國 立 交 通 大 學

電子工程學系 電子研究所

博 士 論 文

適用於可調式小波視訊編碼之訊源機率模型與位元率-失真最佳化方法

Source Modeling and Rate-Distortion Optimization in Scalable Wavelet Video Coder

研 究 生： 蔡 家 揚

指導教授： 杭 學 鳴

中華民國九十九年十月

適用於可調式小波視訊編碼之訊源機率模型與位元率-失真最佳化方法

Source Modeling and Rate-Distortion Optimization in Scalable Wavelet Video Coder

研 究 生：蔡 家 揚　　Student: Chia-Yang Tsai

指導教授：杭 學 鳴 博士　　Advisor: Dr. Hsueh-Ming Hang

國 立 交 通 大 學

電子工程學系 電子研究所

博 士 論 文

A Dissertation

Submitted to Department of Electronics Engineering

& Institute of Electronics

College of Electrical and Computer Engineering

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Electronic Engineering

October 2010

Hsinchu, Taiwan, Republic of China

中華民國九十九年十月

適用於可調式小波視訊編碼之訊源機率模型與位元率-失真最佳化方法

研究生：蔡家揚　　指導教授：杭學鳴

國立交通大學 電子工程學系 電子研究所博士班

摘要

本研究主題可分為兩個主要項目：動態補償(motion-compensated) 差值訊號之訊源模型(source model)以及位元率-失真(rate-distortion)最佳化參數選擇，在第一個項目中，我們發展了零值廣義高斯分佈(ρ-GGD)訊源模型，可準確模擬可調式小波編碼(scalable wavelet coding)中的訊號機率分佈。我們提出了分段式線性方法可有效率估測在零值廣義高斯模型中的型態參數，並且藉由改善零值機率的估測而提高機率模型之精準度。在第二個項目中，我們提出了一個位元率-失真(rate-distortion)模型用以描述可調式小波視訊編碼中的動態預測效率。可調式編碼架構為開放式迴圈(open-loop)並且同時有多種位元率的編碼需求，與傳統非可調式編碼有很大的不同，也因此傳統上廣泛使用的拉格朗日(Lagrangian)最佳化方法無法良好應用於可調式小波視訊編碼上。為了找到在動態資訊與殘存訊號間最好的位元率分配方法，我們提出了動態資訊增益(MIG)做為量測動態預測效率指標。基於這項指標，新的代價函式一同被提出。相較於傳統拉格朗日最佳化作法，我們的實驗結果顯示了所提出的模式決定方法可在 SNR 與畫面率可調式條件下，擁有較佳的 PSNR 表現。

Source Modeling and Rate-Distortion Optimization in Scalable Wavelet Video Coder

Student: Chia-Yang Tsai Advisor: Dr. Hsueh-Ming Hang

Department of Electronics Engineering & Institute of Electronics
National Chiao Tung University

Abstract

There are two key elements in this study, namely, the source modeling of the motion-compensated prediction error signals, and the coding parameter selection to minimize the rate-distortion criterion. For the first item, we develop an accurate ρ-GGD (Generalized Gaussian Distribution) source model for approximating the signal probability distribution in scalable wavelet coding. An efficient piecewise linear expression is designed to estimate the shape parameter of the ρ-GGD. We also improve the model accuracy in matching the real data by modifying the ρ parameter estimation formula. For the second item, a rate-distortion model for describing the motion prediction efficiency in scalable wavelet video coding is proposed. Different from the conventional non-scalable video coding, the scalable wavelet video coding needs to operate under multiple bitrate conditions and it has an open-loop structure. The conventional Lagrangian multiplier, which is widely used to solve the rate-distortion optimization problems in video coding, does not fit well into the scalable wavelet structure. In order to find the rate-distortion trade-off due to different bits allocated to motion and textual residual information, we suggest a motion information gain (MIG) metric to measure the motion prediction efficiency. Based on this metric, a new cost function for mode decision is proposed. Compared with the conventional Lagrangian optimization, our experimental results show that the new mode decision method

generally improves the PSNR performance in the combined SNR and temporal scalability cases.

誌謝

首先要感謝的是我的指導老師杭學鳴教授，在杭老師的指導下，我才能夠逐步學習要如何作研究、要如何培養獨立思考解決問題的能力，在我漫長的研究生生涯中，杭老師總是扮演著貴人的角色。不管是在開會時的討論，論文的修改，或是為人相處上謙沖態度，都讓我學習良多。杭老師尤其提供我許多國際交流的機會，多次參與國際研討會發表，或是兩次在美國 UIUC 交換學生，都讓我在外語能力上有很大的提升，並且感受到不同的學術風氣。另外，還要感謝在 MPEG 合作計畫時，蔣迪豪老師與蔡淳仁老師的指導，讓我有機會可以參與國際標準的寶貴經驗。

還要感謝在這一路上陪伴著我的好友們。感謝 CommLab 的老師們、還有許多已畢業未畢業的成員們,讓我可以自在地以實驗室為家,也祝福你們在未來畢業與工作上順利。感謝每星期聚餐的 Wii 組與麻將組大學好友們，讓我星期五的晚上總是充滿美食跟歡樂。還有感謝我的室友們，一同培養各種第二專長，一同度過許多的 AOM 連線及麥當勞外送的夜晚，也祝你們早日博士班畢業。另外還要感謝 PPS 提供的大量美劇，還有 Ptt 的鄉民們，讓我在苦悶的研究生生涯中多了許多樂趣與話題。

再來要感謝我的未婚妻珮芬，總是在我研究低潮時給我最有力的鼓勵，總是陪我度過許多的高興與悲傷，一起品嘗美食一同遊玩，答應妳的畢業禮物機車包我會記得送妳的，未來的日子我會努力給你幸福。

最後是我親愛的爸媽，沒有你們的支持，我不可能有一絲的成就，你們讓我無後顧之憂安心作研究，並且給我許多寶貴的人生建議，我的學位希望可以榮耀你們，讓你們以我為榮，謝謝你們。

家揚

2010.10.17

交大

目錄

表目錄(List of Tables)

圖目錄(List of Figures)

Chapter 1 Introduction

Over the past few years, multimedia delivery becomes an important class of wireless/wired internet applications, for example, mobile video and digital TV broadcasting. To overcome the constraints on transmission bandwidth and receiver capability, the scalable coding technique was developed and adopted by the recent international video standards. There are two major approaches on scalable video coding: the DCT-based and the wavelet-based coding schemes. These two coding schemes share many similar coding concepts, especially in removing the temporal redundancy. The Scalable Video Coding (SVC) extension of the H.264/AVC is a representative scheme of the DCT-based approach and has been accepted as the ITU/MPEG standards in 2007 [1]. On the other hand, the wavelet-based coding scheme is a relatively new structure and has its potential and advantages [2] as shown during the MPEG competition process for standardization.

Discrete wavelet transform (DWT) has been successfully applied to still image compression. By exploiting the inter-subband or intra-subband correlation, the DWT transformed image signal can be efficiently compressed by a context-based entropy coder, such as EZW [3], SPIHT [4], and EBCOT [5]. Different from the DCT-based JPEG image coding, the multiresolution property of wavelet transform provides a natural way in producing scalable bitstreams. It enables the spatial and the SNR scalability features in the well-known JPEG2000 image coding standard [6]. In addition to the spatial decomposition,

DWT can also be applied along the temporal axis and decomposes video frames into temporal subband signals. Therefore, it provides the temporal scalability for videos. In the past fifteen years, the temporal wavelet decomposition is refined by adopting the motion compensated temporal filtering (MCTF) technique. These schemes were proposed and improved by Ohm [7], Hsiang and Woods [8], Secker and Taubman [9], and Xu et al. [10]. MCTF can efficiently decompose video frames along the motion trajectories. After MCTF and spatial 2-D DWT, the original video frames are transformed to spatio-temporal subband signals and compressed by a context-based entropy coder [9], [11]. This interframe wavelet video coding scheme can achieve temporal, spatial and SNR scalability goals simultaneously. Depending on the processing order in the spatio-temporal domain, the scalable wavelet coding methods can be classified to "t+2D" and "2D+t" structures [12]. In this study, we will focus on the t+2D structure.

The rate-distortion analysis of a scalable interframe wavelet video coder is very different from that of a DCT-based coder owing to the following two issues: inter-scale coding and open-loop coding structure. In DCT-based video coders, such as MPEG-2 or H.264, use the hybrid coding technique; all the temporal and spatial prediction operations are basically block-based. Thus, it is quite straightforward to perform the rate-distortion analysis along the coding operation flow. On the other hand, in the interframe wavelet coders, the temporal MCTF is performed block-wise, but the spatial entropy coding is performed on the

subbands. This inconsistent data partition increases the rate-distortion analysis difficulty drastically. Wang and Schaar proposed a solution in [13] to analyze the rate-distortion behavior across different coding scales for wavelet video coder. The second issue is that the DCT-based video coder has a closed-loop coding structure. The prediction errors within the loop can be controlled by adjusting coding parameters [14]; thus, the optimal rate-constrained motion compensation can be adaptively adjusted [15],[16]. But the interframe wavelet coding has an open-loop prediction structure and the quantization process is performed after all the encoding operations are completed. This open-loop scheme provides more flexibility on bitstream extraction and robustness to transmission errors, but it has no feedback path to provide useful information to adjust prediction parameters in the encoding process. Therefore, it is difficult to achieve the rate-distortion optimization target, especially in the case of allocating bits between the motion and the texture data at multiple operation points all at the same time. How to generate adequate amount of motion information and decide the best prediction modes for MCTF becomes a challenging problem in the scalable interframe wavelet video coding.

Our objective is to develop a rate-distortion optimization method to improve the coding performance of scalable wavelet video coding. For building an efficient rate-distortion model, we propose an accurate source model. Moreover, we also suggest a piecewise linear method to estimate the shape parameter of the Model. Besides, we derive an analytical

model that describes the trade-off between the motion compensation bits and the residual texture coefficients bits. We then allocate bits to each category properly at different scalability dimensions. We first examine the rate-distortion effect due to the increase or decrease of motion information bits. Then we derive a quantitative expression to measure the motion prediction efficiency. Most significantly, we give a theoretical explanation to this metric from the entropy viewpoint. Based on this finding, a new cost function is proposed. By minimizing the proposed cost function, the best prediction mode is decided and the corresponding motion vectors are chosen for the MCTF operation. Compared with the mode decision procedure in the conventional scalable wavelet video coder, the proposed method shows a PSNR improvement for the combined SNR and temporal scalability cases. The proposed methods are also published in [38] and [39].

This thesis is organized as follows. Chapter 2 gives a brief review of interframe wavelet video and the rate-distortion mechanisms in video coding. In Chapter 3, the ρ-GGD source modeling is proposed to approximate the probability distribution of wavelet coefficients. In Chapter 4, we suggest the motion information gain (MIG) metric to measure the motion prediction efficiency. According to our source model, the MIG metric is further discussed from the entropy viewpoint. Extending the work in Chapter 3, the ρ-GGD source model is improved by an enhanced estimation method of the ρ value. The one-sided ρ-GGD is proposed for the texture residual signal in Chapter 5. In Chapter 6, the two concepts, MIG

in Chapter 4 and one-sided ρ-GGD in Chapter 5, are integrated into a complete and working algorithm. The major contributions in this thesis are listed as follows.

Contributions of this Study

(1) An accurate and efficient source model, ρ-GGD, is proposed to approximate the probability distribution of the wavelet coefficients.

(2) A quantitative metric, MIG, is proposed to measure the motion prediction efficiency of MCTF.

(3) Based on the MIG metric, a new rate-distortion cost function is proposed for mode decision. The parameters of the MIG cost function are empirically selected.

(4) To further improve the ρ-GGD model, the one-sided ρ-GGD model and an more reliable estimation method on ρ are proposed to approximate the probability distribution of residual texture signal.

(5) Based on MIG and one-sided ρ-GGD, an integrated MIG mode decision algorithm is developed. The parameters of the cost function are first theoretically derived and then fine-tuned by experimental data.

Chapter 2 Scalable Wavelet Video Coding and Its Rate-Distortion Optimization

2.1 Brief Introduction to Interframe Wavelet Video Coding

The most popular coding structure of interframe wavelet video codec is the so-called "t+2D" structure as shown in Fig. 2-1. The order of "t+2D" implies the encoding operation order: the temporal analysis first and then the spatial analysis. The temporal analysis employs the MCTF technique. It decomposes a group of pictures (GOP) into several temporal high-pass frames and one low-pass frame along the motion vector trajectories. The motion information portion is, in the conventional approach, nonscalable, which is denoted as v in Fig. 2-1. Then, the spatial decomposition operation (2-D DWT) is applied to the low-pass and high-pass frames to form subbands for further quantization and entropy coding. With the help of a scalable entropy coder, these spatio-temporal subbands are compressed to a scalable bitstream, denoted as s in Fig. 2-1. Therefore, the coded output bitstream consists of two parts, one is the scalable bitstream for the texture information (s) and the other is the non-scalable bitstream for the motion information (v); together, they are denoted as $\{s, v\}$. To fulfill the application requirements imposed on the video bitrates, image resolution, and frame rate, the texture bitstream is truncated accordingly but the motion bitstream remains intact. Therefore, the output bitstreams of the bitstream extractor are $\{s_0', v\}$, $\{s_1', v\}$… $\{s_n', v\}$ to match the scalable requirements r_0, r_0,…, r_n, respectively, as shown in Fig. 2-1. The truncation mechanism is designed to collaborate with the scalable

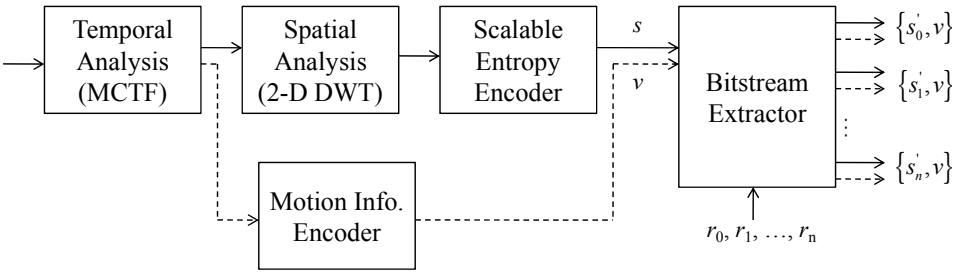


Fig. 2-1 The t+2D coding structure of interframe wavelet encoder. The solid line and dashed line show the data paths of the texture and motion information respectively.

entropy coder.

