attractive feature of the wavelet video coding is using only one fully embedded bitstream to satisfy the aforementioned three coding requirements simultaneously. However, to solve the rate-distortion optimization problem for multiple operation points in wavelet coding is a big challenge.

Manuscript received April 7, 2010; revised July 29, 2009; accepted September 23, 2010. Date of publication March 10, 2011; date of current version November 2, 2011. This work was supported in part by the National Science Council, Taiwan, under Grants 96-2221-E-009-063, 97-2221-E-027-044, and 98-2221-E-009-087. This paper was recommended by Associate Editor B. Yan.

The authors are with the Department of Electronics Engineering and the Institute of Electronics, National Chiao Tung University, Hsinchu 300, Taiwan (e-mail: cytsai.ee94g@nctu.edu.tw; hmhang@mail.nctu.edu.tw).

Digital Object Identifier 10.1109/TCSVT.2011.2125530

A. Introduction to Interframe Wavelet Video Coding

Discrete wavelet transform (DWT) has been successfully applied to image compression, e.g., the well-known image coding standard JPEG2000 [3]. DWT decomposes image pixels into 2-D spatial subbands. The multiresolution representation of wavelet transform allows a natural way to provide spatial scalability. The same concept can be extended to decompose video frames. By adopting the motion-compensated temporal filtering (MCTF) technique, the video frames can also have multiresolution representation and thus temporal scalability is realized. The interframe wavelet coding structure that includes MCTF has been explored by Ohm [4], Hsiang and Woods [5], and Secker and Taubman [6]. The most popular coding structure of interframe wavelet video coder is the so-called "t+2-D" structure shown in Fig. 1. The notion of "t+2-D" indicates the encoding operation order: first, the temporal analysis, MCTF, and then the spatial analysis, 2-D DWT, is applied. After both the temporal and spatial analyses are done, the image frames of a group of pictures (GoP) are transformed to several spatio-temporal subbands. In the meanwhile, the motion information is produced by the MCTF process. By reducing the intersubband or intrasubband redundancy, these subbands can be efficiently compressed to one scalable bitstream by a context-based entropy coder, such as EZW [7], SPIHT [8], and EBCOT [9]. After the entropy coding stage in Fig. 1, the coded bitstream consists of two parts, s and v, representing separately the texture subband information and the motion information.

In most existing schemes, only the texture subband bitstream is scalable, and the motion bitstream is non-scalable. In accordance with the application requirements, such as channel bandwidth and device capability, the texture bitstream is truncated and the motion bitstream remains intact. Therefore, in Fig. 1, the output bitstreams of the bitstream extractor consist of $\{s'_0, v\}, \{s'_1, v\} \ldots \{s'_n, v\}$ according to the scalable bitrate requirements $r_0, r_1, \ldots, r_n$, respectively. The truncation mechanism is designed to match the scalable entropy coder.

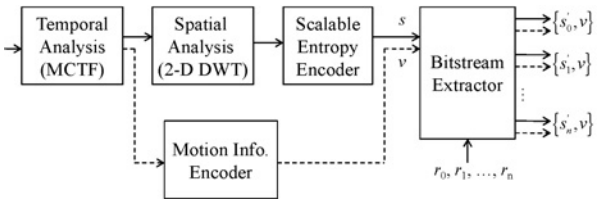

Fig. 1. t + 2-D coding structure of interframe wavelet video coder.

B. Rate-Distortion Optimization Problem in Scalable Wavelet Video Coding

In Fig. 1, the coded output bitstream $\{s, v\}$ is lossless. The original images can be reconstructed when the entire bitstream is received and decoded. Given n scalable requirements, n sub-bitstreams are to be produced by the bitstream extractor. For a well-designed codec, an increased bitrate would lower the coding distortion. An optimal rate-distortion operation point thus exists for each sub-bitstream. In general, a single bitstream $\{s, v\}$ cannot match all the optimal rate-distortion operation points of all sub-bitstreams simultaneously. The bitstream extractor truncates the coded (fixed) bitstream after the coding process is completely done. In other words, it operates outside the coding loop, and this causes the so-called "open-loop" problem [11]. The conventional DCT-based H.264 video coder has a closed-loop coding structure. The motion-compensated prediction errors are controlled by the quantizer inside the loop. The rate-distortion optimization process adjusts parameters targeting at one goal each time. Therefore, the rate-constrained motion estimation is well formulated [12], [13]. However, because of the open-loop coding structure, the scalable wavelet video coder has no feedback path that confines the residual signal quantization errors during encoding. Hence, the motion information bitrate cannot be precisely controlled. Girod [13] showed that (1) warrants the optimal tradeoff between the motion bitrate and texture residual bitrate. That is

$$\frac{\partial D}{\partial R_{rcs}} = \frac{\partial D}{\partial_{R_{mv}}} \tag{1}$$

where D is the distortion, and R_{res} and R_{mv} are the texture residual and the motion bitrates, respectively. However, in our case, the requested multiple bitrates are given after encoding; (1) alone is not sufficient to derive the best motion bitrates for multiple operation points.

The Lagrangian multiplier is a popular and effective method for identifying the optimal motion bitrate (versus the residual texture bitrate). We briefly review its process. The Lagrangian cost function is defined by

$$J = D + \wedge \cdot (R + \Delta R) \tag{2}$$

where ΔR is the motion bitrate, and D and R are the distortion and bitrate of the motion-compensated residual signals, respectively. In a modern hybrid video coding system, many possible prediction modes are available. We choose the best prediction mode by minimizing the Lagrangian cost function in (2). For a fixed ΔR, the Lagrange parameter Δ in (2) can be theoretically derived if the rate-distortion relationship of D and R is known [15]. To control the motion bitrate, ΔR in (2), another Lagrangian cost function is adopted. If the prediction error is measured by the mean of absolute difference (MAD) metric, the rate-constrained motion estimation target is given by

$$\hat{J} = MAD + \hat{\Lambda} \cdot \Delta R \tag{3}$$

where $\hat{\Lambda}$ is chosen empirically as the square root of Δ in (2) [15]. In fact, Λ controls the tradeoff between the prediction error and the motion bitrate. Evidence [15] has shown that the value of Λ strongly depends on the quantization step size. Therefore, to obtain the optimal motion bitrate allocation, Λ should be adjusted at each quantization step size in the encoding process. However, for the scalable wavelet video codec, the encoder performs motion estimation operation without residual quantization information because the quantization step starts after MCTF is done. Hence, the relationship between Λ and quantization step size cannot be exploited in the encoding process. Therefore, the Lagrangian multiplier approach cannot fit into the scalable wavelet video coding structure.

In summary, the optimal motion bitrate allocation problem is a big challenge to the scalable wavelet open-loop coding structure. The conventional Lagrangian optimization process is bitrate dependent. It is hard to adjust it to match the goal of multiple operation points imposed on a single bitstream. To solve the rate-distortion optimization problem of multiple operation points, a new method is needed.

C. Our Previous Papers and Objective of this Paper

In our previous paper [16], we derived a quantitative metric, motion information gain (MIG), for measuring the motion vector efficiency. Based on this metric, we proposed a new cost function for selecting motion vectors and choosing the coding mode. In [17], we found that this metric can also be derived from the entropy definition based on the Laplacian source model. The related encoding parameters are also optimized in [17] by a statistical fitting method. In a separate paper [18], we proposed a ρ-generalized Gaussian distribution (GGD) source model to better approximate the probability distribution of the residual signals (the high-pass spatio-temporal subbands after MCTF). In this paper, these two concepts are integrated into a complete and working algorithm with significant refinements on the proposed process. For example, adaptive schemes for identifying the probability model and the C_0 parameter are designed to match the time-varying characteristics of image sequences. Also, the theoretical foundations of the key parameters are added.

In this paper, we develop a new source model, called as "one-sided ρ-GGD," to approximate the probability distribution of the motion-compensated absolute-valued residual signals. The metric for measuring motion prediction efficiency, so-called MIG in [16] and [17], can now be extended to the more general and accurate one-sided ρ-GGD model for, particularly, the high-pass subbands. In this paper, we also derive a theoretical interpretation of the MIG factor from the entropy viewpoint. Based on the MIG concept, we propose a new cost function to perform rate-distortion optimization, which leads

to a peak signal-to-noise ratio (PSNR) improvement over the conventional Lagrangian optimization method.

This paper is organized as follows. In Section II, the one-sided ρ-GGD source model is investigated for video compression systems. Also, we propose efficient estimation methods for calculating the shape parameter and the ρ parameter with better match. In Section III, we derive the rate-distortion function based on the one-sided ρ-GGD model. The MIG factor is also extended to cover the one-sided ρ-GGD model. Thus, a MIG-based cost function is proposed. The rate-distortion optimization procedure for choosing coding parameters is described in Section IV. Section V shows the experimental results. At the end, a few concluding remarks are given in Section VI.

II. One-Sided ρ-GGD Source Modeling for Motion-Compensated Residual Signal

In the study of motion estimation efficiency, an accurate source model on the motion-compensated residual signal is critical and essential. The results in [18] show that the ρ-GGD source model is more accurate than the Laplacian model. Because we use, typically, a non-negative metric on the prediction errors such as MAD or sum of squared difference, we propose the so-called one-sided ρ-GGD model to approximate the probability distribution of the absolute-valued residual signals. In the modeling process, we propose an efficient linear method to estimate the shape parameter. Furthermore, we increase the modeling accuracy on the real data by proposing an improved ρ value selection method.

A. One-Sided ρ-GGD Function

The probability distribution of the motion-compensated residual signal can be approximated by a zero mean and symmetric probability density function (pdf), and the GGD model is a good example [19]. The GGD pdf is given by

$$P(x) = \frac{1}{2}\left(\frac{\alpha \cdot \eta(\alpha,\sigma)}{\Gamma(\alpha^{-1})}\right)\exp(-[\eta(\alpha,\sigma)\cdot x]^{\alpha}) \tag{4}$$

where

$$\eta(\alpha,\sigma) = \sigma^{-1}\sqrt{\frac{\Gamma(3\alpha-1)}{\Gamma(\alpha^{-1})}} \tag{5}$$

and α is the shape parameter; $\Gamma(\cdot)$ and $\exp(\cdot)$ are the Gamma function and the exponential function, respectively. The σ parameter represents the standard deviation of the residual signal. We now like to approximate the probability distribution of the absolute values of the residual signals. Let the source sample be denoted as $x \in X$, where X is the source alphabet set. Because (4) is a zero-mean and symmetric pdf and X is non-negative, we modify the GGD model to the one-sided GGD with the pdf as follows:

$$P(x) = \left(\frac{\alpha \cdot \eta(\alpha,\sigma)}{\Gamma(\alpha^{-1})}\right)\exp(-[\eta(\alpha,\sigma)\cdot x]^{\alpha}), \quad x \geq 0. \tag{6}$$

The shape parameter α in (6) can be estimated by using the variance and kurtosis of the source signal [19] but the complexity of this approach is very high. We will derive an alternative expression that can be computed from the data samples with much less computation.