The EBCOT [5] image coding algorithm is adopted by the JPEG2000 standard, and similar algorithms are widely adopted by the state-of-art wavelet video codecs [9], [11]. The basic coding flow of an interframe wavelet video coder is as follows. After temporal and spatial analysis, each subband is partitioned into a number of code blocks, and the bitplanes of each block are processed by a few coding paths. The boundary between two consecutive coding paths is a truncation point. These truncation points are characterized by the slopes of the rate-distortion curves at the truncation point. These slope values are recorded and sent to the bitstream extractor. In one extraction unit, such as one GOP, the coding paths with similar slopes are grouped into the same coding layer. A permissible positive slope value is called a rate-distortion threshold. The coding layers with the absolute values of their slopes higher than the rate-distortion threshold are chosen to form an output bitstream. The sum of the bitrates of these chosen coding layers is calculated. If the calculated bitrate is less than the target bitrate, the rate-distortion threshold is adjusted to a smaller value so that more coding layers will be included and the total bitrate increases. On the other hand, the threshold value increases so as to discard some coding layers. By repeating the above operation, the bitrate of the truncated bitstream reaches the target value. Because each bitplane of a code block is split into three coding paths, the bitrate extraction can be quite

accurate. Therefore, the bitrate of the texture bitstream can be precisely controlled by the bitstream truncation mechanism. But the non-scalable motion information imposes a constraint on bitstream scalability. The motion information is typically temporal scalable and can be adapted to different decoding frame rates. However, when the spatial scalability feature is turned on, the motion information is often not adjustable to different decoding picture size during the extraction. In the following sub-section, we will compare the rate-distortion optimization methods for the non-scalable and the scalable video cases, and then develop the methods in the next section to adjust the motion information bitrate.

2.2 Rate-Distortion Mechanism in Video Coding

According to the Shannon's source coding theory [18], the rate-distortion function can be derived from the probability model of a coding source. Based on the rate-distortion function and with the help of optimization methods, an optimal rate-distortion trade-off can be theoretically obtained for a given bitrate or distortion condition.

In a typical hybrid video coding scheme, the coding source is the transformed residual signal after inter or intra predictions. It is well known that the probability distribution of the transformed coefficients can be closely approximated by the Laplacian distribution [21]

$$P(x) = \frac{\Lambda}{2}\exp\{-\Lambda|x|\}, \tag{1}$$

where Λ is the Laplacian parameter and can be estimated from the signal standard deviation σ by $\Lambda = \sqrt{2}/\sigma$. If the probability distribution of the transformed residual signal is a Laplacian source, its rate-distortion function with quantization distortion D and texture coding rate R was derived in [18]. In addition to the texture coding bit rate, the extra side information needed in a hybrid coder is mostly the motion information rate ΔR. According to the optimization theory, the best motion prediction mode can be obtained by minimizing the Lagrangian cost function defined by

$$J_{Mode} = D + \lambda_{Mode}(R + \Delta R), \tag{2}$$

where λ_{Mode} is the Lagrange parameter. For a fixed ΔR, λ_{Mode} can be theoretically derived for a well-defined rate-distortion function in (2). Both the theory and the real data show that

the λ_{Mode} value is strongly related to the quantization step size, which controls the amount of distortion directly [22], [23]. Different λ_{Mode} values are used by several popular reference encoders. These λ_{Mode} values are picked or derived based on their system characteristics and the experimental data [24]. The rate-constrained motion estimation is performed separately by using another Lagrangian cost function given by

$$J_{Motion} = FD + \lambda_{Motion}\Delta R, \tag{3}$$

where *FD* is a function of the frame difference between the original and the reconstructed image blocks. In many practical systems, *FD* is either SSD (sum of squared differences) or SAD (sum of absolute differences). In the MPEG reference encoder, λ_{Motion} is empirically chosen to be λ_{Mode} and $\sqrt{\lambda_{Mode}}$ for SSD and SAD, respectively [22].

From (2) and(3), λ_{Mode} is, clearly, an important factor that balances the weights of rate and distortion in the overall cost (*J*) and it thus affects the bitrates allocated to the texture and the motion information. As discussed earlier, λ_{Mode} depends on the source characteristics, the quantization step size and the bit rate. Several papers [19], [20] show that the statistics of the texture are helpful in selecting the proper λ_{Mode} value. The key for solving the mode decision and bit allocation problem is to find the relationship between quantization step size, texture characteristics and bit rate.

Using only one fully self-embedded bitstream to satisfy different coding requirements simultaneously is the most attractive feature of the scalable video coding technique. In the

scalable interframe wavelet coding, the bitstream generation process and bitstream extraction process are two separate, independent steps. The encoding process generates lossless compressed bitstream. After the encoding, the extractor truncates the lossless bitstream according to the bitrate requirement. In other words, the extractor plays the role of quantizer. This coding structure uses the input source frames, not the *reconstructed frames*, to predict the current frame. It is often referred as "open-loop structure" in the 3D wavelet coding literature [12]. It is very difficult to precisely control the prediction accuracy during the encoding process. Moreover, multiple bitstreams are to be extracted from the same coded bitstream. It is hard to adequately allocate the motion information bitrates at encoder (before the extractor) to satisfy all target operation points simultaneously. A theoretical treatment on the optimum trade-off between the motion information bitrate and the texture signal bitrate for a motion-compensated video codec was earlier explored by Girod [15] and will be discussed in the next section. In practice, most existing scalable wavelet video coding schemes still adopt the cost functions used in the hybrid video coding ((2) and (3)), but the Lagrange parameter in each temporal decomposition stage is manually selected empirically [25]. Because the target bitrate is given after the entire bitstream is coded, the pre-selected, fixed-value Lagrange parameter must be working for a range of bitrates. In other words, we hope it can provide a reasonable overall performance for all the bitrates of interest. The cost function defined by (2) determines the best motion prediction mode. If a

total bitrate is given, we can follow the conventional approach to pick up the Lagrange parameter. But unfortunately, the bitrate is not known at the encoding stage for scalable wavelet video encoding.

To go one step further, we look into the role that the motion vectors play in scalable interframe wavelet coding. The MCTF unit performs the temporal decomposition operation along the motion trajectory; therefore, the accuracy of motion vectors is critical to their motion compensation performance. The low-pass frames produced by temporal filtering will be further decomposed at the next temporal level. Thus, the temporal decomposition layers form a hierarchical structure. The inefficiency in motion prediction propagates along the temporal hierarchy in the same GOP. Therefore, accurate motion vectors tend to decrease the overall distortion. But, a very accurate motion vector often requires more coding bits.

To sum up, the Lagrangian cost function is a very powerful tool in the conventional non-scalable coder. But due to the open-loop coding structure and the requirement of multiple operating points, the use of the Lagrangian cost function in scalable wavelet video coding becomes inadequate. The key problem is finding the proper trade-off between the motion information and the residual texture information for scalable wavelet video coder. The whole scenario becomes even more complicated when we consider the propagation of MCTF inefficiency along temporal hierarchy. Therefore, we propose another approach to

replace the ordinary Lagrangian cost function for scalable wavelet video coding.

Chapter 3 ρ-GGD Source Modeling for Wavelet Coefficients

2-D Image signal can be decomposed twice by a 1-D discrete wavelet transform (DWT) into a 2-D multi-resolution representation. Each 1-D DWT splits the 2-D image signal into low-pass (L) and high-pass (H) subbands along the vertical or the horizontal direction. Typically, the LL subband is further split several times in image coding. In an interframe wavelet video coding structure, another wavelet filter bank is applied along the motion trajectory of moving objects [7]. The temporal L frame is often a moving average of frames, while the temporal H frame contains the frame differences. In video coding, these temporal L and H frames are further decomposed by the spatial 2-D DWT, so all original frames in a GOP are transformed to a temporal-spatial subband representation.

For either image or video coding, the source modeling is critical in the R-D analysis. The pdf (probability density function) of wavelet coefficients has been modeled as a generalized Gaussian distribution (GGD) [26][27]. To construct a GGD source model, the pdf variance and kurtosis have to be calculated first in order to estimate the shape parameter.

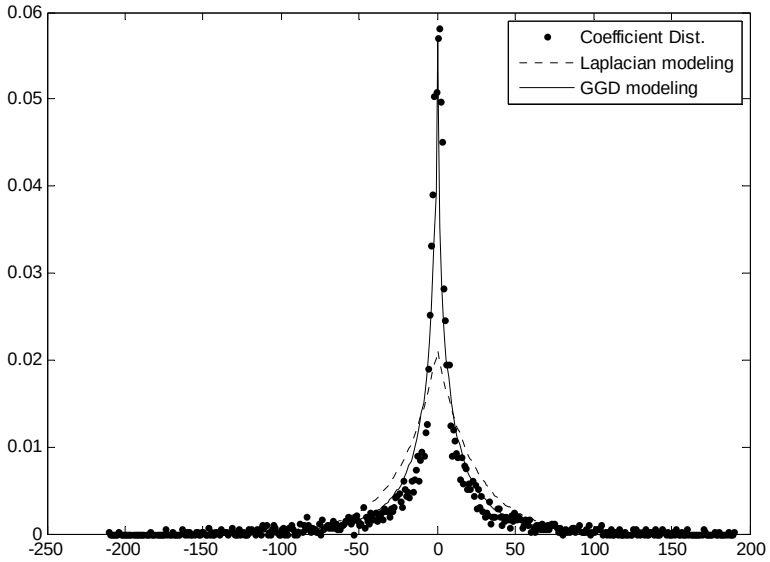


Fig. 3-1 An example of wavelet coefficients modeling. (LL-LL-HL subband of image Lena).

Variance and kurtosis are related to the second and the forth moments. Therefore, the process of constructing a GGD model is rather complicated. To reduce the complexity, the Laplacian distribution is often adopted. Although the Laplacian source model is thus widely used, its coefficients approximation errors are sometimes high as shown in Fig. 3-1. Therefore, we propose a ρ-GGD source model in the next section to achieve the high accuracy of the GGD model but with lower complexity.

3.1 ρ-GGD Source Model Derivation

The pdf of wavelet coefficients typically has zero-mean. Thus, the generalized Gaussian distribution (GGD) source model is given by

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

where

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

Here, σ is the standard deviation of wavelet coefficients, *v* is the shape parameter of the GGD model, and Γ(*x*) is the standard Gamma function. Let ρ be the *probability of zero-value coefficients*. According to(4), ρ is given by

$$\rho \triangleq \alpha \cdot \frac{\eta(\alpha,\sigma)}{2\cdot\Gamma(\alpha^{-1})} = P(0), \tag{6}$$

Therefore, (4) can be rewritten by the following ρ-GGD representation:

$$P_{\rho-GGD}(x) = \rho\cdot\exp\left(-\left(2\cdot\rho\alpha^{-1}\Gamma(\alpha^{-1})\cdot x\right)^{\alpha}\right), \tag{7}$$

In building a ρ-GGD source model, the shape parameter *v* has to be estimated first. From (5) and (6), the product of ρ and σ can be written as

$$\Phi(\alpha) \triangleq \frac{\alpha}{2}\cdot\sqrt{\frac{\Gamma(3\alpha^{-1})}{\Gamma(\alpha^{-1})^{3}}}, \tag{8}$$

(8) shows a mapping relationship between the shape parameter *v* and the product of ρ and σ in the ρ-GGD model; that is, $\rho\sigma = \Phi(\alpha)$. Because parameters ρ and σ can easily be obtained from data, it is convenient to use their product to estimate the value of $\Phi(\alpha)$.

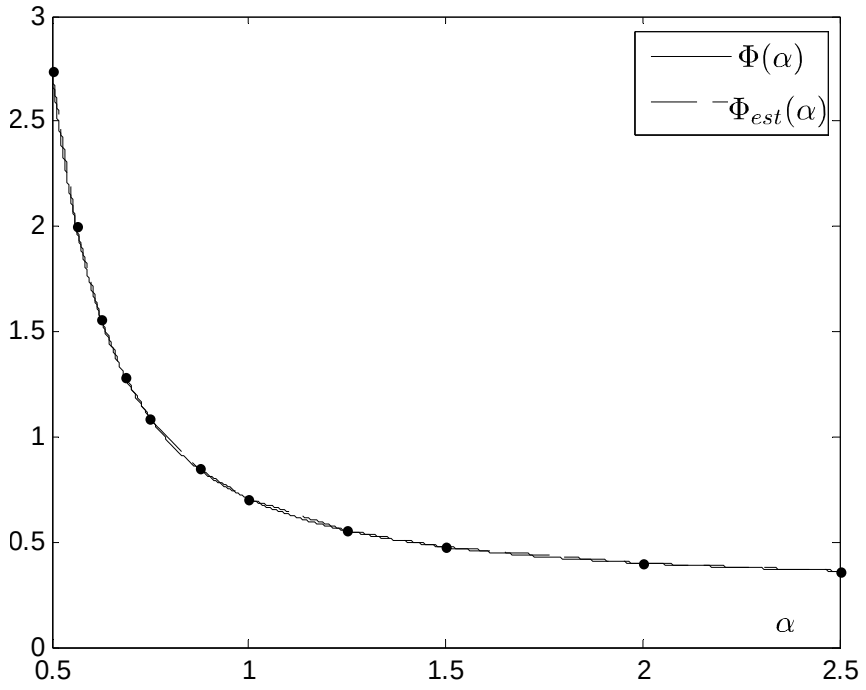


Fig. 3-2 $\Phi(\alpha)$ at $\alpha \in [0.5, 2.5]$ and its piecewise linear approximation.

From experiments, the range of α is [0.5, 2.5] for typical image/video wavelet coefficients. In Fig. 3-2 , the solid line shows the values of $\Phi(\alpha)$ in the range of $\alpha \in [0.5, 2.5]$, an decreasing one-to-one function of α. Therefore, the inverse function of $\Phi(\alpha)$ at $\alpha \in [0.5, 2.5]$ exists and is unique; thus, the shape parameter α can be estimated from $\Phi^{-1}(\rho\sigma)$.

3.2 Piecewise Linear Estimation for the Shape Parameter of Wavelet Coefficients

In Fig. 3-2, $\Phi(\alpha)$ is an exponentially decreasing smooth curve. We found experimentally that $\Phi(\alpha)$ can be approximated accurately for $\alpha \in [0.5, 2.5]$ by piecewise linear approximation. We partition the $\Phi(\alpha)$ curve into ten pieces for $\alpha \in [0.5, 2.5]$. For each piece at $\alpha \in [f_i, f_{i-1}]$, $\Phi_{est}(\alpha)$ is approximated by a linear model as below

$$\Phi_{est}(v) = \frac{\Phi(f_i) - \Phi(f_{i-1})}{f_i - f_{i-1}}(v - f_i) + \Phi(f_i), \tag{9}$$

where i={1,2,…10} and $\{f_0, f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_{10}\}$= {0.5, 0.5625, 0.625, 0.6875, 0.75, 0.875, 1, 1.25, 1.5, 2, 2.5}. Fig. 3-2 shows that $\Phi(\alpha)$ is well approximated by $\Phi_{est}(\alpha)$. And the shape parameter can be estimated by $\alpha_{est} = \Phi_{est}^{-1}(\rho\sigma)$, which is

$$\Phi_{est}^{-1}(\rho\sigma) = \frac{f_i - f_{i-1}}{\Phi(f_i) - \Phi(f_{i-1})}(\rho\sigma - \Phi(f_i)) + f_i, \tag{10}$$

Table 3-1. Look-up table for shape parameter estimation

i	S_i	$\frac{\Phi(f_i)-\Phi(f_{i-1})}{f_i - f_{i-1}}$	$\Phi(f_i)$	f_i
1	[2.739, 2.000]	-11.810	2.000	0.5625
2	[2.000, 1.563]	-7.005	1.563	0.6250
3	[1.563, 1.281]	-4.506	1.281	0.6875
4	[1.281, 1.089]	-3.080	1.089	0.7500
5	[1.089, 0.848]	-1.926	0.848	0.8750
6	[0.848, 0.707]	-1.126	0.707	1.0000
7	[0.707, 0.555]	-0.610	0.555	1.2500
8	[0.555, 0.476]	-0.314	0.476	1.5000
9	[0.476, 0.399]	-0.154	0.399	2.0000
10	[0.399, 0.363]	-0.073	0.363	2.5000

when

$$\rho\sigma \in S_i = [\Phi(f_{i-1}), \Phi(f_i)], \tag{11}$$

Furthermore, a look-up table of the constants used in (10) and (11) can be pre-calculated as shown in Table 3-1. In conclusion, a ρ-GGD model for the pdfs of wavelet coefficients can be constructed by using the following steps:

Step 1: Compute ρ and σ from the wavelet coefficients.

Step 2: Use Table 3-1 to get S_i in (11) based on the product of ρ and σ and also the corresponding model coefficients.

Step 3: Calculate the estimated shape parameter α_{est} from (10) using the model coefficients obtained in Step 2.

Step 4: Obtain the ρ-GGD source model given by (7).

3.3 Modeling Accuracy Evaluation

The difference between two probability distributions can be evaluated by estimating the relative entropy or the said Kullback- Leibler (K-L) divergence [28]. In this thesis, we use the symmetric definition defined by

$$KL(p \parallel q) = \sum_{x \in X} p(x) \log_2 \left(\frac{p(x)}{q(x)} \right) + \sum_{x \in X} q(x) \log_2 \left(\frac{q(x)}{p(x)} \right) , \tag{12}$$

where p is the "true" pdf and q is the "modeling" pdf. A small K-L divergence means a higher modeling accuracy. The experimental results of the ρ-GGD and the Laplacian modeling are compared for several test cases. In the spatial 2-D DWT case, the Daubechies 9/7 biorthogonal wavelet filter [29] popular in image coding is adopted in our experiments. Fig. 3-3 shows the pdfs of the spatial subband coefficients and their models for the test image "Pepper". It is clear that the ρ-GGD model matches the real pdf much better than the Laplacian model in all spatial subbands. Table 3-2 shows the divergence of our model and the real pdf by using the symmetric K-L divergence for the test image "Lena". In general, the ρ-GGD model outperforms the Laplacian significantly in all subbands except for the LH subband.

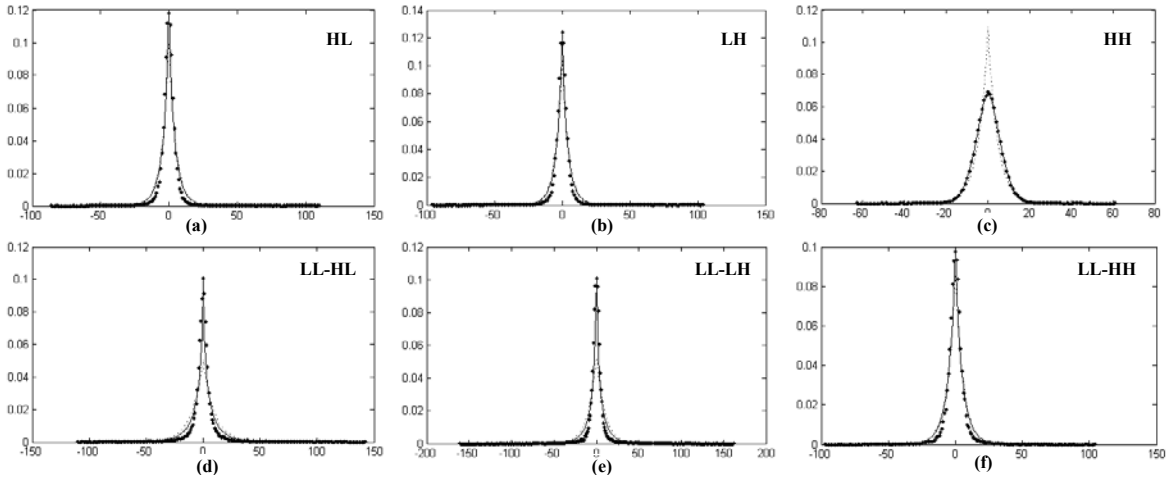


Fig. 3-3 The pdfs of wavelet coefficients (dots) and their approximations by Laplacian (dotted line) and the proposed ρ-GGD (solid line) models in the subbands: (a) HL, (b) LH, (c) HH, (d) LL-HL, (e) LL-LH, (f) LL-HH. The test image is "Pepper".

In the interframe wavelet video case [10], the temporal-spatial subband coefficients are produced by using MCTF and the spatial 2-D DWT. Table 3-3 shows the results of the K-L divergence estimation of two test sequences, "Bus" and "Mobile", with GOP=8 and 16 respectively at CIF resolution. The ρ-GGD model shows a much better modeling accuracy than the Laplacian. In general, the higher subband signals are difficult to model but the ρ-GGD model shows good accuracy in Table 3-3 (b) even at deep temporal subbands.

From the experimental results, the ρ-GGD model shows a very good modeling performance in both spatial 2-D DWT and interframe wavelet video cases. Compared to the Laplacian model, the ρ-GGD has a much better accuracy and consistency in modeling the pdfs of wavelet coefficients.

Table 3-2. K-L divergences of two source models for the 2-D DWT coefficients for image “Lena”.

Band Index / Model	HL	LH	HH	LL-HL	LL-LH	LL-HH
Laplacian	0.22	0.07	0.03	0.68	0.52	0.42
ρ-GGD	0.16	0.08	0.03	0.22	0.21	0.20

Table 3-3. K-L divergence comparison of two source models for temporal-spatial subband coefficients.

Band Index / Model	Temporal Level 2 (H-frame)				Temporal Level 3 (H-frame)			
	LL	HL	LH	HH	LL	HL	LH	HH
Laplacian	0.89	0.38	0.30	0.08	0.70	0.33	0.27	0.10
ρ-GGD	0.28	0.19	0.09	0.07	0.22	0.14	0.07	0.07

(a) “Bus” with GOP=8.

Band Index / Model	Temporal Level 4 (H-frame)				Temporal Level 5 (L-frame)		
	LL	HL	LH	HH	HL	LH	HH
Laplacian	0.85	0.33	0.30	0.19	0.47	0.46	0.38
ρ-GGD	0.20	0.05	0.03	0.02	0.06	0.10	0.07

(b) “Mobile” with GOP=16.

Chapter 4 Motion Information Gain (MIG) and Mode Decision Method

A typical extraction process in scalable wavelet coding truncates only the encoded texture bitstream and maintains the integrity of the entire encoded motion information. For a given bitrate condition, different amounts of motion information lead to different types of residual texture signals, and thus lead to different rate-distortion behavior. Although there are other approximate solutions [29], [31] that select the scalable motion information to match certain very low-bitrate requirements, we focus on the pre-partitioned motion information solution in the following study. That is, the optimal amount of information bits is decided at the encoding stage. We first analyze the rate-distortion behavior of the motion-predicted residual signals. Then, based on this rate-distortion relationship, we derive a quantitative metric that measures the coding efficiency of motion information. Also, a theoretical explanation from the entropy viewpoint is given to our coding efficiency metric.

4.1 Rate-Distortion Model of Motion-Compensated Prediction

For a scalable wavelet video coder, theoretically, we can fix an extraction bitrate and then find the rate-distortion behavior due to the increase/decrease of motion information. In other words, at a given bit rate, if a certain amount of the texture bit rate is shifted to the motion information, will the reconstructed image distortion be reduced or increased? A solution to this problem is searching for the optimal motion information that leads to the optimal R-D performance at different bit rates. For example, is the block size or the motion vector accuracy more important in improving the coded image quality? Clearly, the answer depends on both picture content and bit-rate.

Although the residual frames after MCTF will be further spatially decomposed by 2-D DWT, in this study we focus on the rate-distortion behavior of the texture information at the MCTF stage (not after 2-D DWT) because the motion information coding efficiency is our main concern. Because the consecutive frames are often very similar, the motion-predicted residual signals typically have zero-mean and nearly symmetrical distribution. The residual signals after motion prediction can be modeled as Laplacian sources. Because the temporal high-pass frame is essentially a weighted combination of the motion-predicted residual frames, we next try to construct the rate-distortion model of the motion-compensated residual signals.

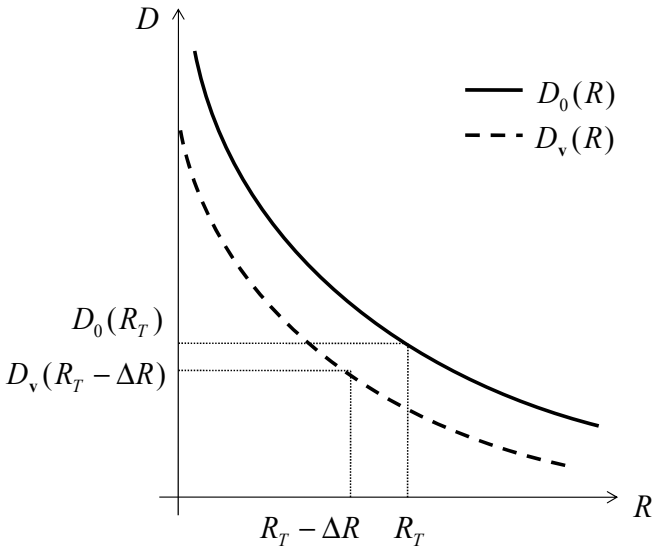


Fig. 4-1 Illustration of rate-distortion curves of texture residual signal before and after motion prediction.

When the residual texture signal is produced by the motion prediction operation, the rate-distortion behavior of this texture information portion is decided. That is, since the residuals are fixed after motion compensation, their rate and distortion trade-off due to quantization and entropy coding is also fixed. However, if we change the motion vectors (mv) used in motion prediction, the residual signals are different and thus, the texture rate-distortion function changes. We like to know the texture rate-distortion function variation before and after the motion prediction being applied to the same coding block.

For a motion-compensated video codec, Girod [15] pointed out that at a given total bit rate, the optimum trade-off point should locate at

$$\frac{\partial D}{\partial R_{texture}} = \frac{\partial D}{\partial R_{mv}}, \tag{13}$$

where the left hand side is the distortion decrease due to texture rate increase and the right

hand is the distortion decrease due to motion information rate increase. Fig. 4-1 gives an illustration of this principle. We use the zero motion vector (no motion-compensation) case as a reference. In Fig. 4-1, $D_0(R)$ is the rate-distortion function of the residual signal produced by using the zero motion vector, and $D_{\mathbf{v}}(R)$ is the rate-distortion function of the residual signals produced with the motion vector set **v**. From the bitrate viewpoint, an extra coding bitrate ΔR is needed for sending the motion vectors **v**. Since the total target bitrate R_T is given, the bitrate available for the texture information is reduced to R_T -ΔR. If this set of mv is beneficial for the overall performance, the quantization error (distortion) of the texture information with mv should be less than that without mv at the same target bitrate. Otherwise, the motion compensation is judged inefficient. Therefore, the distortion with motion prediction is smaller than that without motion prediction:

$$D_v(R_T - \Delta R) < D_0(R_T). \tag{14}$$

Conceptually, (14) is equivalent to (13) in [15]. But different from the motion region partition approach in [15], we try to find an instrumental trade-off measure and a design procedure for adjusting the mv bit rate.

For the Laplacian source described by (1), if the absolute-error distortion measurement is in use, (14) can be rewritten using the rate-distortion functions given in [18] as

$$\frac{1}{\Lambda_{\mathbf{v}}} \cdot 2^{-(R_T-\Delta R)} < \frac{1}{\Lambda_{\mathbf{0}}} \cdot 2^{-R_T}. \tag{15}$$

The Laplacian parameter $\Lambda_{\mathbf{v}}$ and $\Lambda_{\mathbf{0}}$ can be estimated from the residual signal variances,

$\sigma_{\mathbf{v}}^2$ and $\sigma_{\mathbf{0}}^2$, respectively. That is, $\Lambda = \sqrt{2}/\sigma$. Thus, (15) becomes

$$\frac{\log_2(\sigma_{\mathbf{0}}) - \log_2(\sigma_{\mathbf{v}})}{\Delta R} > 1. \tag{16}$$

Let us define the function Φ to be the logarithm value of the signal standard deviation, and let ΔΦ be

$$\Delta\Phi \equiv \Phi_{\mathbf{0}} - \Phi_{\mathbf{v}} = \log_2(\sigma_{\mathbf{0}}) - \log_2(\sigma_{\mathbf{v}}). \tag{17}$$

Then, (16) can be rewritten as

$$\frac{\Delta\Phi}{\Delta R} > 1. \tag{18}$$

From (14) to (18), we can see that the target bitrate term R_T is cancelled because it appears on both sides in (15). This target bitrate elimination gives us a big advantage in the rest of our rate-distortion analysis. Different from the conventional video coding, the target (extraction) bitrate is unknown during the scalable encoding process. In this formulation, the measurement of motion prediction efficiency is extraction bitrate irrelevant. This is true under the assumption that the residual signal probability distribution is Laplacian for both with and without motion-compensated prediction. This Laplacian model is not all accurate in real cases. Here, ΔΦ and ΔR represent the variation of texture statistics and the bitrate cost of adopting motion estimation, respectively. We thus view ΔΦ/ΔR as a *gain* factor in measuring the motion prediction efficiency. Intuitively, the motion prediction operation is preferred if it reduces the texture variance significantly. Furthermore, (18) gives a quantitative metric and specifies a threshold of acceptable ΔΦ/ΔR. This threshold is derived

based on the Laplacian source assumption with absolute-error distortion definition.