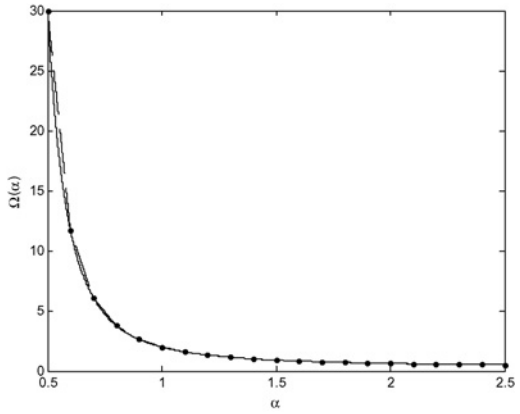

Fig. 2. Solid line and the dashed line are the curves of $\Omega(\alpha)$ and its approximating function $\Omega_e(\alpha)$, respectively. $\Omega_e(\alpha)$ is made of 20 line segments in this example.

We denote the probability of zero in (6) by ρ. That is

$$\rho \triangleq \alpha \cdot \frac{\eta(\alpha,\sigma)}{\Gamma(\alpha^{-1})} = P(0). \tag{7}$$

And then (6) can be rewritten as

$$P_{\rho-GGD}(x) = \rho \cdot \exp\left(-(\rho\alpha^{-1}\Gamma(\alpha^{-1})\cdot x^{\alpha})\right), \quad x \geq 0. \tag{8}$$

We name (8) the one-sided ρ-GGD. There is an interesting property of the proposed one-sided ρ-GGD. From (5) and (7), the product of ρ^2 and σ^2 can be rewritten as

$$\rho^2\sigma^2 = \alpha^2 \cdot \frac{\Gamma(3\alpha^{-1})}{\Gamma(\alpha^{-1})^3}. \tag{9}$$

That is, the product of the square of zero-value probability and the variance is a function of α. We denote this function as

$$\Omega(\alpha) \triangleq \alpha^2 \cdot \frac{\Gamma(3\alpha^{-1})}{\Gamma(\alpha^{-1})^3}. \tag{10}$$

This functional relationship is useful in estimating the shape parameter. As Fig. 2 shows, the mapping between $\Omega(\alpha)$ and α is one-to-one. Therefore, the inverse function of $\Omega(\alpha)$ exists. According to (9) and (10), α can be obtained by

$$\alpha = \Omega^{-1}(\rho^2\sigma^2). \tag{11}$$

Different from the conventional approach, we develop a new and fast method to estimate the shape parameter based on the expression of (11). That is, we use the zero-value probability and the variance value to estimate α.

B. Piecewise Linear Estimation of Shape Parameter

Fig. 2 shows that $\Omega(\alpha)$ is an exponentially decreasing function of the argument α. $\Omega(\alpha)$ can be divided into a number of segments and each segment is approximated by a straight line. The entire range of α is $[\alpha_0, \alpha_n]$. We uniformly partition it into n segments. Thus, $\Omega(\alpha)$ curve is approximated by n pieces of line segments; these line segments are specified by the n

TABLE I
20-SEGMENT SHAPE PARAMETER ESTIMATION TABLE

i	$\Omega(\alpha_{i-1})$	$\Omega(\alpha_i)$	$\Omega(\alpha_i) - \frac{\Omega(\alpha_i)-\Omega(\alpha_{i-1})}{\alpha_i-\alpha_{i-1}}\alpha_i$	$\frac{\Omega(\alpha_i)-\Omega(\alpha_{i-1})}{\alpha_i-\alpha_{i-1}}$
1	30	11.7	121.3	−182.6
2	11.7	6.12	45.49	−56.25
3	6.12	3.8	22.34	−23.18
4	3.8	2.65	13.01	−11.51
5	2.65	2	8.5	−6.5
6	2	1.6	6.023	−4.023
7	1.6	1.33	4.532	−2.667
8	1.33	1.14	3.568	−1.865
9	1.14	1.01	2.911	−1.359
10	1.01	0.91	2.442	−1.024
11	0.91	0.83	2.096	−0.793
12	0.83	0.76	1.833	−0.629
13	0.76	0.71	1.628	−0.508
14	0.71	0.67	1.464	−0.417
15	0.67	0.64	1.332	−0.348
16	0.64	0.61	1.223	−0.293
17	0.61	0.58	1.132	−0.25
18	0.58	0.56	1.056	−0.215
19	0.56	0.54	0.99	−0.187
20	0.54	0.53	0.934	−0.163

sets of boundary points: $\{\Omega(\alpha_0), \Omega(\alpha_1)\}$, $\{\Omega(\alpha_1), \Omega(\alpha_2)\}\ldots$, and $\{\Omega(\alpha_{n-1}), \Omega(\alpha_n)\}$. That is, $\Omega(\alpha)$ is approximated by a piecewise linear function $\Omega_e(\alpha)$. For the ith segment

$$\Omega_e(\alpha) = \frac{\Omega(\alpha_i) - \Omega(\alpha_{i-1})}{\alpha_i - \alpha_{i-1}}(\alpha - \alpha_i) + \Omega(\alpha_i) \tag{12}$$

where $\alpha \in [\alpha_{i-1}, \alpha_i]$. Generally, the approximation is more accurate for large n. Fig. 2 shows the example of $n = 20$, and $\Omega(\alpha)$ is rather accurately approximated by $\Omega_e(\alpha)$ in this case.

The linear function defined by (12) clearly has an inverse. We can thus estimate the shape parameter α_e using (11). If both ρ and σ^2 are known, then

$$\begin{aligned}\alpha_c &= \Omega_c^{-1}(\rho^2\sigma^2) \\ &= \left(\rho^2\sigma^2 - \left(\Omega(\alpha_i) - \frac{\Omega(\alpha_i) - \Omega(\alpha_{i-1})}{\alpha_i - \alpha_{i-1}} \cdot \alpha_i\right)\right) / \\ &\quad \left(\frac{\Omega(\alpha_i) - \Omega(\alpha_i - 1)}{\alpha_i - \alpha_{i-1}}\right)\end{aligned} \tag{13}$$

for

$$\rho^2\sigma^2 \in [\Omega(\alpha_{i-1}), \quad \Omega(\alpha_i)]. \tag{14}$$

One may notice that the coefficients in (13) are independent of data and can thus be calculated in advance and recorded on a table. Table I shows the example of $n = 20$. Therefore, for the ith line segment, the coefficients can be retrieved from Table I, and then the shape parameter can be estimated by using (13).

C. Improved Estimation of ρ Value

In the above discussion, ρ is defined as the zero-value probability of the one-sided ρ-GGD. In the one-sided ρ-GGD model, ρ also represents the highest probability value of the model. However, for some residual image macroblocks (MBs), zero is not the most probable value. In this case, using the zero probability to estimate ρ does not lead to good approximation. Therefore, we modify the ρ estimation formula for this special case.

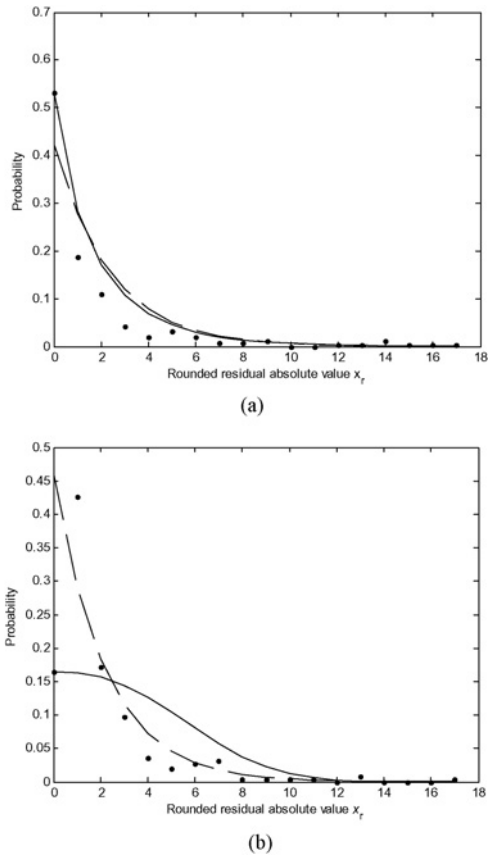

Fig. 3. Dots are the probability distribution of the residual absolute-valued signal, x_r. The dashed line and the solid line show the approximation results by one-sided Laplacian and ρ-GGD modeling, respectively. The ρ value of the ρ-GGD modeling is estimated based on only the zero probability. Two different cases are shown here; the highest probabilities of the distributions are located at (a) $x_r = 0$ and (b) $x_r = 1$, respectively.

Fig. 3 shows two cases. To plot the probability derived from data, the residual absolute-valued signal is rounded to its nearest integer and is denoted by x_r; the probability distribution of x_r and its modeling results are shown in Fig. 3. In the case of Fig. 3(a), the zero probability, $P\{x_r = 0\}$, is the highest probability, and thus the one-sided ρ-GGD can well approximate the data distribution. However, in the case of Fig. 3(b), because $P\{x_r = 0\}$ is not the peak probability, it results in poor approximation. Therefore, we propose a modified estimation formula for ρ. Although the mean of the real residual signal may not be zero, it is not far away from zero based on our collected data. We thus use both the probability of zero, $P\{x_r = 0\}$, and the probability of one, $P\{x_r = 1\}$, to estimate ρ, i.e., ρ is the linear combination of two probabilities as follows:

$$\rho = a \cdot P\{x_r = 0\} + (1 - a) \cdot P\{x_r = 1\} \tag{15}$$

and $0 \leq a \leq 1$. In order to find the optimal a value, we test the following a values, $a \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$, and examine the one-sided ρ-GGD modeling results for each a value.