4.2 Motion Information Gain (MIG)

According to the last sub-section, ΔΦ represents the variation of texture statistics due to motion-compensated prediction. We are going to show next that ΔΦ represents the difference between two differential entropies. For the Laplacian source X, its differential entropy $h(X)$ is given below [18].

$$\begin{aligned} h(X) &= -\int_X P(X)\log_2(P(X))dx \\ &= -\int_{-\infty}^{\infty} \frac{\Lambda}{2}e^{-\Lambda|x|}\cdot\log_2\left(\frac{\Lambda}{2}e^{-\Lambda|x|}\right)dx \\ &= 1+\log_2\left(\frac{e}{\Lambda}\right) \end{aligned}, \tag{19}$$

where Λ is the Laplacian parameter. Thus, the differential entropies of the residual signals $X_{\mathbf{0}}$ and $X_{\mathbf{v}}$ produced by the zero motion vector and the motion vector set $\mathbf{v}$ are, respectively,

$$\begin{aligned} h(X_{\mathbf{0}}) &= 1+\log_2\left(\frac{e}{\Lambda_{\mathbf{0}}}\right) \\ h(X_{\mathbf{v}}) &= 1+\log_2\left(\frac{e}{\Lambda_{\mathbf{v}}}\right). \end{aligned} \tag{20}$$

Although the differential entropy does not represent the actual bitrate, the difference between two differential entropies represents the bitrate difference estimation of these two sources. Since the Laplacian parameter can be estimated from the signal variance, we thus obtain the following equation:

$$h(X_{\mathbf{0}})-h(X_{\mathbf{v}})=\log_2\left(\frac{\Lambda_{\mathbf{v}}}{\Lambda_{\mathbf{0}}}\right)=\log_2\left(\frac{\sigma_{\mathbf{0}}}{\sigma_{\mathbf{v}}}\right). \tag{21}$$

Comparing (21) with (17), as a consequence of rate-distortion theory on the Laplacian source, we find that these two equations are the same. Therefore, ΔΦ represents the reduction of residual signal entropy in encoding the residual signals before and after motion-compensated prediction. Thus, the interpretation of ΔΦ/ΔR is as follows.

$$\frac{\Delta\Phi}{\Delta R}\sim\frac{\text{decrease in residual signal entropy}}{\text{increase in motion information bitrate}}. \tag{22}$$

From (22), we can see that $\Delta\Phi/\Delta R$ is the ratio of the "reward" and the "cost" due to the use of motion-compensated prediction. The "cost" is the extra bitrate for encoding the motion vectors, and the "reward" is the entropy reduction of the residual texture signals. Therefore, $\Delta\Phi/\Delta R$ is called the "motion information gain", abbreviated as MIG. It is thus used to measure the motion prediction efficiency. We denote this MIG function due to the motion vector set **v** by

$$\phi(\mathbf{v}) \triangleq \frac{\Delta\Phi}{\Delta R} \ . \tag{23}$$

This gain factor implicitly represents the trade-off between the residual signal bitrate and motion information bitrate. The fundamental concept behind (23) is similar to that (13) in [15] as discussed earlier. But through our preceding lengthy derivation, we show that the total target bitrate disappears in the final MIG expression. Thus, the MIG metric fits well for applying to the scalable wavelet video coding structure.

Let us extend the original criterion (18) to a more general form. When we consider the advantage of using motion- prediction in scalable wavelet video coding, the MIG metric of the candidate motion vector set **v** should satisfy

$$\frac{\Delta\Phi}{\Delta R} > C, \tag{24}$$

where C is a chosen threshold value. In the original derivation, C is 1. Here we investigate the range of C values in real video coding cases. Because a practical entropy coder cannot approach the entropy bound, both the compressed texture and the compressed motion information would need more bits to code. Therefore, the motion prediction is not as effective as (14) shows. The distortion reduction by the motion bitrate ΔR, measured in bits/pixel, is less than the expected value; that is, $D_{\mathbf{v}}$ should be larger in real cases. Therefore, (14) is modified to

$$a \cdot D_v(R_T - \Delta R) < D_0(R_T), \tag{25}$$

where $a > 1$. Using the above equation, we can follow the same derivation process in section III.A to obtain the MIG lower bound. Consequently, an inequality similar (16) is derived:

$$\frac{\log_2(\sigma_{\mathbf{0}}) - \log_2(\sigma_{\mathbf{v}})}{\Delta R} > 1 + \frac{\log_2(a)}{\Delta R}. \tag{26}$$

Because $a > 1$, the right term of the above equation, the lower bound of *C*, is larger than 1. When ΔR is small or $\log_2(a)$ is large, *C* becomes much larger than 1.

4.3 MIG Cost Function

Since our motion mode and vector selection process is applied only to image blocks with non-zero optimal motion vectors, the denominator of (23) is non-zero. There are a few interesting properties associated with $\phi(\mathbf{v})$.

1) $\phi(\mathbf{v}) \geq 0$. Clearly, we will not use an mv that produces a negative ΔΦ value. For a given image block, if the zero mv is the best mv in the sense that any non-zero mv cannot reduce the residual signal variance, then the $\phi(\mathbf{v})$ value associated with this block is assigned to be 0 and the best coding mode is the one with the zero motion vector.

2) $\phi(\mathbf{v})$ is bounded. In digital image coding, the residual signal has a finite variance. The best non-zero mv can, at the best, reduce the residual variance to zero. The variance difference before and after employing mv is thus finite. In other words, the $\phi(\mathbf{v})$ value saturates and cannot be further improved when a proper mv is identified.

3) In the following sections, we deal mainly with the case that $\phi_{max} \geq \phi(\mathbf{v}) > C$. That is, the useful mv, **v**, should produce a $\phi(\mathbf{v})$ value greater than 0 and less than or equal to ϕ_{max}. Ideally, the parameter *C* is 1 and is independent of image contents and target bit rate if the Laplacian rate-distortion model holds. However, as discussed earlier, practically *C* is not 1 and is bitrate dependent.

Intuitively, the MIG metric $\phi(\mathbf{v})$ with the constraint, $\phi_{max} \geq \phi(\mathbf{v}) > C$, can be the cost function used for searching for the optimal mv. However, the *C* value is unknown and to be

identified in real image coding. Thus, for the convenience in computation, we use the following equivalent form. We expand (24) with the aid of (17) and (23). The inequality becomes

$$\sigma_{\mathbf{0}}^2 > \sigma_{\mathbf{v}}^2 \cdot 2^{2 \cdot C \cdot \Delta R}. \tag{27}$$

A large MIG value implies a large ΔΦ and/or a small Δ*R*. In (17), a large ΔΦ value implies that the difference between $\sigma_{\mathbf{0}}$ and $\sigma_{\mathbf{v}}$ is large. Thus, the right term in (27), $\sigma_{\mathbf{v}}^2 \cdot 2^{2 \cdot C \cdot \Delta R}$, should be as small as possible. Therefore, we propose a so-called "MIG cost function" to measure the prediction cost. For a coding source **s**, the motion vector set **v** produces the residual signals with variance $\sigma_{\mathbf{s}}^2(\mathbf{v})$ and its average information bitrate (for representing **v**) is Δ*R*(**v**). The MIG cost function *J* is defined as

$$J(\mathbf{s}, \mathbf{v}|C) = \sigma_{\mathbf{s}}^2(\mathbf{v}) \cdot 2^{2 \cdot C \cdot \Delta R(\mathbf{v})}, \tag{28}$$

where *C* is generally source and bit-rate dependent. We include it explicitly in the argument of the *J* function to emphasize its role in our rate control algorithm. The problem now becomes looking for **v** that minimizes *J*.

We need to identify the value of *C* in (28). According to our previous discussions, the *C* value is decided by the coding system and the source signal **s** in (23). In practice, the source signal **s** is the temporal high-pass frames generated by MCTF. Indeed, the probability distributions of the different temporal layers have different shapes [32]. We conduct the following experiments to characterize *J* and also to identify the value of *C*.

We start with a fixed C value and simply use (28) as the cost function in performing motion estimation and mode decision in encoding. The detailed procedure of mode decision will be described in the next sub-section. After the encoding process is done, the encoded bitstream is truncated to a fixed bitrate, for example, 256kbps, and then we decode the truncated bitstream. The mean-squared error (MSE) between the decoded and the original images is calculated; thus, one test point of a MSE and C pair is obtained. The data are collected from 32 frames of the Mobile sequence at CIF resolution.

Repeating the above steps with different C values, we obtain a MSE vs. C curve at 256Kbps as shown in Fig. 4-2 (a). By changing the truncation bitrates settings, the MSE vs. C curves at 384Kbps and 800Kbps are obtained as shown in Fig. 4-2 (c) and (d) respectively. Each of Fig. 4-2 (a)(b)(c) shows that the MSE is minimal when C reaches a certain value. This is equivalent to the performance saturation phenomenon we discuss earlier. When C is large, only the very effective mv's can make positive contribution and their value is diminishing as C gets larger; and thus the MSE goes up again as shown in Fig. 4-2 (a)(b)(c). Although the theory predicts that MIG is independent of bit rate, in reality, however, the coding system efficiency and the source probability distribution are bitrate and temporal level dependent. Indeed, the best C value that leads to the minimum MSE tends to be smaller at higher bitrates. This is consistent with the known observation that the mathematical model matches the real rate-distortion relationship at higher rates. For

example, the rate-distortion relationship of a quantizer approximates the asymptotical R-D function at high bitrates [18]. If the optimum C value does not change significantly, we prefer to use a constant C to cover the bitrates of our interests. We pick up seven target bitrates, 256k, 384k, 512k, 800k, 1024k, 1200k, and 1500k, and their average behavior (MSE vs. C) is shown in Fig. 4-2 (d). In conclusion, the C value generally falls in the range of [4, 10].

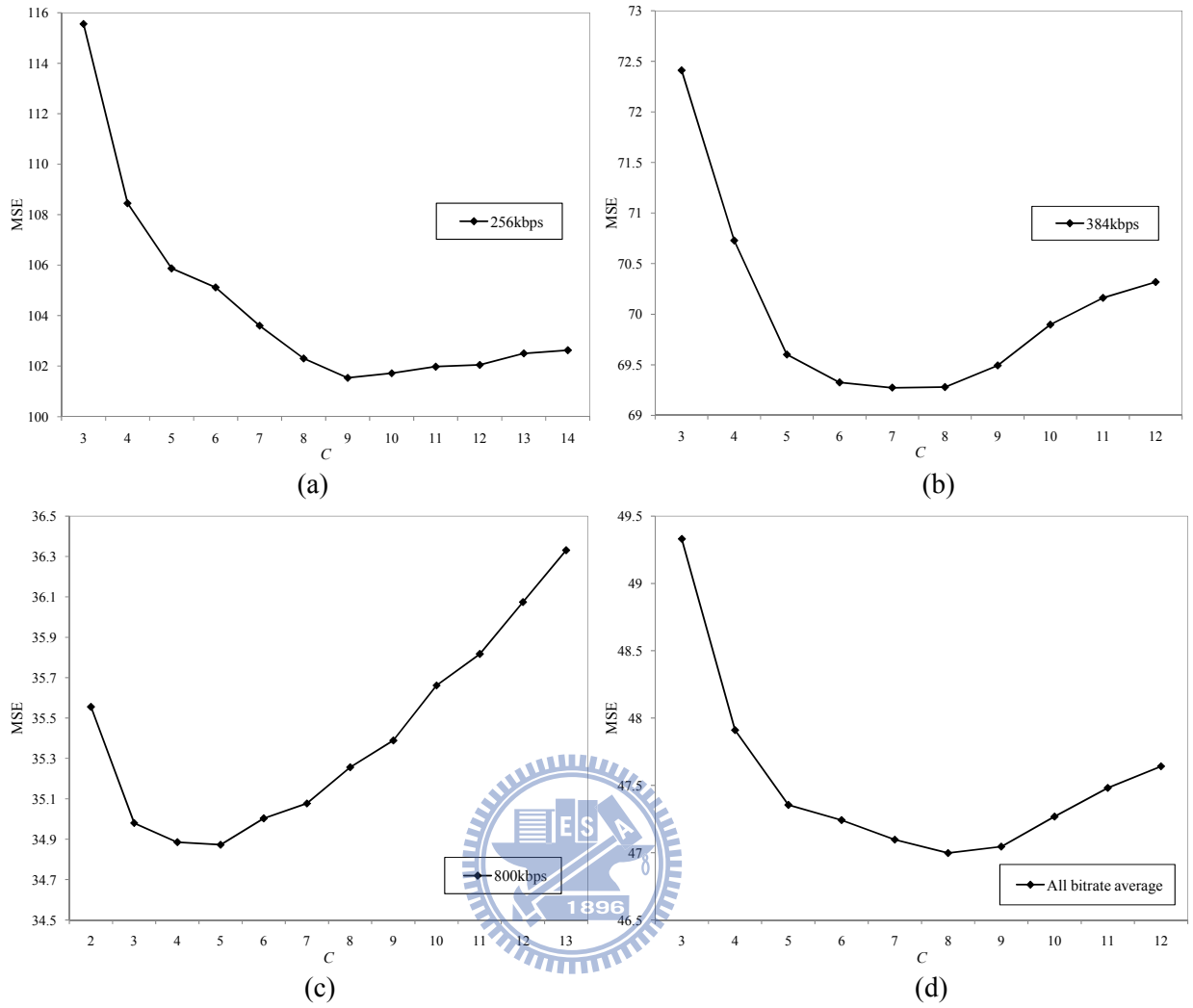


Fig. 4-2 MSE vs. *C* value in the MIG cost function at (a) 256Kbps, (b) 284Kbps, and (c) 800Kbps truncation bitrates, and (d) the average MSE for 7 bitrates. (Mobile, CIF resolution).

4.4 Block-Based Mode Decision Procedure

The MIG cost function can be used to decide the coding mode. It tells us the trade-off between the motion information and the texture information. Based on MIG, we develop a mode decision procedure. In a conventional non-scalable video coder, the best motion vector and coding mode are decided by minimizing the Lagrangian cost function ((2) and (3)) for a given single bitrate. As discussed in the previous sub-sections, with the MIG cost function we are able to choose the most appropriate coding mode (including mv) by minimizing its value. The basic steps in the proposed mode decision procedure are similar to that in the conventional scheme. In the existing scalable wavelet video coding schemes, the mv search is block-based and the variable block-size motion compensation technique is used to find the best macroblock coding mode. Each macroblock coding mode represents a partition of macroblock into a certain combination of sub-blocks. Fig. 4-3 illustrates the proposed mode decision procedure, which consists of three steps as described below.

1) Step 1: Select the appropriate MIG cost function parameters

The proposed MIG cost function contains one parameter, *C*. According to our previous discussions, *C* can be empirically chosen from the intervals, [4, 10]

2) Step2: Search for the best motion vector set for each block mode

There are many possible sub-block combinations for motion compensation in one macroblock. For example, a typical 16x16 size macroblock has 16x16, 16x8, 8x16, and

8x8 block modes; and each 8x8 block can be further partitioned to 8x4, 4x8, and 4x4 sub-blocks. Assuming that a macroblock can be partitioned to N_m sub-blocks for mode m, the mv's ($\mathbf{v}_i$) associated with all sub-blocks (b_i) form two N_m-tuple vectors, $\mathbf{v}_m$ and $\mathbf{b}_m$, respectively, where

$$\begin{aligned} \mathbf{v}_m &= (\mathbf{v}_1, \cdots, \mathbf{v}_{N_m}) \\ \mathbf{b}_m &= (b_1, \cdots, b_{N_m}) \end{aligned}. \tag{29}$$

For each sub-block, to find the best mv, all the mv candidates within the search range **S** are examined. These candidate motion vectors can have forward, backward or bi-directional prediction directions. By minimizing the MIG cost function in (28), the best motion vector v_i^* for sub-block b_i is obtained. Mathematically, it is identified by performing the following optimization procedure.

$$\begin{aligned} v_i^* &= \arg\min_{v \in \mathbf{S}} \{J_{Motion}(b_i, v|C)\} \\ \text{with } J_{Motion}(b_i, v|C) &= \sigma_{b_i}^2(v) \cdot 2^{2 \cdot C \cdot \Delta R(v)} \end{aligned}. \tag{30}$$

Then, the best mv for the macroblock is the collection of all the best motion vectors for mode m; i.e.,

$$\mathbf{v}_m^* = (v_1^*, \ldots, v_{N_m}^*). \tag{31}$$

The residual signal is modeled as a Laplacian source with zero-mean. After all the sub-blocks finish the motion estimation process for mode m, the residual variance $\sigma_{\mathbf{b_m}}^2(\mathbf{v}_m^*)$ and the average motion information bitrate $\Delta R(\mathbf{v}_m^*)$ of a macroblock can be, respectively, expressed as

$$\sigma^2_{\mathbf{b}_m}(v^*_m) = \frac{1}{N_m}\sum_i^{N_m}\sigma^2_{b_i}(v^*_i)$$
$$\Delta R(v^*_m) = \frac{1}{N_m}\sum_i^{N_m}\Delta R(v^*_i) + r_m \quad , \tag{32}$$

where r_m is the average extra bits needed to record the coding mode information. Both ΔR and r_m are in bits/pixel.

3) *Step 3*: Choose the best block mode with the minimum MIG cost

Assuming that the block mode *m* in *Step 2* belongs to the mode set **M,** which contains all possible block modes, the MIG cost function in (28) is used again to choose the best macroblock mode. Hence, the best block mode is decided by minimizing the MIG cost function:

$$m^* = \arg\min_{m\in\mathbf{M}}\{J_{Mode}(\mathbf{b}_m, \mathbf{v}^*_m|C)\}$$
$$\text{with } J_{Mode}(\mathbf{b}_m, \mathbf{v}^*_m|) = \sigma^2_{\mathbf{b}_m}(\mathbf{v}^*_m)\cdot 2^{2\cdot C\cdot\Delta R(\mathbf{v}^*_m)} \tag{33}$$

Therefore, the best block mode and its associated motion vectors of a macroblock are obtained.

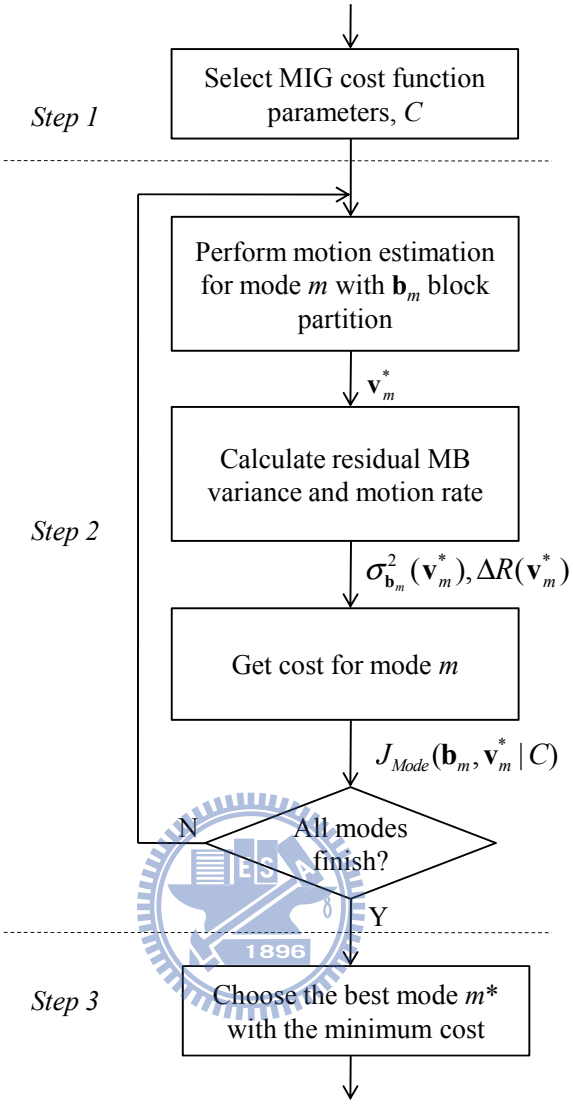


Fig. 4-3 Flow chart of the proposed mode decision procedure using the MIG cost function

Chapter 5 One-Sided ρ-GGD Source Modeling for Residual Signals

In the study of motion estimation efficiency, an accurate source model on the motion-compensated residual signal is critical and essential. The results in [32] 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 SSD (Sum of Squared Difference), we propose the so-call 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.

5.1 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 [27]. The GGD pdf is given by

$$P(x) = \frac{1}{2}\left(\frac{\alpha \cdot \eta(\alpha,\sigma)}{\Gamma(\alpha^{-1})}\right)\exp\left(-[\eta(\alpha,\sigma)\cdot x]^{\alpha}\right), \tag{34}$$

where

$$\eta(\alpha,\sigma) = \sigma^{-1}\sqrt{\frac{\Gamma(3\alpha^{-1})}{\Gamma(\alpha^{-1})}}, \tag{35}$$

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 (34) is a zero-mean and symmetric pdf and X is non-negative, we modify the GGD model to the one-sided GGD with the following pdf:

$$P(x) = \left(\frac{\alpha \cdot \eta(\alpha,\sigma)}{\Gamma(\alpha^{-1})}\right)\exp\left(-[\eta(\alpha,\sigma)\cdot x]^{\alpha}\right), \quad x \geq 0. \tag{36}$$

The shape parameter α in (36) can be estimated by using the variance and kurtosis of the source signal [27] but the complexity of this approach is very high. We will derive an

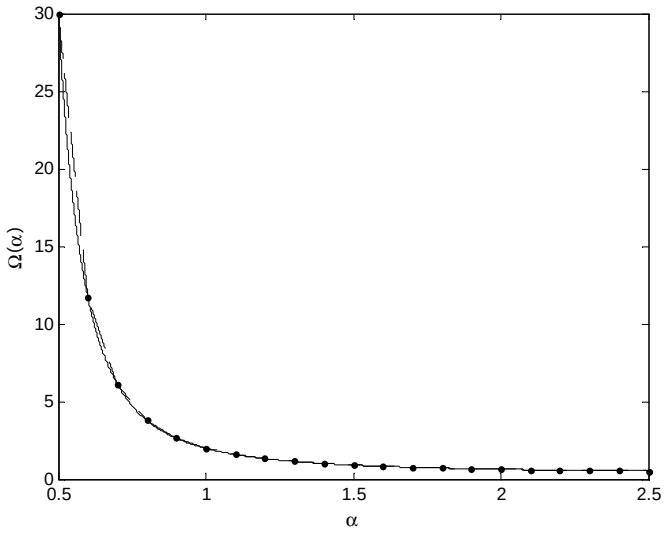


Fig. 5-1 The solid line and the dashed line are the curves of Ω(α) and its approximating function $\Omega_e(\alpha)$, respectively. $\Omega_e(\alpha)$ is made of 20 line segments in this example.

alternative expression that can be computed from the data samples with much less computation.

We denote the probability of zero in (36) by ρ. That is,

$$\rho \triangleq \alpha \cdot \frac{\eta(\alpha,\sigma)}{\Gamma(\alpha^{-1})} = P(0). \tag{37}$$

And then (36) can be rewritten as

$$P_{\rho-GGD}(x) = \rho \cdot \exp\left(-\left(\rho\alpha^{-1}\Gamma(\alpha^{-1})\cdot x\right)^{\alpha}\right), \quad x \geq 0. \tag{38}$$

We name (38) the one-sided ρ-GGD. There is an interesting property of the proposed one-sided ρ-GGD. From (35) and (37), the product of ρ^2 and σ^2 can be rewritten as

$$\rho^2\sigma^2 = \alpha^2 \cdot \frac{\Gamma(3\alpha^{-1})}{\Gamma(\alpha^{-1})^3} \ . \tag{39}$$

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

$$\Omega(\alpha) \triangleq \alpha^2 \cdot \frac{\Gamma(3\alpha^{-1})}{\Gamma(\alpha^{-1})^3} \quad . \tag{40}$$

This functional relationship is useful in estimating the shape parameter. As Fig. 5-1 shows, the mapping between $\Omega(\alpha)$ and α is one-to-one. Therefore, the inverse function of $\Omega(\alpha)$ exists. According to (39) and (40), α can be obtained by

$$\alpha = \Omega^{-1}(\rho^2\sigma^2). \tag{41}$$

Different from the conventional approach, we develop a new and fast method to estimate the shape parameter based on the expression of (41). That is, we use the zero-value probability and the variance value to estimate α.

5.2 Piecewise Linear Estimation of Shape Parameter of Residual Signal

Fig. 5-1 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 sets of boundary points: $\{\Omega(\alpha_0), \Omega(\alpha_1)\}$, $\{\Omega(\alpha_1), \Omega(\alpha_2)\}$ …,and $\{\Omega(\alpha_{n-1}), \Omega(\alpha_n)\}$. That is, $\Omega(\alpha)$ is approximated by a piecewise linear function $\Omega_e(\alpha)$. For the i-th segment,

$$\Omega_e(\alpha) = \frac{\Omega(\alpha_i) - \Omega(\alpha_{i-1})}{\alpha_i - \alpha_{i-1}}(\alpha - \alpha_i) + \Omega(\alpha_i) \ , \tag{42}$$

where $\alpha \in [\alpha_{i-1}, \alpha_i]$. Generally, the approximation is more accurate for large n. Fig. 5-1 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 (42) clearly has an inverse. We can thus estimate the shape parameter α_e using (41). If both ρ and σ^2 are known, then

$$\begin{aligned} \alpha_e &= \Omega_e^{-1}(\rho^2\sigma^2) \\ &= \left(\rho^2\sigma^2 - \left(\Omega(\alpha_i) - \frac{\Omega(\alpha_i) - \Omega(\alpha_{i-1})}{\alpha_i - \alpha_{i-1}} \cdot \alpha_i\right)\right) \Big/ \left(\frac{\Omega(\alpha_i) - \Omega(\alpha_{i-1})}{\alpha_i - \alpha_{i-1}}\right) , \end{aligned} \tag{43}$$

Table 5-1. A 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

for

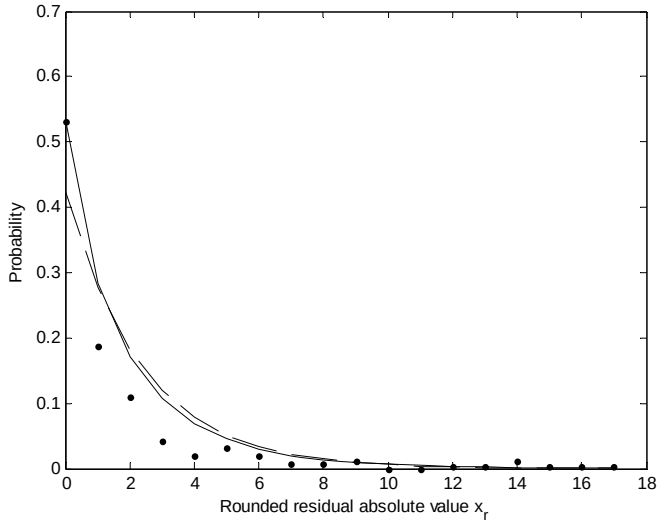
$$\rho^2\sigma^2 \in \left[\Omega(\alpha_{i-1}), \Omega(\alpha_i)\right] . \tag{44}$$

One may notice that the coefficients in (43) are independent of data and can thus be calculated in advance and recorded on a table. **Table 5-1** shows the example of n=20. Therefore, for the i-th line segment, the coefficients can be retrieved from Table I, and then the shape parameter can be estimated by using (43).

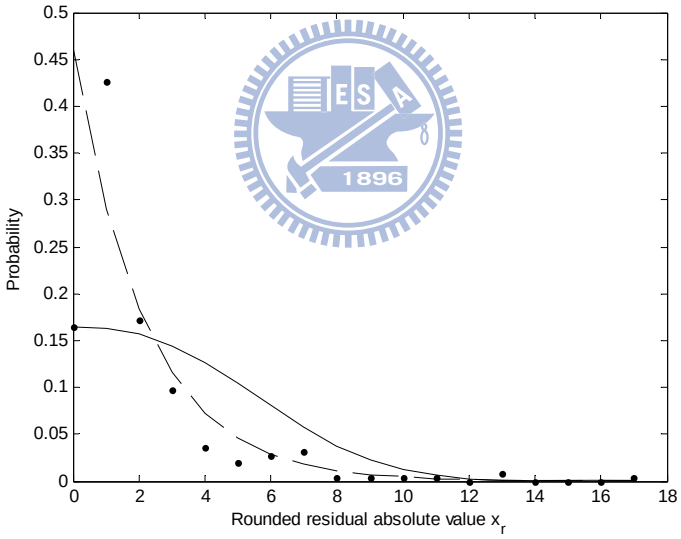
5.3 Improved ρ Estimation

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, 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. 5-2 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. 5-2. In the case of Fig. 5-2 (a), the zero probability, $P\{x_r = 0\}$, is the highest probability, and thus the one-sided ρ-GGD can well approximates the data distribution. However, in the case of Fig. 5-2 (b), because $P\{x_r = 0\}$ is not the peak probability and 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: that is, ρ is the linear combination of two probabilities,



(a)



(b)

Fig. 5-2 The dots are the probability distribution of the residual absolute-valued signal, x_r. The dashed line and 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 x_r=0 (a) and x_r=1 (b), respectively.

$$\rho = a \cdot P\{x_r = 0\} + (1 - a) \cdot P\{x_r = 1\}, \tag{45}$$

and $0 \le a \le 1$. 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 K-L (Kullback-Leibler) divergence as (12). Therefore, for each residual macroblock, we can choose the best *a* value, denoted by *a**,

$$a^* = \arg\min_{a \in A} \left\{ KL\big(P(x) \parallel P_{\rho-GGD}(x;a)\big) \right\}, \tag{46}$$

where *P* is the probability distribution of the residual absolute-valued signal; $P_{\rho-GGD}$ is defined by (38) and its ρ value is estimated using (45). Although (17) can be used in the off-line 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 macroblocks, 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. 5-3. 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,

$$a = \begin{cases} 1 & P\{x_r = 0\} > P\{x_r = 1\} \\ 0 & \text{otherwise} \end{cases}. \tag{47}$$