The a value that leads to the most accurate approximation is chosen to calculate the ρ value. To evaluate the modeling accuracy, we use the Kullback–Leibler (K–L) divergence [20] as the measure, that is

$$KL(p||q) = \sum_{x\in X} P(x)\log_2\left(\frac{p(x)}{q(x)}\right) + \sum_{x\in X} q(x)\log_2\left(\frac{q(x)}{p(x)}\right) \tag{16}$$

where p and q are the "true" and "modeling" probability distribution, respectively. A smaller K–L divergence means more accurate modeling. Therefore, for each residual MB, we can choose the best a value, denoted by a^* as follows:

$$a^* = \arg\min_{a\in A}\{KL(P(x)||P_{\rho-GGD}(x;a))\} \tag{17}$$

where P is the probability distribution of the residual absolute-valued signal; $P_{\rho-GGD}$ is defined by (8) and its ρ value is estimated using (15). Although (17) can be used in the offline analysis, it is impractical in processing real data. We thus develop an efficient method for determining the a^* value.

We separate all events into two cases: $P\{x_r = 0\} > P\{x_r = 1\}$ and the opposite. At each temporal level, we collect the a^* values of all MBs, and separate them into two bins according to the preceding two cases. The probability distributions of a^* of these two cases are shown in Fig. 4. In the case of $P\{x_r = 0\} > P\{x_r = 1\}$, the most probable a^* value is 1 and its probability is over 90%. Therefore, when the first case occurs, a^* is chosen to be 1. Otherwise, 0 is chosen to be the value of a*. In other words

$$a = \begin{cases} 1 & P\{x_r = 0\} > P\{x_r = 1\} \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

In summary, the probability distribution of the residual absolute-valued signal can be approximated by the proposed one-sided ρ-GGD source model by the following steps.

Step 1: Calculate the variance σ^2 from the motion-compensated residual signals.
Step 2: Estimate the ρ value using (15) and (18).
Step 3: Compute the product of ρ^2 and σ^2.
Step 4: Using Table I, we can find the interval $[\Omega(\alpha_{i-1}), \Omega(\alpha_i)]$ that the $\rho^2\sigma^2$ value belongs to.
Step 5: Pick up the ith segment coefficients from Table I. The shape parameter α_e is estimated by using (13).
Step 6: Insert α_e and ρ into (8). The one-sided ρ-GGD modeling is done.

III. MIG for One-Sided ρ-GGD Source Model

In this section, we study the coding efficiency of the motion information when the codec has multiple operation points. Extending our previous work in [16] and [17], we derive the rate-distortion function based on the one-sided ρ-GGD source model. We define the so-called MIG to measure motion efficiency in [16] and [17]. Now in this paper, a similar metric is defined for the one-sided ρ-GGD source model and in fact, it leads to a more general theoretical implication than our previous paper [16], [17]. Based on this metric, we propose a new cost function to perform motion estimation and mode decision procedures.

A. Prediction Efficiency for Rate-Constrained Motion Estimation

We start with the conventional rate-constrained motion estimation case. It is obvious that the motion information rate affects the rate-distortion behavior of the residual signal. Here $D_0(R)$ and $D_{\mathbf{v}}(R)$ denote the rate-distortion functions at rate R for the residual signals produced by zero motion vector and motion vector $\mathbf{v}$, respectively. For a target coding rate R_T, if the predicted frame is directly compensated from the reference frame without motion information, i.e., motion vector is zero, then the maximum available rate allocated to the residual signal is R_T. In this case, we denote the residual signal distortion as $D_0(R_T)$. On the contrary, if the motion vector $\mathbf{v}$ is used in prediction, the maximum available rate allocated to the residual signal is $R_T - \Delta R$, where ΔR is the motion information rate, and the residual signal distortion is denoted as $D_{\mathbf{v}}(R_T - \Delta R)$. Apparently, reducing the available rate of residual signal is worthwhile if its distortion after motion compensation can be reduced. Therefore, an efficient rate-constrained motion compensation case should satisfy the condition as follows:

$$D_{\mathrm{v}}(R_T - \Delta R) < D_0(R_T). \tag{19}$$

B. Rate-Distortion Function of One-Sided ρ-GGD Source Model

Now we will derive the rate-distortion function for the one-sided ρ-GGD source model. The source signal is denoted by $x \in X$ with probability distribution function $P_{\rho-GGD}(x)$ defined by (8). According to the Shannon's rate-distortion theory [10], the Shannon lower bound for the magnitude-error criterion is

$$R_L(D) = \Phi(X) - \log(2eD) \tag{20}$$

where D is the distortion, e is the Euler's number, $\log(\cdot)$ is the natural logarithm function, and $\Phi(X)$ is the differential entropy of X. Based on (66) in the appendix, the differential entropy of the one-sided ρ-GGD source model can be written as

$$\begin{aligned}\Phi(X) &= \rho^{\alpha-1}\left(\rho\alpha^{-1}\Gamma(\alpha^{-1})\right)^{-1}\Gamma(\alpha^{-1})(\alpha^{-1} - \log\rho). \\ &= \alpha^{-1} - \log\rho\end{aligned} \tag{21}$$

where α and ρ are the shape parameter and the zero-value probability of the source model, respectively. Replace $\Phi(X)$ in (20) by (21) as follows:

$$\begin{aligned}R_L(D) &= \alpha^{-1} - \log\rho - \log(2eD) \\ &= -\log(2\rho e^{(1-\alpha^{-1})}).D.\end{aligned} \tag{22}$$

If the conditions given in [10] are satisfied, $R_L(D)$ becomes $R(D)$, the true rate-distortion function, and can be rewritten as follows:

$$D(R) = \frac{e^{-R}}{2\rho e^{(1-\alpha^{-1})}}. \tag{23}$$

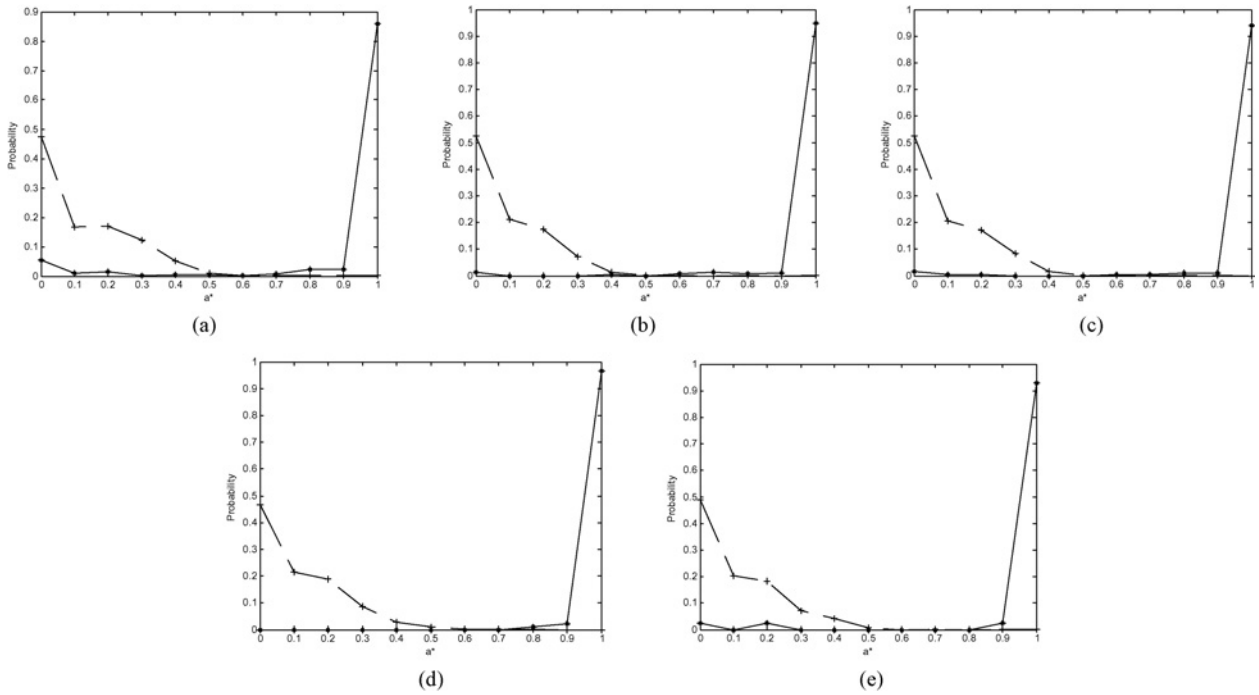

Fig. 4. Solid line and dashed line are the probability distributions of the best a value, denoted by a*, of the following two cases. The first case is (solid) and the second case is the opposite (dashed). The five figures show the results at five temporal levels. (a) $t = 0$. (b) $t = 1$. (c) $t = 2$. (d) $t = 3$. (e) $t = 4$. The test sequence is *Foreman* (CIF, 30 f/s).

TABLE II
AVERAGE FRAME-LEVEL C VALUES USING THE PROPOSED ADAPTIVE SCHEME

Test Sequence	Average C Value
Tempete	7.75
Mobile	7.43
Foreman	7.37
Container	7.99
Waterfall	7.12
Irene	6.43

C. MIG

We now try to find relationship connection between the residual signal statistics and the motion bitrate. As discussed earlier, $\rho_{\mathbf{v}}$ and $\alpha_{\mathbf{v}}$ denote, respectively, the zero-value probability and the shape parameter in one-sided ρ-GGD model of the residual signal using motion vector $\mathbf{v}$. Thus, ρ_0 and α_0 are the residual signal statistics when $\mathbf{v} = \mathbf{0}$. We substitute (23) into (19) with the corresponding parameters, and (19) becomes

$$\frac{e^{-(R-\Delta R)}}{2\rho_{\mathbf{v}} \cdot e^{(1-\alpha_{\mathbf{v}}^{-1})}} < \frac{e^{-R}}{2_{\rho 0} \cdot e(1\alpha_0^{-1})}. \tag{24}$$

Equation (19) can be simplified to (25) as follows:

$$\frac{\alpha_0^{-1} - \alpha_{\mathbf{v}}^{-1} + \log(\rho_{\mathbf{v}}/\rho_0)}{\Delta R} > 1. \tag{25}$$

Interestingly, the target coding rate term, R_T, in (19) is eliminated. This elimination implies that (25) is a rate-independent criterion for checking the motion prediction efficiency. Therefore, in theory, this criterion is applicable in the multiple operation rate situations, such as scalable interframe wavelet coding. However, this criterion needs to be adjusted to match the real video data.