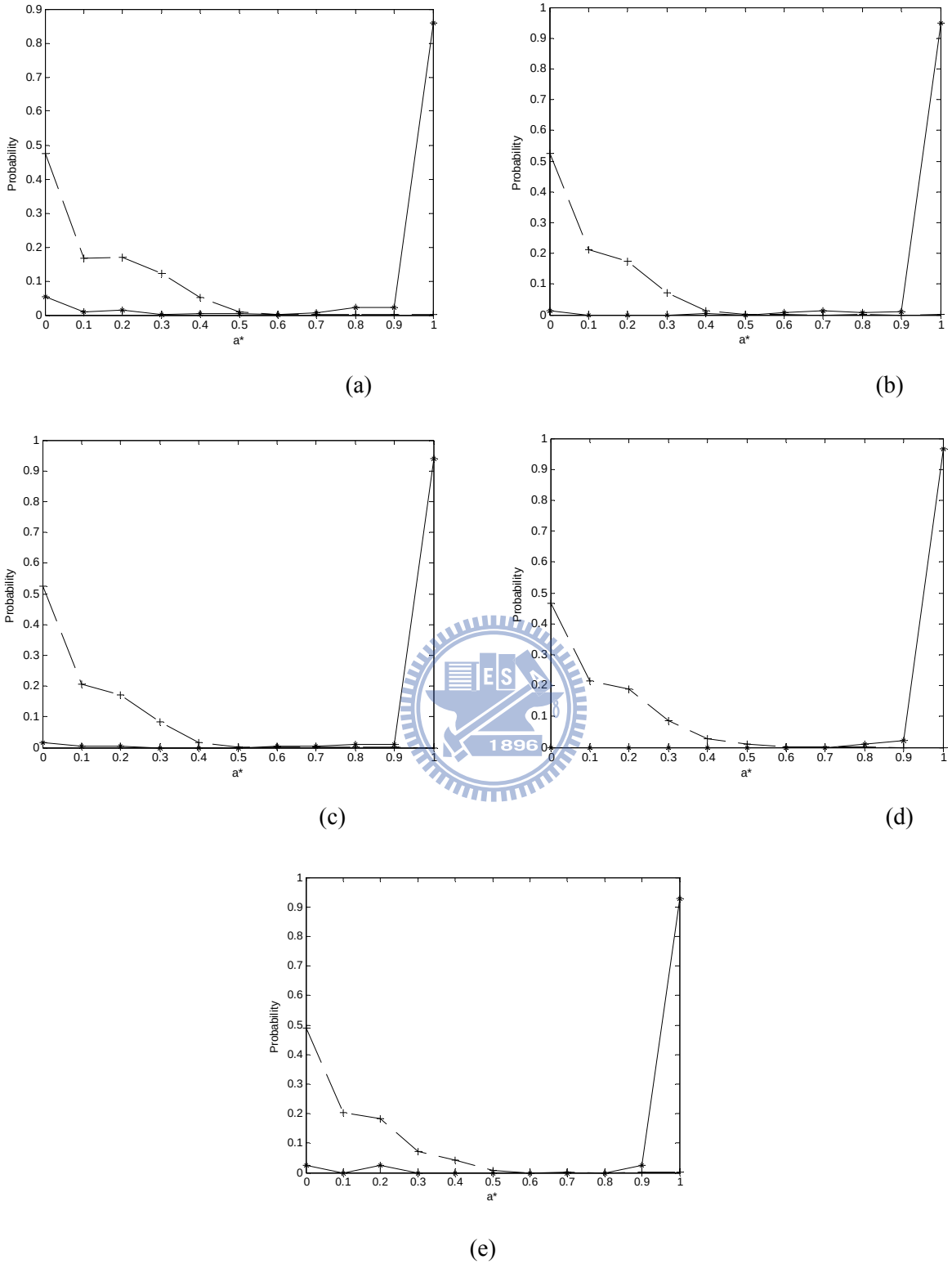


Fig. 5-3. The solid line and dashed line are the probability distributions of the best *a* value, denoted by a*, of the following two cases. The first case is $P\{x_r = 0\} > P\{x_r = 1\}$ (solid) and the second case is the opposite (dashed). The five figures show the results at 5 temporal levels: (a) *t*=0, (b) *t*=1, (c) *t*=2, (d) *t*=3, and (e) *t*=4. The test sequence is Foreman (CIF, 30fps).

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.

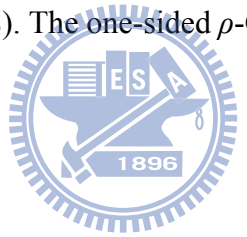
Step 2: Estimate the ρ value using (45) and (47).

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 *i*-th segment coefficients from TABLE I. The shape parameter α_e is estimated by using (43).

Step 6: Insert α_e and ρ into (38). The one-sided *ρ*-GGD modeling is done.



5.4 Experimental Results

In Section 5.1, 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 5.3. 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 macroblock 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 macroblocks in one frame. Fig. 5-4 (a) and Fig. 5-5(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 30fps). From Fig. 5-4 (a) and Fig. 5-5 (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. Fig. 5-4 (b)-(e) and Fig. 5-5 (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.

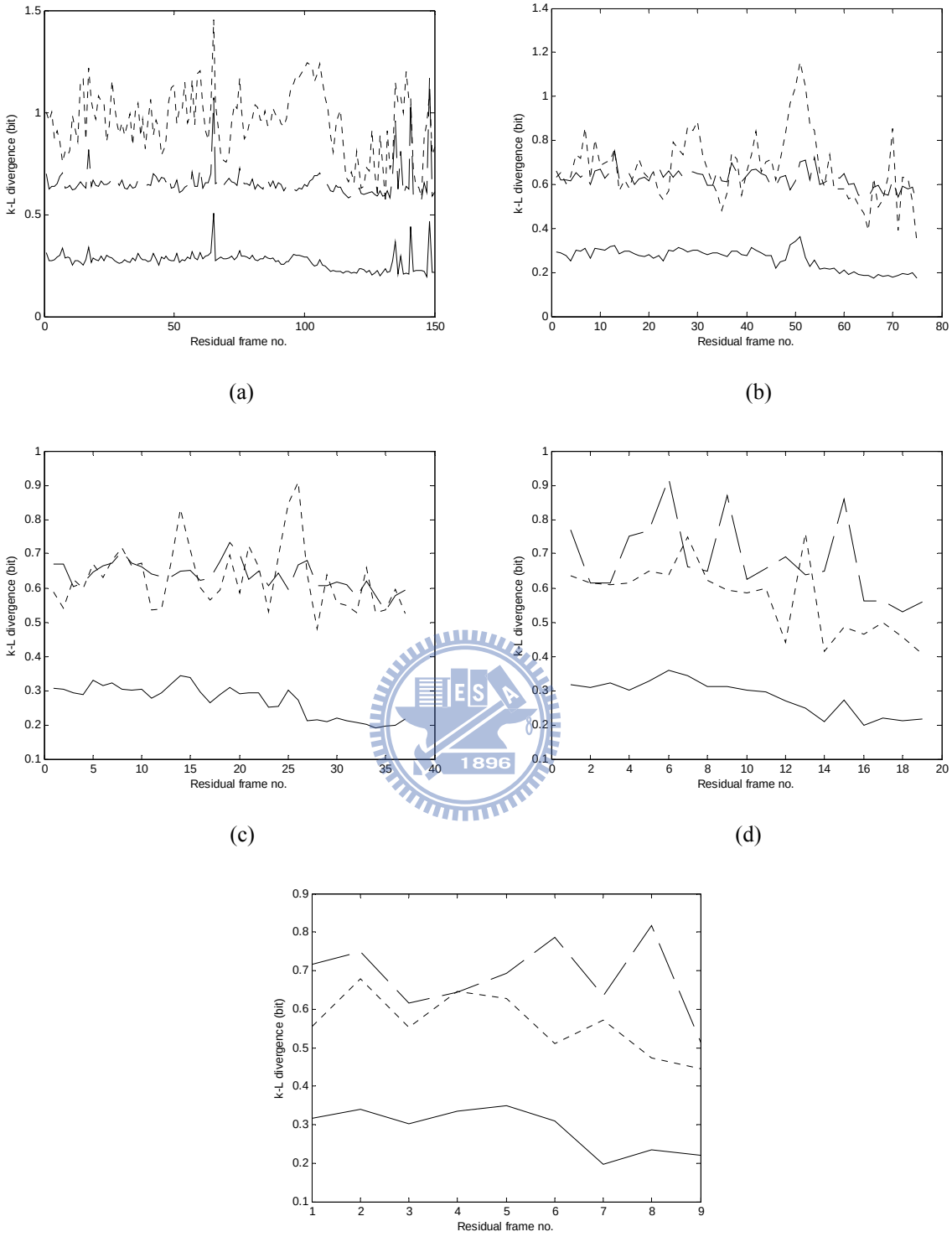


Fig. 5-4. The dotted, dashed and solid lines show the K-L divergence between the probability distributions of the absolute-valued signal and three approximations. These three approximations are Laplacian distribution (dotted), one-sided ρ-GGD (dashed), and one-sided ρ-GGD with the improved ρ estimation (solid), respectively. (a)-(e) figures are the results at different temporal levels (*t*): (a) *t*=0, (b) *t*=1, (c) *t*=2, (d) *t*=3, and (e) *t*=4. The test sequence is Foreman (CIF, 30fps).

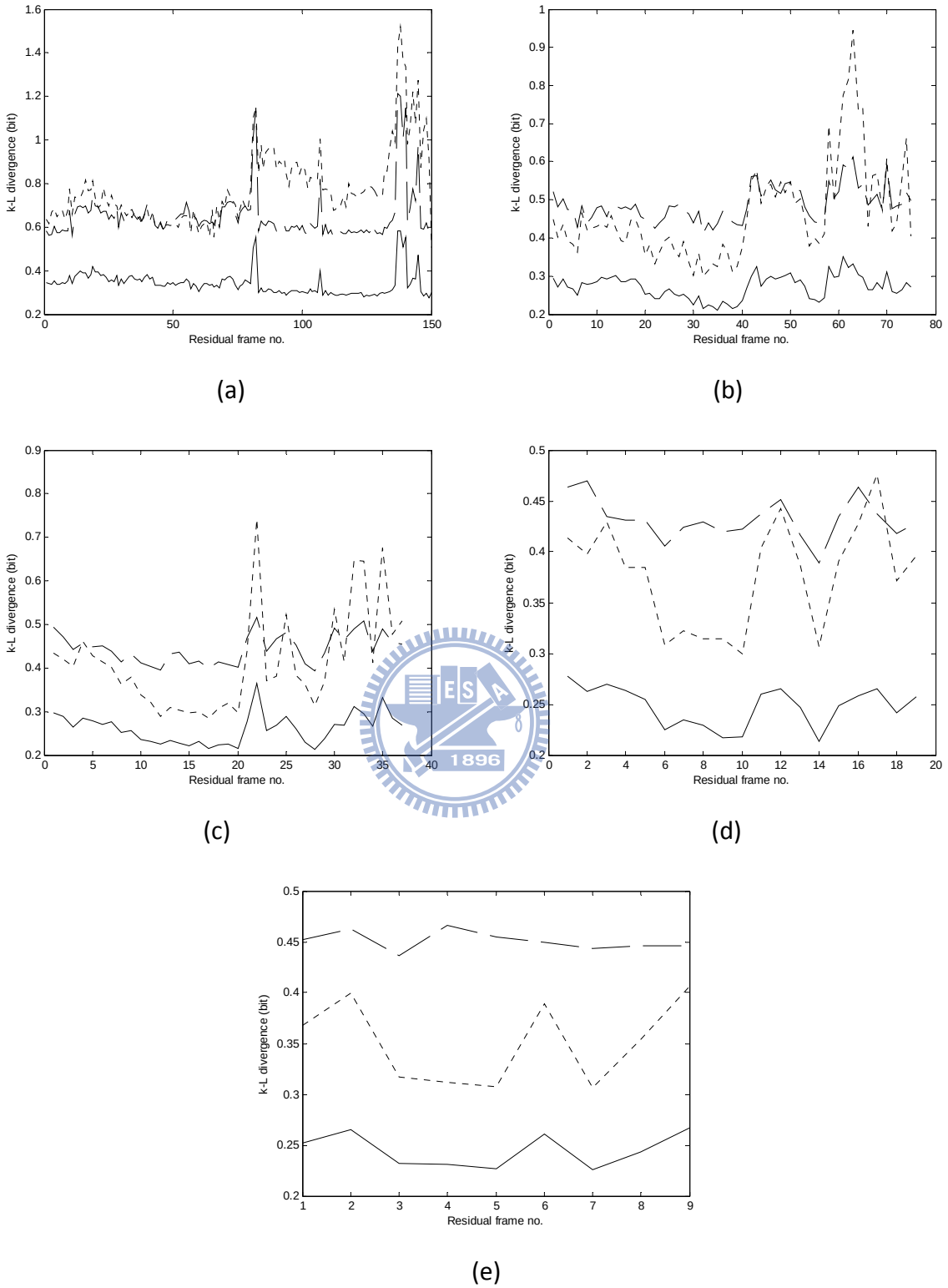


Fig. 5-5. The dotted, dashed and solid lines show the K-L divergence between the probability distributions of the absolute-valued signal and three approximations. These three approximations are Laplacian distribution (dotted), one-sided ρ-GGD (dashed), and one-sided ρ-GGD with the improved ρ estimation (solid), respectively. (a)-(e) figures are the results at different temporal levels (*t*): (a) *t*=0, (b) *t*=1, (c) *t*=2, (d) *t*=3, and (e) *t*=4. The test sequence is Mobile (CIF, 30fps).

Chapter 6 Generalized MIG Derivation and Improved Mode Decision Method

In this chapter, extending our previous work in Chapter 4, we improve the MIG mode decision method by two ways. First, we generalize the MIG derivation by using high-dimensional probability model. Second, we improve the mode decision method by introducing a new temporal weighting factor to the cost function.

6.1 Rate-Distortion Function of ρ-GGD

The source signal is denoted by $x \in X$ with probability distribution function $P_{\rho-GGD}(x)$ defined by (38). According to the Shannon's rate-distortion theory [18], the Shannon lower bound for the magnitude-error criterion is

$$R_L(D) = \Phi(X) - \log(2eD), \tag{48}$$

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 (99) in Appendix, the differential entropy of the one-sided ρ-GGD source model can be written as

$$\begin{aligned}\Phi(X) &= \rho\alpha^{-1}\left(\rho\alpha^{-1}\Gamma(\alpha^{-1})\right)^{-1}\Gamma(\alpha^{-1})(\alpha^{-1} - \log\rho) \\ &= \alpha^{-1} - \log\rho\end{aligned}. \tag{49}$$

where α and ρ are the shape parameter and the zero-value probability of the source model, respectively. Replace $\Phi(X)$ in (48) by (49),

$$\begin{aligned} R_L(D) &= \alpha^{-1} - \log\rho - \log(2eD) \\ &= -\log\left(2\rho e^{(1-\alpha^{-1})} \cdot D\right) \end{aligned} \quad . \tag{50}$$

If the conditions given in [18] are satisfied, $R_L(D)$ becomes $R(D)$, the true rate-distortion function, and can be rewritten as (51)

$$D(R) = \frac{e^{-R}}{2\rho e^{(1-\alpha^{-1})}} \quad . \tag{51}$$

6.2 Generalized MIG Derivation

We now try to find relationship connection between the residual signal statistics and the motion bitrate. As discussed earlier, ρ_{v} and α_{v} denote, respectively, the zero-value probability and the shape parameter in one-sided ρ-GGD model of the residual signal using motion vector v. Thus, ρ_0 and α_0 are the residual signal statistics when v=0. We substitute (51) into (14) with the corresponding parameters, and (14) becomes

$$\frac{e^{-(R-\Delta R)}}{2\rho_{\mathbf{v}} \cdot e^{(1-\alpha_{\mathbf{v}}^{-1})}} < \frac{e^{-R}}{2\rho_0 \cdot e^{(1-\alpha_0^{-1})}} \quad . \tag{52}$$

(52) can be simplified to (53),

$$\frac{\alpha_0^{-1} - \alpha_{\mathbf{v}}^{-1} + \log(\rho_{\mathbf{v}}/\rho_0)}{\Delta R} > 1. \tag{53}$$