We can examine (25) from a different perspective. Let the residual signal produced by using motion vector $\mathbf{v}$ be $x \in X_{\mathbf{v}}$. Similar to the derivation of (21), the differential entropies of X_0 and $X_{\mathbf{v}}$ are expressed, respectively, as

$$\begin{aligned} \Phi(X_0) &= \alpha_0^{-1} - \log \rho^0 \\ \Phi(X_{\mathbf{v}}) &= \alpha_{\mathbf{v}}^{-1} - \log \rho_{\mathbf{v}}. \end{aligned} \tag{26}$$

If motion vector $\mathbf{v}$ results in good motion compensation, the differential entropy of the residual signal should be smaller than that obtained by using the zero motion vector. The positive difference of the differential entropies of X_0 and $X_{\mathbf{v}}$ is as follows:

$$\begin{aligned} \Delta\Phi(X\mathbf{v}) &= \Phi(X_0) - \Phi(X\mathbf{v}) \\ &= \alpha_0^{-1} - \alpha_{\mathbf{v}}^{-1} + \log(\rho_{\mathbf{v}}/\rho_0). \end{aligned} \tag{27}$$

We can find that (27) is exactly the numerator of the left term in (25). Thus, (25) is reduced to

$$\frac{\Delta\Phi}{\Delta R} > 1. \tag{28}$$

This formula, (28), is equivalent to the efficiency criterion (19). Furthermore, from the entropy viewpoint, the left term of (28) has the interpretation as follows:

$$\frac{\Delta\Phi}{\Delta R} \sim \frac{\text{Decrease of residual signal information}}{\text{Increase of motion information}}. \tag{29}$$

Ideally, the optimal motion compensation method can reduce the maximum residual signal by the minimum motion rate. Therefore, $\Delta\Phi$ and ΔR represents a "reward" and "cost" relationship during the motion estimation process. Also, (29)

can also be regarded as normalizing $\Delta\Phi$ by ΔR, i.e., the residual signal entropy reduction per motion bit. Hence, the ratio of $\Delta\Phi$ to ΔR can be viewed as a gain factor of motion information. Therefore, we define the ratio of $\Delta\Phi$ to ΔR as MIG, denoted by φ, that is

$$\varphi \triangleq \frac{\Delta\Phi}{\Delta R}. \tag{30}$$

In [17], a similar conclusion was obtained based on the Laplacian source assumption. Now, we show that the MIG definition is also valid for the higher dimensional cases such as one-sided ρ-GGD source model.

An efficient rate-constrained motion compensation case should satisfy (19). In our previous discussions, by replacing the distortion term in (19) with the rate-distortion function in (23), we derive the MIG lower bound, which is 1, in (28). However, this result does not match the real-world situation due to at least two factors: one is that a practical coder cannot achieve the rate-distortion bound predicted by the information theory and the other factor is that the real video data do not completely satisfy the mathematical assumptions in theory such as stationarity and probability distribution. Thus, the theoretically derived rate-distortion function may not accurately represent the relationship between the produced coding rate and the real distortion. Therefore, we modified (28) as follows:

$$\frac{\Delta\Phi}{\Delta R} > C \tag{31}$$

where C is the MIG lower bound in the real world. Due to this divergence problem, C is not 1 for a practical wavelet coder applied to the test video data. Therefore, two parameters are introduced and inserted into (19) to reflect the model divergence problem. We rewrite (19) as

$$D_{real,\mathbf{v}}(R_T - \Delta R) < D_{real,\mathbf{0}}(R_T) \tag{32}$$

where $D_{real,\mathbf{v}}$ is the "real distortion" measured from the quantized residual signal compensated using motion vector $\mathbf{v}$, $D_{real,\mathbf{v}}$ is the "ideal distortion" derived from the rate-distortion function of the source model in (19), and a new parameter $\beta_{\mathbf{v}}$ is introduced to compensate for the difference between $D_{real,\mathbf{v}}$ and $D_{real,\mathbf{v}}$. In other words, $\beta_{\mathbf{v}} D_{ideal,\mathbf{v}} = D_{real.\mathbf{v}}$ or

$$\beta_{\mathbf{v}} = \frac{D_{real,\mathbf{v}}}{D_{ideal,\mathbf{v}}}. \tag{33}$$

Here, we assume that a (nearly) constant multiplication factor is adequate for compensating the model divergence. Since this factor is introduced to bridge the gap between the ideal case and the real-world case, it is to be verified by the test data. Then, $D_{real,\mathbf{0}}$ $D_{ideal,\mathbf{0}}$ and $\beta_{\mathbf{0}}$ are similarly defined for using the $\mathbf{0}$ motion vector. Hence, (32) can be rewritten as

$$\beta_{\mathbf{v}} \cdot D_{ideal,\mathbf{v}}(R_T - \Delta R) < \beta_{\mathbf{0}} \cdot D_{ideal,\mathbf{0}}(R_T). \tag{34}$$

By replacing $D_{real,\mathbf{v}}$ by the rate-distortion function in (23), (34) gives

$$\frac{\Delta\Phi}{\Delta R} > 1 + \frac{\log_2(\beta_{\mathbf{v}}/\beta_{\mathbf{0}})}{\Delta R}. \tag{35}$$

Equation (35) is very similar to (28). In the ideal case, the "ideal distortion" would be equal to the "real distortion," which makes $\beta_{\mathbf{v}} = 1$ and $\beta_{\mathbf{0}} = 1$ and (35) would fall back to (28). Therefore, for the real case, the MIG lower bound C becomes

$$C = 1 + \frac{\log_2(\beta_{\mathbf{v}}/\beta_{\mathbf{0}})}{\Delta R}. \tag{36}$$

Let $X^*_{\mathbf{v}}$ denote the quantized residual signal. According to (23), $D_{real,\mathbf{v}}$ is calculated by

$$D_{ideal,\mathbf{v}} = \frac{2^{-H(X'_{\mathbf{v}})}}{2\rho_{\mathbf{v}} e^{(1-\alpha_{\mathbf{v}}^{-1})}} \tag{37}$$

where $H(X'_{\mathbf{v}})$ is the entropy of the quantized residual signal. Using (33) and (37), (36) can be rewritten as

$$C = 1 + \frac{1}{\Delta R}\left(\alpha_0^{-1} - \alpha_v^{-1} + \log\left(\frac{\rho_v}{\rho_0}\right) - H(X'_{\mathbf{0}}) + H(X'_{\mathbf{v}}) + \log_2\left(\frac{D_{real,\mathbf{v}}}{D_{real,\mathbf{0}}}\right)\right). \tag{38}$$

Based on (38), the C value can be found using statistical analysis. How to obtain the quantized residual signal $X'_{\mathbf{v}}$ and $X'_{\mathbf{0}}$ is an issue. The scalable encoder does not have the bitstream extraction condition at the MCTF stage. Due to this reason, it becomes very tricky to select a quantization step size to generate $X'_{\mathbf{v}}$ and $X'_{\mathbf{0}}$. However, the purpose of generating the quantized residual signal is to simulate the divergence problem of the rate-distortion function. We conjecture that there exists a certain range of the quantization step sizes that are representative. Therefore, we take an engineering solution to find a proper quantization step size for deriving the C value. We ran exhaustive experiments for all sequences and found that 8 is generally a good quantization step size for estimating C in (38).

Therefore, we design an adaptive C-value updating scheme. In our scheme, there are two levels in the C value adaptation: frame level and GoP level. In the frame level, we collect the statistics of the MBs with nonzero motion vector and calculate the frame-level C value using (38). This new C value is then used for the next frame. If the encoding frame is the last frame of the GoP, the GoP-level C value is updated by averaging all frame-level C values in that GoP. Then, we explain the connection between the frame-level and the GoP-level adaptations. The newly derived frame-level C value is limited to the range of $[C_{GoP} - \Delta C.C_{GoP} + \Delta C]$, where C_{GoP} is the current GoP-level C value and is used to prevent from the extreme values due to noise or insufficient data in the adaptation process. Also, the GoP-level C value is also limited in the same rage in the adaptation process. For example, if the newly derived GoP-level C value is larger than the previous C_{GoP} plus ΔC, the new GoP-level C value is set to $C_{GoP}+\Delta C$. In our experiments, ΔC is chosen to be 0.5 empirically.

Table II shows the average frame-level C values using this adaptive approach. We can see that the average C value is around 7, which is consistent with our previous finding [17]—in the range of [4, 10]. The proposed adaptive scheme verifies that our previously used offline-trained C value is adequate. Now we compare the rate-distortion performance of the adaptive C scheme and fixed C scheme. We pick up four common intermediate format (CIF) test sequences: *Mobile*, *Container*, *Waterfall*, and *Irene*. The test bitrate points are

TABLE III
AVERAGE PSNR RESULTS OF TWO DIFFERENT C VALUE SCHEME

Test Sequence	Offline-Trained C Value	Adaptive C Value
Mobile	33.625	33.631
Container	45.347	45.351
Waterfall	41.038	41.046
Irene	41.441	41.461

256 kb/s, 384 kb/s, 512 kb/s, 800 kb/s, 1024 kb/s, 1200 kb/s, and 1500 kb/s. The average PSNR results of seven test points of these two schemes are shown in Table III. As Table III shows, their PSNR performances are very similar. However, from (38), we can see that the adaptive scheme requires a lot of additional encoding operations. In the experiment section of this paper, the results are obtained using the offline-trained C value, which is 7, and it still outperforms the conventional Lagrangian method.