Interestingly, the target coding rate term, R_T, in (14) is eliminated. This elimination implies that (53) 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 (53) from a different perspective. Let the residual signal produced by using motion vector v be $x \in X_{\mathrm{v}}$. Similar to the derivation of (49), the differential entropies of X_0 and X_{v} are expressed, respectively, as

$$\begin{aligned} \Phi(X_0) &= \alpha_0^{-1} - \log \rho_0 \\ \Phi(X_{\mathbf{v}}) &= \alpha_{\mathbf{v}}^{-1} - \log \rho_{\mathbf{v}} \end{aligned} \quad . \tag{54}$$

If motion vector 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_v is as follows.

$$\begin{aligned}\Delta\Phi(X_{\mathbf{v}}) &= \Phi(X_0) - \Phi(X_{\mathbf{v}}) \\ &= \alpha_0^{-1} - \alpha_{\mathbf{v}}^{-1} + \log(\rho_{\mathbf{v}}/\rho_0).\end{aligned} \tag{55}$$

We can find that (55) is exactly the numerator of the left term in (53). Thus, (53) is reduced to

$$\frac{\Delta\Phi}{\Delta R} > 1 \ . \tag{56}$$

In (18), a similar conclusion was obtained based on the Laplacian source assumption. However, as discussed in Section 4.2, 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 (56) to

$$\frac{\Delta\Phi}{\Delta R} > C, \tag{57}$$

where C is the MIG lower bound in 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 (14) to reflect the model divergence problem. We rewrite (14)

as

$$D_{real,\mathbf{v}}(R_T - \Delta R) < D_{real,\mathbf{0}}(R_T), \tag{58}$$

where $D_{real,\mathbf{v}}$ is the "real distortion" measured from the quantized residual signal compensated using motion vector v. $D_{ideal,\mathbf{v}}$ is the "ideal distortion" derived from the rate-distortion function of the source model in (14). And a new parameter β_v is introduced to compensate for the difference between $D_{real,\mathbf{v}}$ and $D_{ideal,\mathbf{v}}$. In other words, $\beta_{\mathbf{v}} D_{ideal,\mathbf{v}} = D_{real,\mathbf{v}}$. Or,

$$\beta_{\mathbf{v}} = \frac{D_{real,\mathbf{v}}}{D_{ideal,\mathbf{v}}}. \tag{59}$$

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 β_0 are similarly defined for using the 0 motion vector. Hence, (58) can be rewritten as

$$\beta_{\mathbf{v}} \cdot D_{ideal,\mathbf{v}}(R_T - \Delta R) < \beta_{\mathbf{0}} \cdot D_{ideal,\mathbf{0}}(R_T), \tag{60}$$

By replacing $D_{ideal,\mathbf{v}}$ by the rate-distortion function in (51), (60) gives

$$\frac{\Delta\Phi}{\Delta R} > 1 + \frac{\log_2(\beta_{\mathbf{v}}/\beta_0)}{\Delta R}. \tag{61}$$

(61) is very similar to (56). 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 (A.4) would fall back to (56). Therefore, for the real case, the MIG lower bound *C* becomes

$$C = 1 + \frac{\log_2(\beta_{\mathbf{v}}/\beta_{\mathbf{0}})}{\Delta R}. \tag{62}$$

Let $X'_{\mathbf{v}}$ denotes the quantized residual signal. According to (51), $D_{ideal,\mathbf{v}}$is calculated by

$$D_{ideal,\mathbf{v}} = \frac{2^{-H(X'_{\mathbf{v}})}}{2\rho_{\mathbf{v}} e^{(1-\alpha_{\mathbf{v}}^{-1})}} \ , \tag{63}$$

where $H(X'_{\mathbf{v}})$ is the entropy of the quantized residual signal. Use (59) and (63), (62) can be rewritten as

$$C = 1 + \frac{1}{\Delta R}\left(\alpha_0^{-1} - \alpha_{\mathbf{v}}^{-1} + \log\left(\frac{\rho_{\mathbf{v}}}{\rho_0}\right) - H(X'_0) + H(X'_{\mathbf{v}}) + \log_2\left(\frac{D_{real,\mathbf{v}}}{D_{real,0}}\right)\right). \tag{64}$$

Based on (64), 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 (64).

Table 6-1. The 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

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 macroblocks with non-zero motion vector and calculate the frame-level *C* value using (64). 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 ΔC 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 6-2. The 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

Table 6-1 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 -- in the range of [4, 10] (in Section 4.3). 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 CIF test sequences: *Mobile*, *Container*, *Waterfall*, and *Irene*. The test bitrate points are 256kbps, 384kbps, 512kbps, 800kbps, 1024kbps, 1200kbps, and 1500kbps. The average PSNR results of 7 test points of these two schemes are shown in Table 6-2. As Table 6-2 shows, their PSNR performances are very similar. However, from (64), we can see that the adaptive scheme requires a lot of additional encoding operations. In the experiment section of this chapter, the results are obtained using the offline-trained *C* value, which is 7, and it still outperforms the conventional Lagrangian method.

6.3 Improved MIG Cost Function

We follow the similar process in Section 4.3 to derive MIG cost function for ρ–GGD. Therefore, (27) is rewritten as

$$\begin{aligned} (\alpha_0^{-1} - \log \rho_0) &\geq (\alpha_{\mathbf{v}}^{-1} - \log \rho_{\mathbf{v}}) + C \cdot \Delta R \\ \Rightarrow 2 \cdot \log\left(e^{\alpha_0^{-1}}/\rho_0\right) &\geq 2 \cdot \log\left(e^{\alpha_{\mathbf{v}}^{-1} + C \cdot \Delta R}/\rho_{\mathbf{v}}\right) \\ \Rightarrow \frac{e^{2/\alpha_0}}{{\rho_0}^2} &\geq \frac{e^{2/\alpha_{\mathbf{v}}}}{{\rho_{\mathbf{v}}}^2} \cdot e^{2 \cdot C \cdot \Delta R} \end{aligned} . \tag{65}$$

When an MV produces a smaller right-side term in (65), it leads to a larger φ. Hence, we look for the best MV that achieves the minimum right term value in (65). Also, when ΔR equals to zero, the right term reaches its maximum value ${e^{2/\alpha_{\mathbf{v}}}}/{\rho_{\mathbf{v}}}^2$ and there is no singular problem. Therefore, for source signal s and motion vector v, the proposed MIG cost function is defined as

$$J(\mathbf{s}, \mathbf{v}|C) = \frac{e^{2/\alpha_{\mathbf{s}}}}{{\rho_{\mathbf{s}}}^2} \cdot e^{2 \cdot C \cdot \Delta R(\mathbf{v})}, \tag{66}$$

where α_s and ρ_s are the shape parameter and zero-value probability of the source signal s and ΔR(v) is the MV bit rate. On the other hand, from (39) and (40), we have

$${\rho_{\mathbf{s}}}^2 {\sigma_{\mathbf{s}}}^2 = \Omega(\alpha_{\mathbf{s}}), \tag{67}$$

where $\sigma_{\mathbf{s}}^2$ is the residual signal variance. Hence, (66) can be rewritten as

$$J(\mathbf{s}, \mathbf{v}|C) = \frac{e^{2/\alpha_{\mathbf{s}}}}{\Omega(\alpha_{\mathbf{s}})} \cdot {\sigma_{\mathbf{s}}}^2 \cdot e^{2 \cdot C \cdot \Delta R(\mathbf{v})} . \tag{68}$$

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

$$\tau(\alpha) = \frac{e^{2/\alpha}}{\Omega(\alpha)} ; \tag{69}$$

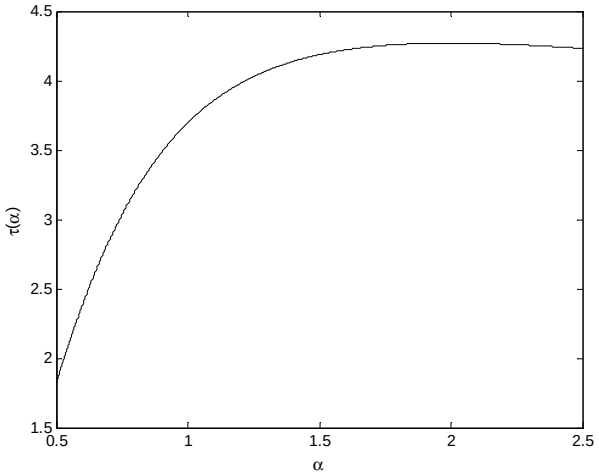


Fig. 6-1. The cost weighting function $\tau(\alpha)$ for $\alpha \in [0.5, 2.5]$.

and thus,

$$J(\mathbf{s}, \mathbf{v}|C) = \tau(\alpha_{\mathbf{s}}) \cdot {\sigma_{\mathbf{s}}}^2 \cdot e^{2 \cdot C \cdot \Delta R(\mathbf{v})}. \tag{70}$$

The function values of $\tau(\alpha)$ are shown in Fig. 6-1. It increases as α increases but saturates at about α=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", (70) has another equivalent form as follows:

$$J(\mathbf{s}, \mathbf{v}|C) = \tau(\alpha_{\mathbf{s}}) \cdot {\sigma_{\mathbf{s}}}^2 \cdot 2^{2 \cdot C \cdot \Delta R(\mathbf{v})}. \tag{71}$$

In the case of Laplacian source model in Section 4.3, (71) is reduced to

$$J_{\text{Laplacian}}(\mathbf{s}, \mathbf{v}|C) = {\sigma_{\mathbf{s}}}^2 \cdot 2^{2 \cdot C \cdot \Delta R(\mathbf{v})}. \tag{72}$$

The difference between (71) and (72) is $\tau(\alpha_s)$. 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. 6-1 shows, when α is small, so is $\tau(\alpha)$. Thus, the proposed MIG cost function in the form of (71) provides a richer interpretation, which links to the pdf shape.

6.4 Temporal Weighting for MIG Lower Bound

After motion-compensated prediction, the relationship between the pixels on the predicted and the reference frames can be classified to three types: *connected*, *unconnected*, and *multi-connected* [8]. During the MCTF process, because the temporal correlation between the low-pass frames at the deep temporal level is relatively small, the unconnected pixel percentage increases, which implies that the prediction effectiveness decreases. Furthermore, the connection relationship leads to the distortion propagation along the tree structure generated by the temporal filtering process after quantization, which is the so-called "quantization noise propagation" problem in MCTF [13],[33]. Here we follow the notations defined by [13] in modeling the noise propagation process. The average distortions of the low-pass frame and the high-pass frame at temporal level t are denoted as $\overline{d}_L^{(t)}$ and $\overline{d}_H^{(t)}$, respectively. When the Harr wavelet filter is adopted in MCTF, Wang and Schaar [13] show that $\overline{d}_L^{(t)}$ and $\overline{d}_H^{(t)}$ are related to $\overline{d}_L^{(t-1)}$ by the following equation,

$$\overline{d}_L^{(t-1)} = \frac{1}{2}\overline{d}_L^{(t)} + \left(\frac{3}{4} - \frac{r_c}{4}\right)\cdot\overline{d}_H^{(t)}, \tag{73}$$

where r_c is the ratio of the connected pixels. It is obvious that r_c determines the severity of the distortion propagation problem. There are two major factors affecting the r_c value: the picture characteristics and the motion estimation method. By minimizing the MIG cost function with the pre-chosen $C^{(t)}$ parameter (the C value at the t temporal level), the best motion vector set $\mathrm{v}^{(t)}$ can be obtained, and thus r_c is decided. The frames are temporally

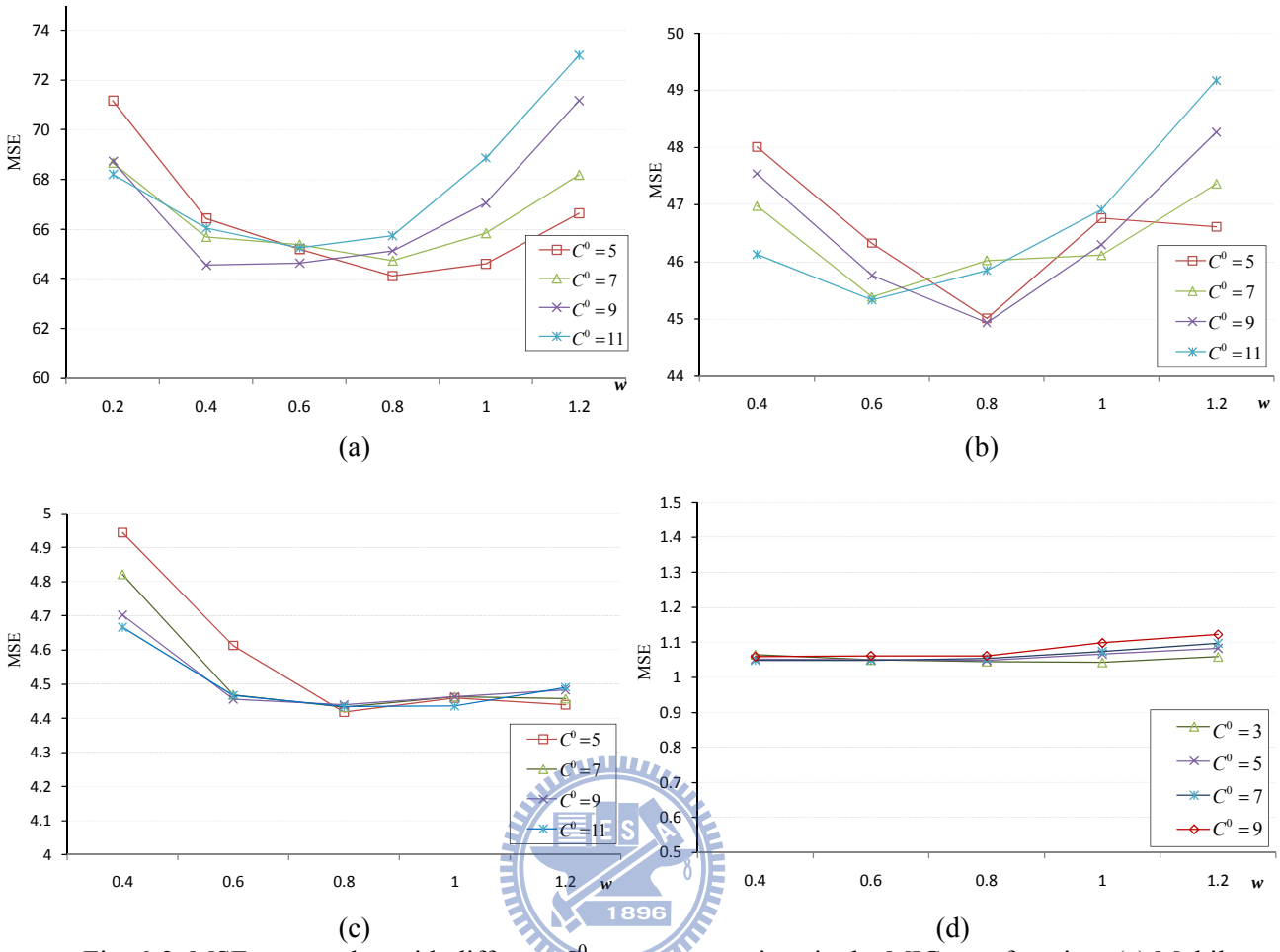


Fig. 6-2. MSE vs. *w* value with different C^0 parameter settings in the MIG cost function: (a) Mobile (b) Tempete, (c) Container, and (d) Akiyo, all in CIF resolution.