D. Proposed MIG Cost Function

Let us assume $\varphi \geq C$, C is the target lower bound of φ. If the φ value produced by a motion vector (MV) is smaller than C, this MV is not cost-effective. We substitute (27) for φ. Then $\varphi \geq C$ becomes

$$\begin{aligned} &\left(\alpha_0^{-1} - \log \rho_0\right) \geq \left(\alpha_{\mathbf{v}}^{-1} - \log \rho_{\mathbf{v}}\right) + C \cdot \Delta R \\ \Rightarrow\ &2 \cdot \log\left(e^{\alpha_0^{-1}}/\rho_0\right) \geq 2 \cdot \log\left(e^{\alpha_{\mathbf{v}}^{-1} + C \cdot \Delta R}/\rho_{\mathbf{v}}\right) \\ &\Rightarrow \frac{e^{2/\alpha_0}}{{\rho_0}^2} \geq \frac{e^{2/\alpha_{\mathbf{v}}}}{{\rho_{\mathbf{v}}}^2} \cdot e^{2 \cdot C \cdot \Delta R}. \end{aligned} \tag{39}$$

When an MV produces a smaller right-side term in (39), it leads to a larger φ. Hence, we look for the best MV that achieves the minimum right term value in (39). Also, when ΔR equals to zero, the right term reaches its maximum value and there is no singular problem. Therefore, for source signal $\mathbf{s}$ and motion vector $\mathbf{v}$, the proposed MIG cost function is defined as

$$J(\mathbf{s}, \mathbf{v}|C) = \frac{e^{2/\alpha_{\mathbf{s}}}}{{\rho_{\mathbf{s}}}^2} \cdot \mathbf{e}^{2 \cdot C \cdot \Delta R(\mathbf{v})} \tag{40}$$

where α_{s} and ρ_{s} are the shape parameter and zero-value probability of the source signal $\mathbf{s}$ and $\Delta R(\mathbf{v})$ is the MV bitrate. On the contrary, from (9) and (10), we have

$${\rho_s}^2 {\sigma_s}^2 = \Omega(\alpha_s) \tag{41}$$

where $\sigma_{\mathbf{s}}^2$ is the residual signal variance. Hence, (40) can be rewritten as

$$J(\mathbf{s}, \mathbf{v}|C) = \frac{e^{2/\alpha_{\mathbf{s}}}}{\Omega\alpha_{\mathbf{s}}} \cdot {\sigma_{\mathbf{s}}}^2 \cdot e^{2 \cdot C \cdot \Delta R(\mathbf{v})}. \tag{42}$$

Let us define a new weighting function $\tau(\alpha)$ as

$$\tau(\alpha) = \frac{e^{2/\alpha}}{\Omega(\alpha)} \tag{43}$$

and thus

$$J(s, \mathbf{v}|C) = \tau(\alpha_{\mathbf{s}}) \cdot {\sigma_{\mathbf{s}}}^2 \cdot e^{2 \cdot C \cdot \Delta R(\mathbf{v})}. \tag{44}$$

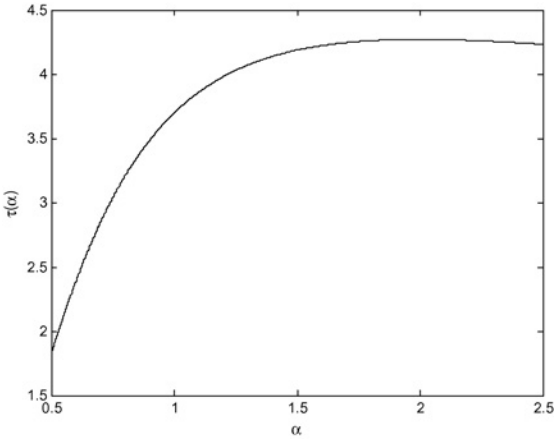

Fig. 5. Cost weighting function $\tau(\alpha)$ for $\alpha \in [0.5, 2.5]$.

The function values of $\tau(\alpha)$ are shown in Fig. 5. It increases as α increases but saturates at about $\alpha = 2$.

In the preceding discussions, the entropy function value is in the unit of "nat." In practice, "bit" is the most common unit used for sending digital data. If the motion rate, $\Delta R(\mathbf{v})$, is measured in "bit," (44) has another equivalent form as follows:

$$J(\mathbf{s}, \mathbf{v}|C) = \tau(\alpha_{\mathbf{s}}) \cdot {\sigma_{\mathbf{s}}}^2 \cdot 2^{2 \cdot C \cdot \Delta R(\mathbf{v})}. \tag{45}$$

In the case of Laplacian source model in [16], (45) is reduced to

$$J_{\mathrm{Laplacian}}(\mathbf{s}, \mathbf{v}|C) = {\sigma_{\mathbf{s}}}^2 \cdot 2^{2 \cdot C \cdot \Delta R(\mathbf{v})}. \tag{46}$$

The difference between (45) and (46) is $\tau(\alpha_{\mathbf{s}})$. It represents the impact of the pdf shape parameter on the MIG cost function. If the residual signals cluster around the zero value, which implies effective motion compensation and the shape parameter, α, in the one-sided ρ-GGD model becomes small. As Fig. 5 shows, when α is small, so is $\tau(\alpha)$. Thus, the proposed MIG cost function in the form of (45) provides a richer interpretation, which links to the pdf shape.

The temporal wavelet decomposition has a tree structure, and, therefore, the coding error propagates along the inversed motion vector direction during decoding. Let us look at an MCTF example at temporal level t. After motion estimation, the high-pass and low-pass frames are generated by using the chosen motion vectors. The low-pass frames will be used in the next stage $(t+1)$ temporal decomposition. At the decoder side, the temporal frames are synthesized by the quantized high-pass and low-pass frames along the inversed motion vector direction. Obviously, the quantization error at temporal level $t+1$ propagates to temporal level t in the synthesis process. Therefore, in the same GoP, the error propagates from the bottom to the top (in the MCTF tree) and thus affects the MCTF coding performance (image quality). This phenomenon is called the "quantization noise propagation" problem. In [14], Wang and van der Schaar proposed an analytic model based on the Lagrangian multiplier method to model this phenomenon for the single bitrate case. However, this technique is hard to extend and apply to the multi-bitrate operating case, especially

under the open-loop coding structure. Therefore, we take a different but feasible approach.

In (45), C controls the tradeoff between the residual signal and the motion information. By adjusting C at different temporal levels, the quality loss due to error propagation can be compensated. To emphasize different C value at level t, we denote it as C_t. Therefore, (45) is rewritten as

$$J(\mathbf{s}, \mathbf{v}|C_t) = \tau(\alpha_{\mathbf{s}}) \cdot {\sigma_{\mathbf{s}}}^2 \cdot 2^{2 \cdot C_t \cdot \Delta R(\mathbf{v})}. \tag{47}$$

Because the decoding bitrate is not pre-specified at the encoding time, it is very difficult to solve this problem at the encoder side. To solve this problem, the rate-distortion behavior at the decoder side has to be considered. Because the synthesis gain is used to allocate the bitrate among different subbands so that the overall distortion can be minimized [22], C_t in (47) is highly related to the so-called synthesis gain. Let gL denote the synthesis gain of the temporal low-pass frame. If the high-pass frame is losslessly decoded, the mean-squared distortion after the inverse MCTF is a function of gL times the mean-squared distortion of the temporal low-pass frame. Following the spirit of [14], because the MIG definition consists of the magnitude-error, we conjecture that the same relationship between the MIG values of different temporal levels would exist. Therefore, at temporal level t, (3) is modified to

$$\frac{\Delta\Phi}{\Delta R} \cdot (\sqrt{gL})^t > C_0 \tag{48}$$

where C_0 is the target MIG lower bound at the first temporal level ($t = 0$). Or, (48) can be rewritten to an equivalent form as follows:

$$\frac{\Delta\Phi}{\Delta R} > C_t \tag{49}$$

where

$$C_t = \left(\frac{1}{\sqrt{gL}}\right)^t \cdot C_0 = w^t C_0. \tag{50}$$

For example, if the 5/3 wavelet filter is used for temporal decomposition

$$gL = (0.5)^2 + (1)^2 + (0.5)^2 = 1.5. \tag{51}$$

Thus, $\omega = 1/\sqrt{1.5} = 0.817$. This theoretically derived ω value is consistent with the finding in our previous work: ω value generally falls in the range of [0.6, 0.9]. In the experiment section of this paper, the results are obtained using the offline-trained ω value in [17], which is 0.8. In summary, (47) is now the cost function used for both motion estimation and mode decision. Their detailed steps are described in the next section.

IV. Prediction Mode Decision Procedure

In the previous section, we propose an MIG cost function which is nearly bitrate independent. It is the target function in our multi-operation-point optimization procedure. The inter-prediction process in a scalable wavelet video codec is very similar to that in H.264/AVC. We take the well-known scalable wavelet codec, Vidwav [23], as an example. The basic prediction unit is MB. Its motion compensation mode consists of a MB partition. The sub-block size can be 16×16, 16×8, 8×16, 8×8, 8×4, 4×8, and 4×4 for a MB in the Vidwav coder. Therefore, for mode m, there are N_m sub-blocks in a MB. The motion-compensated MB residuals and the associated motion vectors can be expressed by two N_m-tuple vectors as follows:

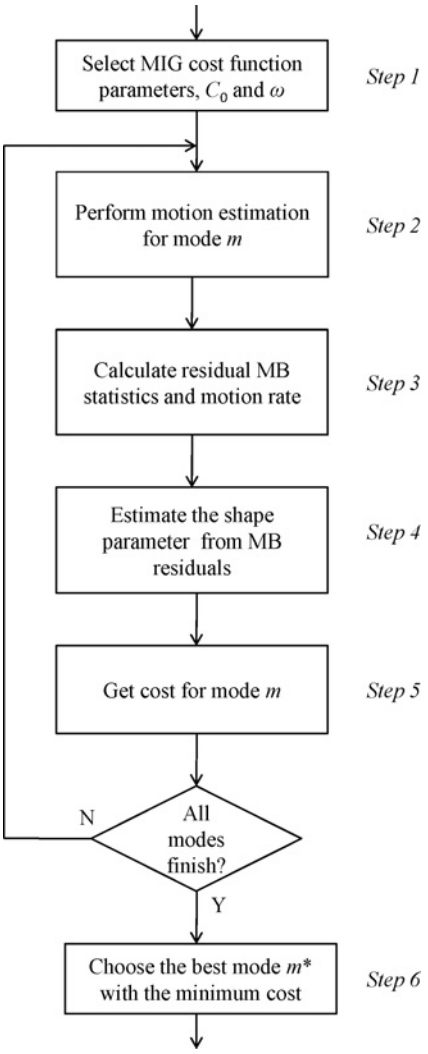

Fig. 6. Flow chart of the proposed mode decision procedure.

$$\begin{aligned} \mathbf{b}_m &= (b_1, \cdots, b_{N_m}) \\ \mathbf{v}_m &= (v_1, \cdots, v_{N_m}) \end{aligned} \tag{52}$$

where b_i and v_i represent the ith sub-block residual signal and its MV, respectively. Assume $\mathbf{M}$ is the mode candidate set, i.e., $m \in \mathbf{M}$. As Fig. 6 shows, there are six steps in deciding the best prediction mode.

Step 1: *Select the MIG cost function parameters:* The proposed MIG cost function (47) contains one parameter, C_t. According to (50), C_t is further split to two parameters, C_0 and ω. As discussed earlier, we empirically choose C_0 and ω from the range of [4, 10] and [0.6, 0.9], respectively.

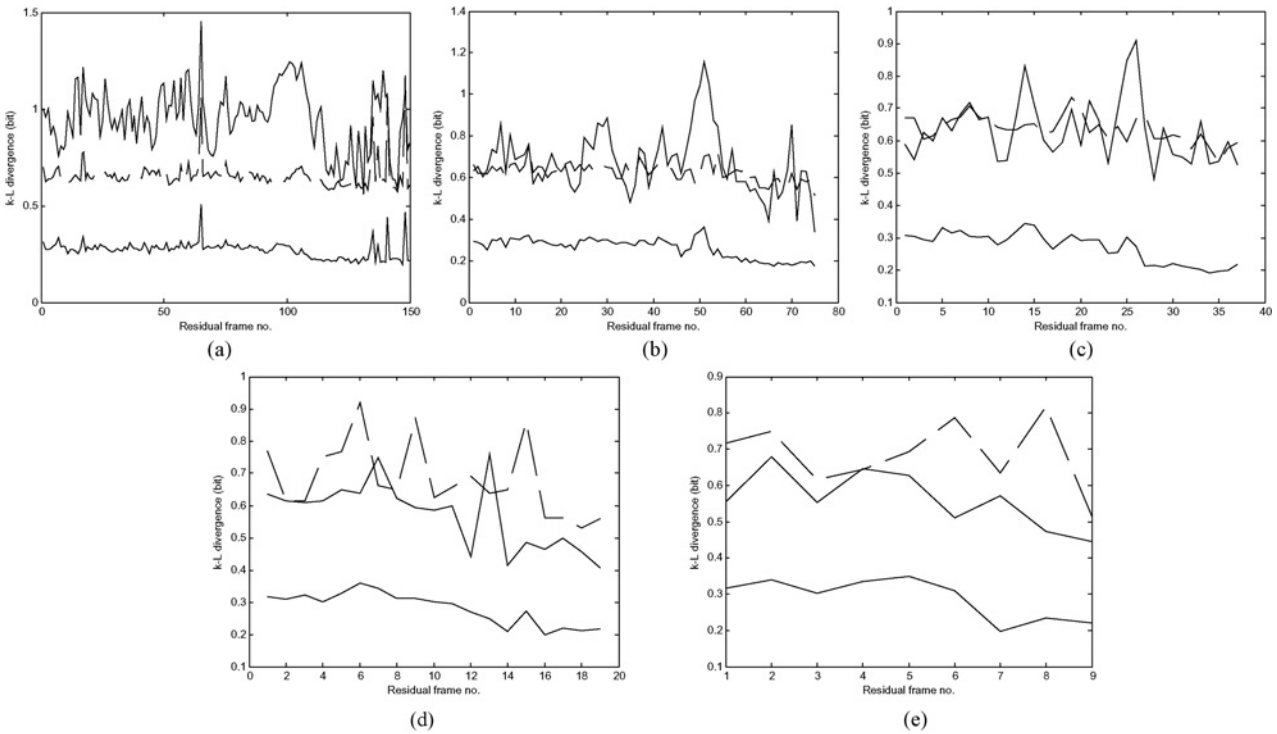

Fig. 7. Dotted, dashed, and solid lines show the K–L divergence between the probability distributions of the absolute-valued signal and three approximations. These three approximations are Laplacian distribution (dotted), one-sided ρ-GGD (dashed), and one-sided ρ-GGD with the improved ρ estimation (solid), respectively. Figures (a)–(e) are the results at different temporal levels (t). (a) $t = 0$. (b) $t = 1$. (c) $t = 2$. (d) $t = 3$. (e) $t = 4$. The test sequence is *Foreman* (CIF, 30 f/s).

Step 2: *Perform motion estimation for mode m:* Given a candidate mode m, the current MB is partitioned to N_m sub-blocks. Thus, we have to find the best motion vector for each sub-block and combine them into the motion vector set for this MB. For the ith sub-block, we test motion vector v for motion compensation and obtain the residual sub-block b_i. The residual signal variance is calculated and denoted as ${\sigma_{b_i}}^2(v)$; the zero-value probability of the one-sided ρ-GGD model is estimated by (15) and (18) and is denoted as ρ_{b_i}. According to (13), the shape parameter of the sub-block b_i can be obtained by

$$\alpha_{b_i} = {\Omega_e}^{-1}\left({\rho_{b_i}}^2(v) \cdot {\sigma_{b_i}}^2(v)\right). \tag{53}$$

Therefore, the MIG cost for motion vector v is

$$J_{\mathrm{mv}}(b_i, v|C_t) = \tau(\alpha_{b_i}) \cdot {\sigma_{b_i}}^2(v) \cdot 2^{2 \cdot C_t \cdot \Delta R(v)} \tag{54}$$

where $\triangle R(v)$ is the motion bitrate. If the entire MV candidate set (search range) is denoted as $\mathbf{S}$, for all motion vector $v \in \mathbf{S}$, the best motion vector for the sub-block b_i can be found by

$$v_i^* = \arg\min_{v \in \mathbf{s}}\{J_{\mathrm{mv}}(b_i, v|C_t)\}. \tag{55}$$

This is the most time-consuming process in our procedure. Repeating the same process for all N_m sub-blocks, we obtain all the MVs needed for mode m. The resultant motion vector set of mode m is

$$\mathbf{v}_m^* = (v_1^*, \cdots, v_{N_m}^*). \tag{56}$$

Step 3: *Calculate the residual MB statistics and the motion rate:* The MB residual signal $\mathbf{b_m}$ for mode m is obtained in Step 2 after performing motion compensation using the MV set $\mathbf{v}_m^*$. To construct the one-sided ρ-GGD model for $\mathbf{b_m}$, we need to calculate the variance and estimate the zero-value probability. Let $\rho_{\mathbf{b}_m}$ and ${\rho_{\mathbf{b}_m}}^2$ denote the zero-value probability and the variance of $\mathbf{b}_m$, respectively. ${\sigma_{\mathbf{b}_m}}^2$ is computed by

$${\sigma_{b_m}}^2(\mathbf{v}_m^*) = \frac{1}{N_m}\sum_{i=1}^{N_m} \sigma {b_i}^2(v_i^*) \tag{57}$$

where $\mathbf{V}_m^*$ is the best motion vector set for mode m in Step 2; $\rho\mathbf{b}_m(\mathbf{v}_m^*)$ is estimated by (15) and (18). Next, the motion bitrate for this MB is given by

$$\Delta R(\mathbf{v}_m^*) = \frac{1}{N_m}\sum_{i=1}^{N_m} \Delta R(v_i^*) + r_m \tag{58}$$

where $\Delta R(v_i^*)$ is the bitrate of encoding MV, and v_i^* and r_m is the average bitrate for recording the MB mode information.

Step 4: *Estimate the shape parameter from MB residuals:* According to (13), the shape parameter of $\mathbf{b}_m$ is estimated by

$$\alpha_{\mathbf{b}_m} = {\Omega_e}^{-1}\left({\rho_{\mathbf{b}_m}}^2(\mathbf{v}_m^*) \cdot {\sigma_{\mathbf{b}_m}}^2(\mathbf{v}_m^*)\right). \tag{59}$$

Step 5: *Calculate the MIG cost for mode m:* Using the parameter values calculated in Steps 1–5, we can

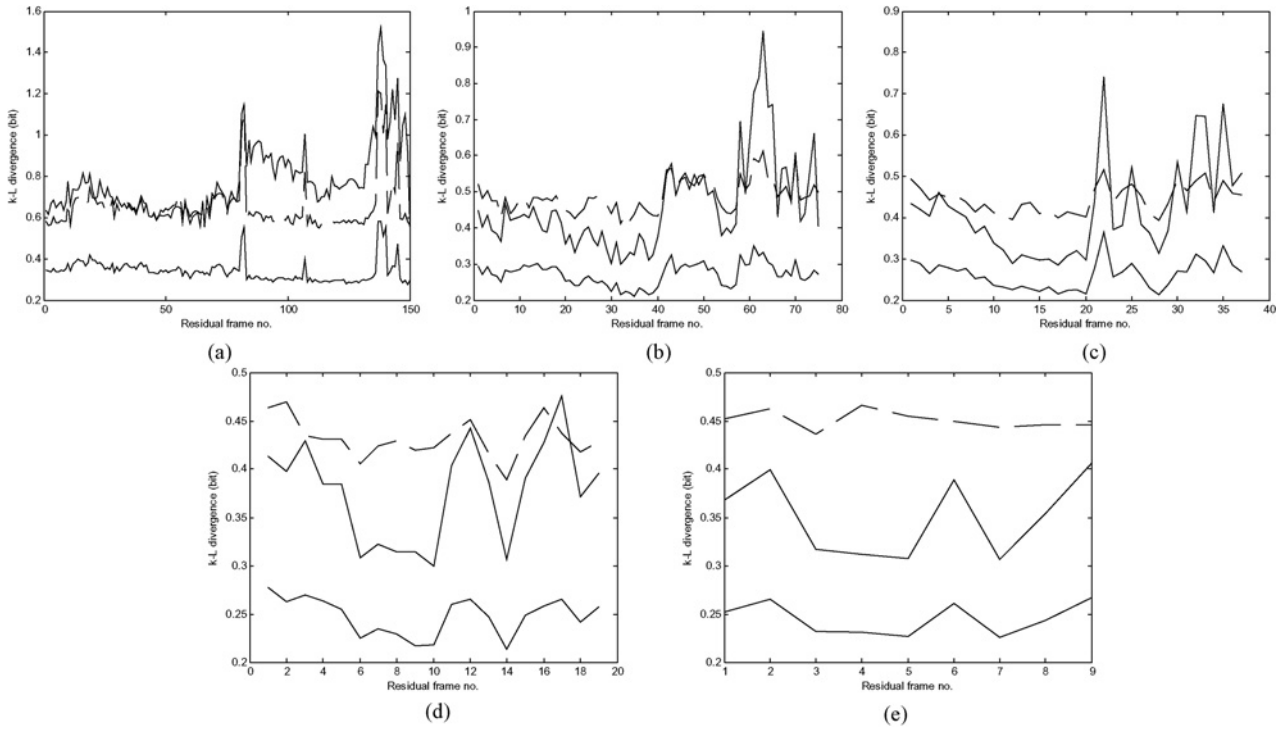

Fig. 8. Dotted, dashed, and solid lines show the K–L divergence between the probability distributions of the absolute-valued signal and three approximations. These three approximations are Laplacian distribution (dotted), one-sided ρ-GGD (dashed), and one-sided ρ-GGD with the improved ρ estimation (solid), respectively. Figures (a)–(e) are the results at different temporal levels (t). (a) $t = 0$. (b) $t = 1$. (c) $t = 2$. (d) $t = 3$. (e) $t = 4$. The test sequence is *Mobile* (CIF, 30 f/s).

compute the MIG cost for mode m as follows:

$$J_{\text{mode}}(\mathbf{b}_m, \mathbf{v}_m^* | C_t) = \tau(\alpha_{\mathbf{b}_m}) \cdot {\sigma_{\mathbf{b}_m}}^2(\mathbf{v}_m^*) \cdot 2^{2 \cdot C_t \cdot \Delta R(\mathbf{v}_m^*)}. \quad (60)$$

If mode m is the last mode in $\mathbf{M}$, go to Step 6 to decide the best prediction mode; if not, go to Step 2 to perform the same operation for the next candidate mode.

Step 6: *Choose the best mode* m* *with the minimum cost:* After all MIG costs for all $m \in \mathbf{M}$ are obtained, the best mode m^* is obtained by

$$m^* = \arg\min_{m \in M} \left\{ J_{\text{mode}}(\mathbf{b}_m, \mathbf{v}_m^*) | C_t \right\}. \quad (61)$$

V. Experimental Results and Discussions

In this section, we conduct two experiments: 1) the source modeling performance of the proposed one-sided ρ-GGD, and 2) the rate-distortion performance of the proposed MV selection and mode decision scheme. We implement our methods on the Vidwav reference software [23]. Other than the MV/mode decision part, all the other parts of Vidwav are not altered.

A. Source Modeling Performance of the Proposed One-Sided ρ-GGD

In Section II-B, we propose the one-sided ρ-GGD model and an efficient estimation method on the shape parameter. Furthermore, an improved ρ estimation method is proposed in Section II-C. In this experiment, we compare the modeling results using three different methods; they are one-sided Laplacian, the proposed one-sided ρ-GGD, and the proposed one-sided ρ-GGD with improved ρ estimation. We use the K–L divergence to measure the modeling accuracy. A small K–L divergence value means a more accurate approximation. For each MB in a frame, the K–L divergence between the probability distribution of the residual absolute-valued signal and its approximation is calculated. Then, we take the average of the K–L divergences of all MBs in one frame. Figs. 7(a) and 8(a) show the average K–L divergences of all residual frames at the first temporal level of two test sequences, *Foreman* and *Mobile*, respectively (CIF format, and 30 f/s). From Figs. 7(a) and 8(a), the proposed one-sided ρ-GGD shows a better modeling accuracy than Laplacian. Also, with the improved ρ estimation, the approximation accuracy of the one-sided ρ-GGD is further improved. Because the low-pass frame quality degrades after temporal decompositions, the motion compensation efficiency is also reduced at deep temporal level. In the meanwhile, modeling the probability distribution of residual signal becomes more difficult. Figs. 7(b)–(e) and 8(b)–(e) show the modeling performance of the residual frames for the rest of temporal levels. We can see that the proposed one-sided ρ-GGD with the improved ρ estimation consistently maintains good approximation accuracy at all temporal levels.

B. Rate-Distortion Performance of the Proposed MV Selection and Mode Decision Scheme

In this experiment, we compare the rate-distortion performance of the proposed MV selection and mode decision scheme with that of the conventional Lagrangian method in the original Vidwav. Based on the one-sided ρ-GGD source

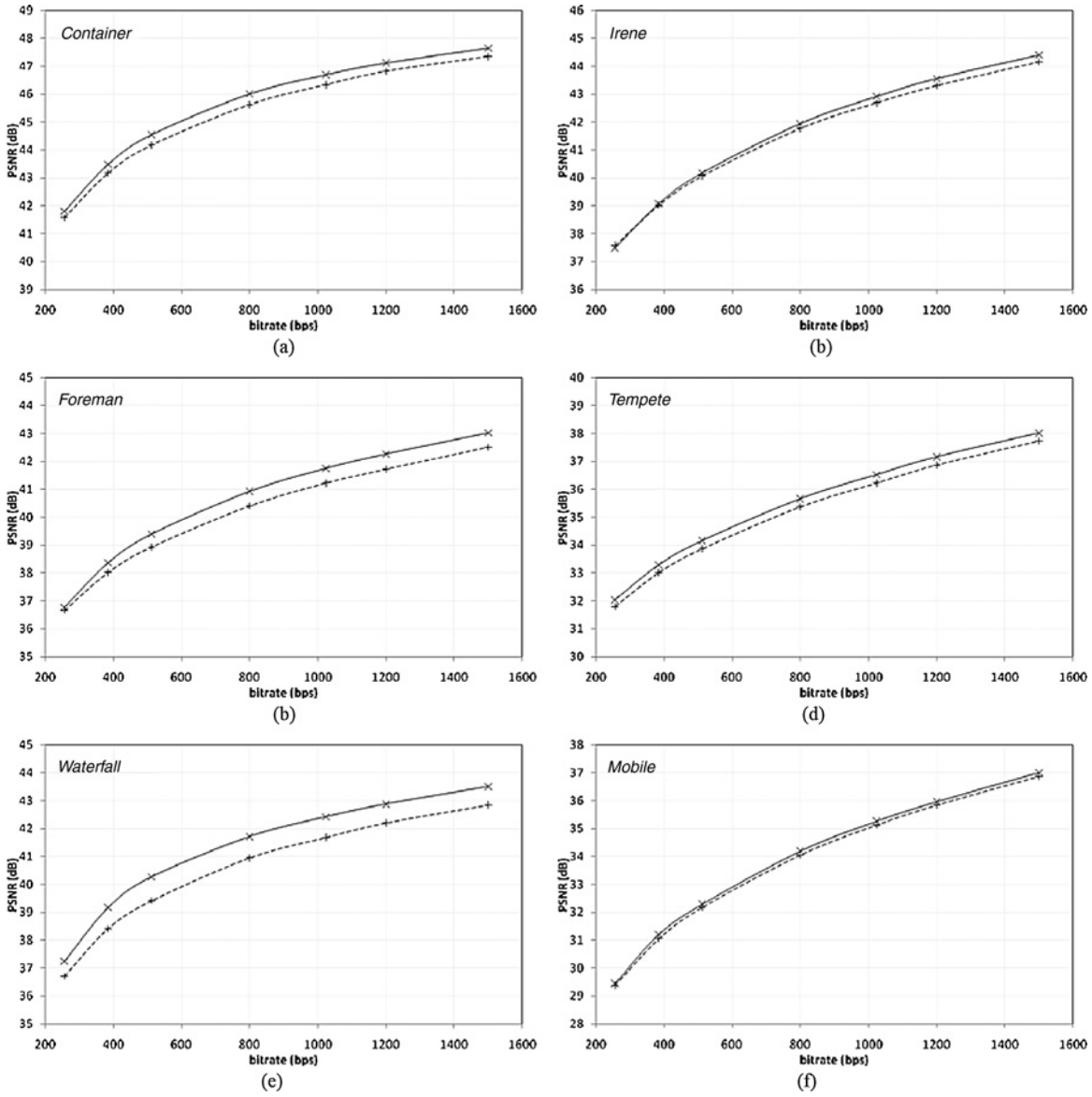

Fig. 9. PSNR comparison between the proposed MIG cost method (solid line) and the conventional Lagrangian method (dashed line). The test sequences are (a) *Container*, (b) *Irene*, (c) *Foreman*, (d) *Tempete*, (e) *Waterfall*, and (f) *Mobile*. (CIF, 30 f/s.)

model, we derive its MIG cost function and use it to decide the best MV and prediction mode. The MCTF parameters of the conventional Lagrangian method are given in Table IV. Our proposed method uses the same motion search range and motion vector accuracy settings in Table IV. The parameters, C_0 and ω, are empirically selected and will be given below. We focus on the mid bitrate to high bitrate cases. There are two scenarios in this experiment.

The first scenario is the SNR scalability test. We select six test sequences: *Container*, *Irene*, *Foreman*, *Tempete*, *Waterfall*, and *Mobile*. All are in the CIF format and 30 f/s. In this scenario, C_0 and ω of the MIG cost function are 7 and 0.8, respectively. The operation bitrates are 256 kb/s, 384 kb/s, 512 kb/s, 800 kb/s, 1024 kb/s, 1.2 Mb/s, and 1.5 Mb/s. For each test sequence, seven bitstreams are extracted according to the bitrate conditions from the same losslessly coded bitstream, and then each extracted bitstream is decoded to obtain the PSNR at various selected bitrate points. Fig. 9 shows the PSNR comparison between the two coding methods for the six test sequences. Compared with the conventional Lagrangian method, our method shows 0.1–0.9 dB PSNR improvements.

TABLE IV
DEFAULT PARAMETER SETTINGS [24] OF MCTF IN VIDWAV CODER

	Motion Search Range (pel)	Motion Vector Accuracy (pel)		Lagrange Parameter	
		CIF	4 CIF	CIF	4 CIF
$t = 0$	32	1/4	1/4	16	16
$t = 1$	64	1/2	1/2	32	50
$t = 2$	128	1/2	1	64	150
$t = 3$	128	1/2	1	64	150
$t = 4$	128	1/2	1	64	150

The second scenario is the combined temporal and SNR scalability test. In this scenario, in addition to the CIF videos

TABLE V
PSNR COMPARISON BETWEEN THE PROPOSED MIG COST METHOD AND THE CONVENTIONAL LAGRANGIAN METHOD IN COMBINED TEMPORAL AND SNR SCALABILITY TEST FOR FIVE TEST SEQUENCES (4 CIF, 60 F/S)

Sequence (4 CIF)	GoP Size	Mode Decision Method	750 kb/s 15 f/s	1024 kb/s 15 f/s	1200 kb/s 30 f/s	1500 kb/s 30 f/s	2048 kb/s 60 f/s	3000 kb/s 60 f/s
City	32	Lagrangian	36.39	37.33	37.42	37.98	38.49	39.33
		Proposed	36.72	37.70	37.81	38.42	38.86	39.63
Crew	32	Lagrangian	36.39	37.30	36.74	37.34	37.18	38.20
		Proposed	36.41	37.35	36.87	37.50	37.38	38.34
Harbor	32	Lagrangian	33.91	34.97	34.96	35.59	36.25	37.50
		Proposed	33.94	35.02	34.99	35.65	36.29	37.53
Soccer	32	Lagrangian	36.28	37.22	36.92	37.61	38.00	39.20
		Proposed	36.52	37.52	37.18	37.94	38.20	39.42
Ice	16	Lagrangian	40.51	41.65	41.25	42.00	42.41	43.62
		Proposed	40.88	42.05	41.75	42.51	42.84	44.06

in the first scenario, we also test five high-resolution test sequences: *City*, *Crew*, *Harbour*, *Soccer*, and *Ice*. All are in the 4 CIF format and 60 f/s. The operation points include six bitrates combined with three frame rates. The C_0 value is empirically selected within [7] and [10], and ω is 0.8. Table V lists the PSNR results of the proposed MIG and the conventional Lagrangian methods. Our proposed method shows 0.1 to 0.5 dB PSNR improvements on all 30 test points.

VI. CONCLUSION

The main theme of this paper was to develop the "one-sided ρ-GGD" source model to approximate the probability distribution of the residual signals in the scalable wavelet video codec. Extended from our earlier findings, we suggested a fast scheme that constructs the one-sided ρ-GGD based on the zero-value probability (ρ) and the source signal variance. Also, we proposed a piecewise linear expression to estimate the shape parameter of the source model. Furthermore, an improved ρ estimation scheme is proposed to increase the model accuracy.

In the second half of this paper, we followed our previous approach [17] to derive the rate-distortion function for the wavelet video coder based on the one-sided ρ-GGD model. The notion of MIG in [17] is carried over and a similar mode decision procedure is developed. This mode decision procedure is less bitrate dependent and thus is suitable for solving the multi-operation-point problem in scalable wavelet video coding. Our simulation results showed that the one-sided ρ-GGD-based mode decision algorithm provided a 0.1–0.5 dB PSNR improvements over the conventional Lagrangian method on both the SNR scalability and the combined SNR and temporal scalability tests.

APPENDIX

DIFFERENTIAL ENTROPY OF THE HIGH-ORDER EXPONENTIAL PDF

Let $p(x)$ be a high-order exponential probability distribution function given by

$$p(x) = \gamma \exp(-\beta x^{\alpha}), \quad x \geq 0 \tag{62}$$

where $\exp(\cdot)$ is the exponential function. α, β, and γ are positive constants. By definition, the differential entropy of $x \in X$ is

$$\Phi(X) = -\int_X p(x) \cdot \log(p(x)) dx \tag{63}$$

where $\log(\cdot)$ is the natural logarithm function. $\Phi(X)$ can be derived as

$$\begin{aligned}\Phi(X) &= -\int_0^{\infty} \gamma \exp(-\beta x^{\alpha}) \cdot \log(\gamma \exp(-\beta x^{\alpha})) dx. \\ &= \gamma \left(\beta \int_0^{\infty} x^{\alpha} \exp(-\beta x^{\alpha}) dx - \log\gamma \cdot \int_0^{\infty} \exp(-\beta x^{\alpha}) dx \right).\end{aligned} \tag{64}$$

Here, we rewrite $\Phi(X)$ as

$$\Phi(X) = \gamma(\beta \cdot A - \log\gamma \cdot B) \tag{65}$$

where

$$\begin{aligned}A &= \int_0^{\infty} x^{\alpha} \cdot \exp(-\beta x^{\alpha}) dx \\ B &= \int_0^{\infty} \exp(-\beta x^{\alpha}) dx.\end{aligned} \tag{66}$$

Let us derive A and B first, and then substitute the results into $\Phi(X)$ in (65). We use a new variable $t = -\beta x^{\alpha}$ to replace the variable x in A. Thus, A is derived as

$$\begin{aligned}A &= \int_0^{\infty} \beta^{-1} t \exp(-t) \alpha^{-1} \beta^{-1/\alpha_t 1/\alpha - 1} dt \\ &= \alpha^{-1} \beta^{-(1/\alpha+1)} \int_0^{\infty} \exp(-t) t^{(1/\alpha+1)-1} dt \\ &= \alpha^{-1} \beta^{-(1/\alpha+1)} \cdot \Gamma(\alpha^{-1} + 1)\end{aligned} \tag{67}$$

where $\Gamma(\cdot)$ is the standard Gamma function. With the similar procedure, B in (66) is derived as

$$\begin{aligned}B &= \alpha^{-1} \beta^{-1/\alpha} \int_0^{\infty} \exp(-t) \cdot t^{1/\alpha - 1} dt \\ &= \alpha^{-1} \beta^{-1/\alpha} \cdot \Gamma(\alpha^{-1}).\end{aligned} \tag{68}$$

By using (67) and (68), $\Phi(X)$ can be rewritten as

$$\begin{aligned}\Phi(X) &= \gamma\left(\beta\alpha^{-1}\beta^{-(1/\alpha+1)}\Gamma(\alpha^{-1}+1) - \log\gamma\cdot\alpha^{-1}\beta^{-1/\alpha}\Gamma(\alpha^{-1})\right)\\ &= \gamma\alpha^{-1}\beta^{-1/\alpha}\left(\Gamma(\alpha^{-1}+1) - \log\gamma\cdot\Gamma(\alpha^{-1})\right)\\ &= \gamma\alpha^{-1}\beta^{-1/\alpha}\Gamma(\alpha^{-1})(\alpha^{-1} - \log\gamma\cdot)(\text{nat}). \qquad (69)\end{aligned}$$

Therefore, the differential entropy of the high-order exponential probability distribution function is derived.

REFERENCES

[1] *ISO/IEC 14496-10/Amd.3 Scalable Video Coding*, document JVT-X201, ITU-T SG16 Q.6, Jul. 2007.

[2] R. Leonardi, T. Oelbaum, and J.-R. Ohm, *Status Report on Wavelet Video Coding Exploration*, document N8043, ISO/IEC JTC1/SC29/WG11 MPEG, Apr. 2006.

[3] D. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Boston, MA: Kluwer, 2002.

[4] J.-R. Ohm, "3-D subband coding with motion compensation," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 559–571, Sep. 1994.

[5] S.-T. Hsiang and J. W. Woods, "Embedded video coding using invertible motion compensated 3-D subband wavelet filter bank," *Signal Process.: Image Commun.*, vol. 16, pp. 705–724, May 2001.

[6] A. Secker and D. Taubman, "Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1530–1542, Dec. 2003.

[7] J. M. Shapiro, "An embedded wavelet hierarchical image coder," in *Proc. IEEE ICASSP*, Mar. 1992, pp. 657–660.

[8] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.

[9] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1158–1170, Jul. 2000.

[10] T. Berger, *Rate Distortion Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1984, sec. 4.3, pp. 92–94.

[11] W.-H. Peng, C.-Y. Tsai, T. Chiang, and H.-M. Hang, "Advances of MPEG scalable video coding standards," in *Intelligent Multimedia Data Hiding*. Berlin-Heidelberg, Germany: Springer-Verlag, 2007, ch. 3, pp. 55–80.

[12] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences," *IEEE J. Select. Areas Commun.*, vol. SAC-5, no. 7, pp. 1140–1154, Aug. 1987.

[13] B. Girod, "Rate-constrained motion estimation," in *Proc. Int. Symp. Vis. Commun. Image Process.*, Nov. 1994, pp. 1026–1034.

[14] M. Wang and M. van der Schaar, "Operational rate-distortion modeling for wavelet video coders," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3505–3517, Sep. 2006.

[15] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Nov. 1998.

[16] C.-Y. Tsai and H.-M. Hang, "Rate-distortion model for motion prediction efficiency in scalable wavelet video coding," in *Proc. Packet Video Workshop*, May 2009, pp. 1–9.

[17] C.-Y. Tsai and H.-M Hang, "A rate-distortion analysis on motion prediction efficiency and mode decision for scalable video coding," *J. Vis. Commun. Image Representat.*, vol. 21, no. 8, pp. 917–929, Nov. 2010.

[18] C.-Y. Tsai and H.-M. Hang, "ρ-GGD source modeling for wavelet coefficients in image/video coding," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jun. 2008, pp. 601–604.

[19] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Trans. Image Process.*, vol. 8, no. 12, pp. 1688–1701, Dec. 1999.

[20] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.

[21] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.

[22] R. Xiong, J. Xu, and F. Wu, "Optimal subband rate allocation for spatial scalability in 3-D wavelet video coding with motion aligned temporal filtering," in *Proc. VCIP*, Jul. 2005, pp. 381–392.

[23] R. Xiong, X. Ji, D. Zhang, and J. Xu, *Vidwav Wavelet Video Coding Specifications*, document M12339, ISO/IEC JTC1/SC29/WG11 MPEG, 2005.

[24] R. Xiong, J. Xu, and F. Wu, *Coding Performance Comparison Between MSRA Wavelet Video Coding and JSVM*," document M11975, ISO/IEC JTC1/SC29/WG11 MPEG, 2005.

Chia-Yang Tsai (M'09) received the B.S., M.S., and Ph.D. degrees in electronics engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2001, 2003, and 2010, respectively.

He joined the MediaTek Incorporated, Hsinchu, in 2010. From 2003 to 2005, he contributed to the development of the MPEG-21 Part 12 Test Bed for the MPEG-21 Resource Delivery standard. Since 2010, he has been actively participating in the high efficiency video coding standardization process. He holds three patents (Taiwan and U.S.) and has published over 20 technical papers in the field of video signal processing. His current research interests include video compression algorithms, scalable video coding, multimedia communication, and image/video signal processing.

Hsueh-Ming Hang (F'02) received the B.S. and M.S. degrees in control engineering and electronics engineering from National Chiao Tung University (NCTU), Hsinchu, Taiwan, in 1978 and 1980, respectively, and the Ph.D. degree in electrical engineering from Rensselaer Polytechnic Institute, Troy, NY, in 1984.

From 1984 to 1991, he was with AT&T Bell Laboratories, Holmdel, NJ. He joined the Department of Electronics Engineering, NCTU, in December 1991. From 2006 to 2009, he took a leave from NCTU and was appointed as the Dean of the College of Electrical Engineering and Computer Science, National Taipei University of Technology, Taipei, Taiwan. He is currently a Distinguished Professor with the Department of Electrical Engineering, NCTU, and an Associate Dean with the College of Electrical and Computer Engineering, NCTU. He has been actively involved in the international MPEG standards since 1984. He holds 13 patents (Taiwan, U.S., and Japan) and has published over 170 technical papers related to image compression, signal processing, and video codec architecture. His current research interests include multimedia compression, image/signal processing algorithms and architectures, and multimedia communication systems.

Dr. Hang was the recipient of the IEEE Third Millennium Medal. He was an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1992 to 1994, and of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 1997 to 1999. He is currently an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING again. He is a co-editor and contributor of the *Handbook of Visual Communications* (New York: Academic Press). He is a Fellow of the Institution of Engineering and Technology and a member of Sigma Xi.