decomposed along the $v^{(t)}$ trajectory. Hence, $\bar{d}_L^{(t)}$ and $\bar{d}_H^{(t)}$ are the functions of $v^{(t)}$. We rewrite (73) as

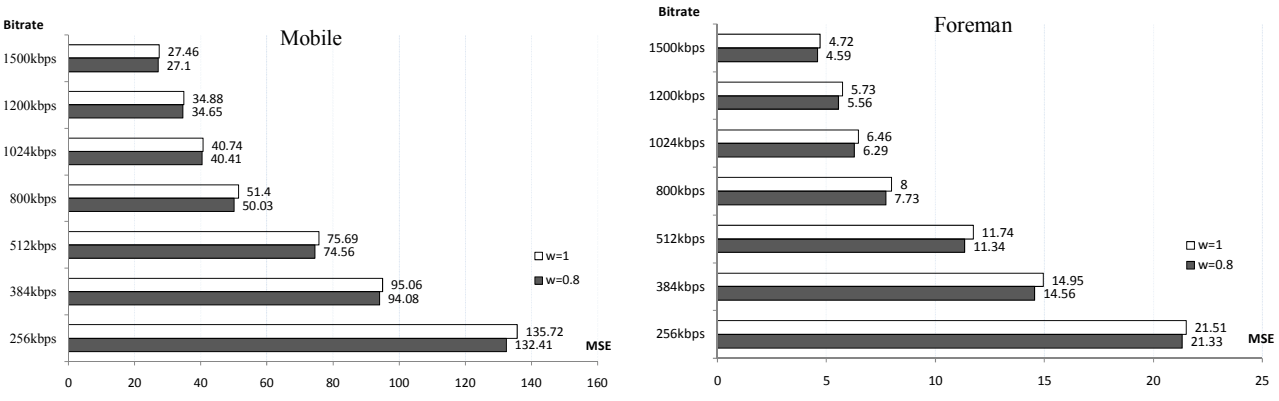
$$\bar{d}_L^{(t-1)} = \frac{1}{2}\bar{d}_L^{(t)}(\mathbf{v}^{(t)}|C^{(t)}) + \left(\frac{3}{4} - \frac{r_c(\mathbf{v}^{(t)}|C^{(t)})}{4}\right)\cdot\bar{d}_H^{(t)}(\mathbf{v}^{(t)}|C^{(t)}), \tag{74}$$

in which the notation $(.|C^{(t)})$ is inserted to emphasize the result depends on the $C^{(t)}$ value. Thus, in the Haar wavelet filter case, (74) shows that the rate-distortion behavior of the low-pass frame at temporal level *t*-1 is affected by the motion vectors at temporal level *t*.

Theoretically, to find the optimal solution of mv, the effects of the quantized/truncated residual signals at all the previous temporal levels have to be considered. Practically, because of the open-loop structure and the complexity of the inter-scale coding system, it is very difficult to construct an analytical model, or even an experimental model, to describe the relationship between the distortion propagation and the motion information. A feasible approach is to adjust the C value of (28) along with the increased temporal level. Also, this adjustment changes the values of $\sigma_{\mathbf{s}}^2(\mathbf{v})$ and $\Delta R(\mathbf{v})$ according to their located MCTF decomposition layer and thus it can effectively compensate for the propagation distortion loss. Therefore, the MIG cost function of (71) is modified to

$$J(\mathbf{s},\mathbf{v}|C^{(t)}) = \tau(\alpha_{\mathbf{s}}) \cdot \sigma_{\mathbf{s}}^2(\mathbf{v}) \cdot 2^{2\cdot C^{(t)}\cdot \Delta R(\mathbf{v})}, \tag{75}$$

where the superscript t is the temporal level index in MCTF. It is shown that the statistical relationship between consecutive subband signals can be modeled by a hidden Markov



(a) (b)

Fig. 6-3. The MSE comparison between the cases with temporal weighting, w=0.8 and w=1, in the MIG cost function at different truncation bitrates. Test sequences are (a) Mobile and (b) Foreman. (CIF resolution).

model [34]. Similarly, a Markov-like relationship seems to exist between consecutive temporal decomposition layers. Thus, the optimally decided distortion values of these layers are correlated. Therefore, we conjecture that a simple linear predictor can describe the relationship of the C parameters among temporal layers. That is, for two consecutive temporal levels,

$$C^{(t)} = w \cdot C^{(t-1)}. \tag{76}$$

Consequently, if C^0 is given for the first temporal level, (76) becomes

$$C^{(t)} = w^t \cdot C^0. \tag{77}$$

In practice, the weighting factor w can be found by extensive experiments. We start with a pair of C^0 and w values and use (75) to perform motion search and mode decision. Repeating the same experimental steps for Fig. 4-2(d) with different w values, we obtain the MSE vs. w curves using different C^0 values. The experimental results are shown in Fig. 6-2. Because the motion information percentage in fast-motion pictures is larger than that in the slow-motion pictures, the error propagation problem is severe. Hence, the benefit of using our temporal weighting adjustment is more significant in the fast-motion cases. Fig. 6-2 (a) and Fig. 6-2 (d) show the results of Mobile and Akiyo test sequences respectively. Compared with Akiyo, Mobile is a relatively fast-motion test sequence, and thus the distortion in Fig. 6-2 (a) is more sensitive to the w value than that in Fig. 6-2 (d). In contrast, the temporal weight adjustment makes little difference in MSE for the Akiyo test sequence.

Fig. 6-2 shows that the average MSE is a convex function in w and the minimal MSE appears at around [0.6, 0.9]. According to the collected data, w =0.8 seems to be a good value for most cases. To verify the effectiveness of our chosen temporal weighting factor, we tested Mobile and Foreman videos and adopted the MIG cost function with weightings, w=0.8 and w=1. In these simulations, the C^0 parameter is set to 7. Fig. 6-3 shows that applying the temporal weighing factor can improve the overall MSE at different bitrates.

In addition to the empirical selection method, we have also derived the w value from the viewpoint of decoder rate-distortion behavior. Because the decoding bit rate 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 [37], $C^{(t)}$ in (75), is highly related to the so-called synthesis gain. Let g_L 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 g_L times the mean-squared distortion of the temporal low-pass frame. Following the spirit in [13], 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, (24) is modified to

$$\frac{\Delta\Phi}{\Delta R}\cdot(\sqrt{g_L})^t > C^0, \tag{78}$$

where C^0 is the target MIG lower bound at the first temporal level (t=0). Or, (78) can be rewritten to an equivalent form:

$$\frac{\Delta\Phi}{\Delta R} > C^t, \tag{79}$$

where

$$C^t = \left(\frac{1}{\sqrt{g_l}}\right)^t \cdot C^0 = w^t C^0. \tag{80}$$

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

$$g_L = (0.5)^2 + (1)^2 + (0.5)^2) = 1.5. \tag{81}$$

Thus, $w = 1/\sqrt{1.5} = 0.817$. This theoretically derived w value is consistent with the finding in our previous work: w value generally falls in the range of [0.6, 0.9]. In the experiment section of this chapter, the results are obtained using the offline-trained w value in [17], which is 0.8. In summary, (75) is now the cost function used for both motion estimation and mode decision. Their detailed steps are described in the next section.

6.5 Improved 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 [25], as an example. The basic prediction unit is macroblock (MB). Its motion compensation mode consists of a MB partition. The sub-block size can be 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, and 4x4 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

$$\begin{aligned} \mathbf{b}_m &= (b_1, \cdots, b_{N_m}) \\ \mathbf{v}_m &= (v_1, \cdots, v_{N_m}) \end{aligned} \tag{82}$$

where b_i and v_i represents the i-th sub-block residual signal and its MV, respectively. Assume **M** is the mode candidate set, that is, $m \in \mathbf{M}$. As **Fig. 6-4** shows, there are six steps in deciding the best prediction mode.

1) *Step 1:* Select the MIG cost function parameters

The proposed MIG cost function (75) contains one parameter, C_t. According to (80), 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.

2) *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 i-th 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 (45) and (47) and is denoted as ρ_{b_i}. According to (43), the shape parameter of the sub-block b_i can be obtained by

$$\alpha_{b_i} = {\Omega_e}^{-1}\left({\rho_{b_i}}^2(v) \cdot {\sigma_{b_i}}^2(v)\right) . \tag{83}$$

Therefore, the MIG cost for motion vector v is

$$J_{\mathrm{mv}}(b_i, v|C_t) = \tau(\alpha_{b_i}) \cdot {\sigma_{b_i}}^2(v) \cdot 2^{2 \cdot C_t \cdot \Delta R(v)} , \tag{84}$$

where $\Delta R(v)$ 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

$$v_i^* = \arg\min_{v \in \mathbf{S}} \{ J_{\mathrm{mv}}(b_i, v|C_t) \}. \tag{85}$$

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

$$\mathbf{v}_m^* = (v_1^*, \cdots, v_{N_m}^*). \tag{86}$$

3) *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 ${\sigma_{\mathbf{b}_m}}^2$ denote, respectively, the zero-value probability and the variance of $\mathbf{b}_m$. ${\sigma_{\mathbf{b}_m}}^2$ are computed by

$$ {\sigma_{\mathbf{b}_m}}^2(\mathbf{v}_m^*) = \frac{1}{N_m} \sum_{i=1}^{N_m} {\sigma_{b_i}}^2(v_i^*) \ , \tag{87} $$

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 (45) and (47). Next, the motion bitrate for this MB is given by

$$ \Delta R(\mathbf{v}_m^*) = \frac{1}{N_m} \sum_{i=1}^{N_m} \Delta R(v_i^*) + r_m \ , \tag{88} $$

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

4) *Step 4:* Estimate the shape parameter from MB residuals

According to (43), the shape parameter of $\mathbf{b}_m$ is estimated by

$$ \alpha_{\mathbf{b}_m} = {\Omega_e}^{-1}\left({\rho_{\mathbf{b}_m}}^2(\mathbf{v}_m^*) \cdot {\sigma_{\mathbf{b}_m}}^2(\mathbf{v}_m^*) \right). \tag{89} $$

5) Step 5: Calculate the MIG cost for mode *m*

Using the parameter values calculated in *Steps 1* to *5*, we can compute the MIG cost for mode *m*:

$$ J_{\text{mode}}(\mathbf{b}_m, \mathbf{v}_m^* | C_t) = \tau(\alpha_{\mathbf{b}_m}) \cdot {\sigma_{\mathbf{b}_m}}^2(\mathbf{v}_m^*) \cdot 2^{2 \cdot C_t \cdot \Delta R(\mathbf{v}_m^*)}. \tag{90} $$

If mode *m* is the last mode in **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.

6) *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

$$m^* = \arg \min_{m \in \mathbf{M}} \{ J_{\text{mode}}(\mathbf{b}_m, \mathbf{v}_m^* | C_t) \}. \tag{91}$$

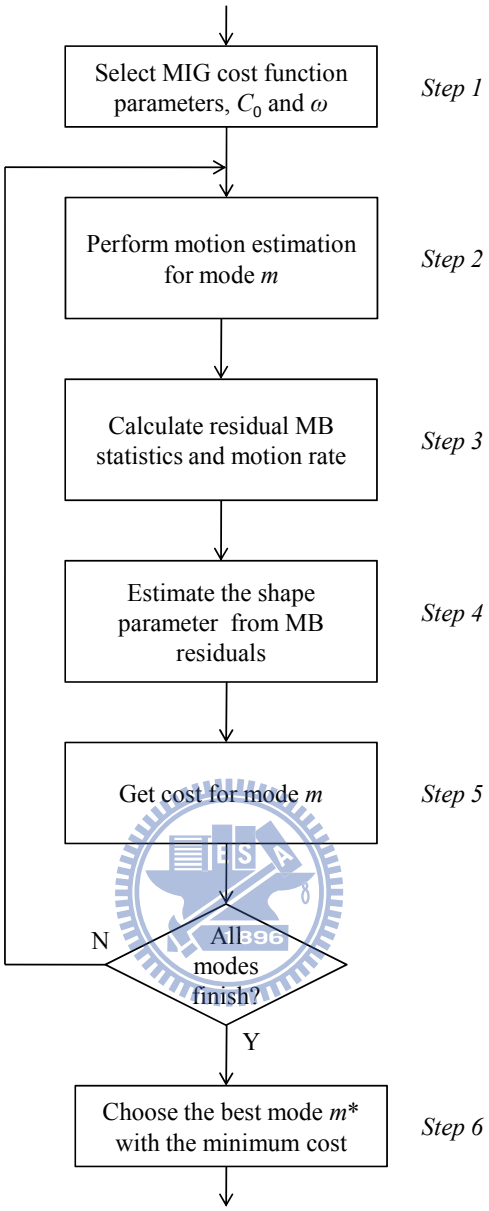


Fig. 6-4. Flow chart of the proposed mode decision procedure

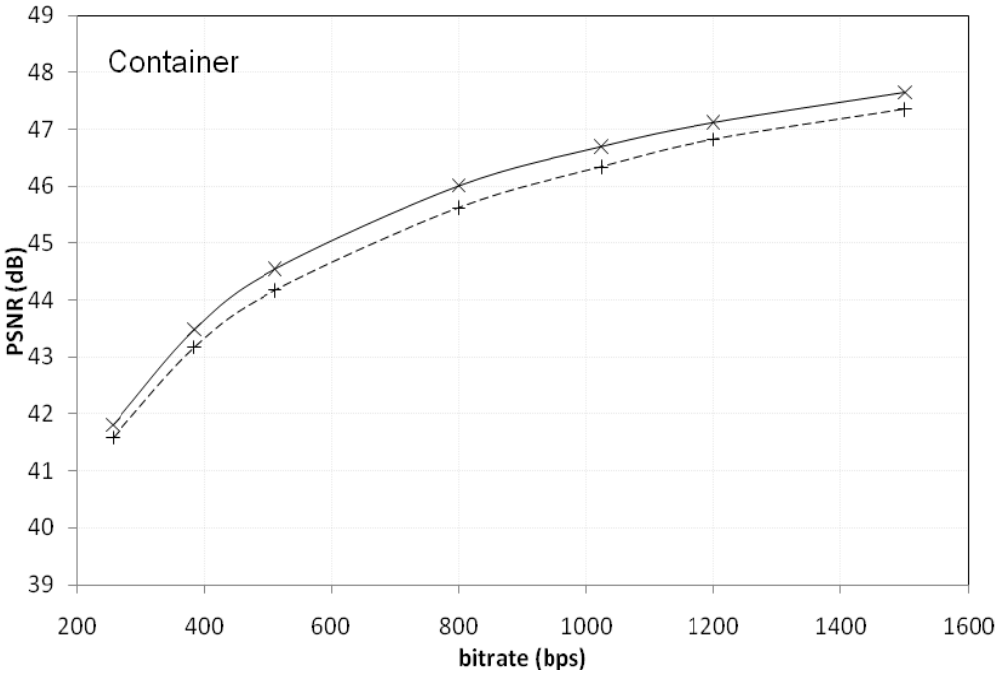
6.6 Experimental Results

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 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 6-3. Our proposed method use the same motion search range and motion vector accuracy settings in Table 6-3. 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.

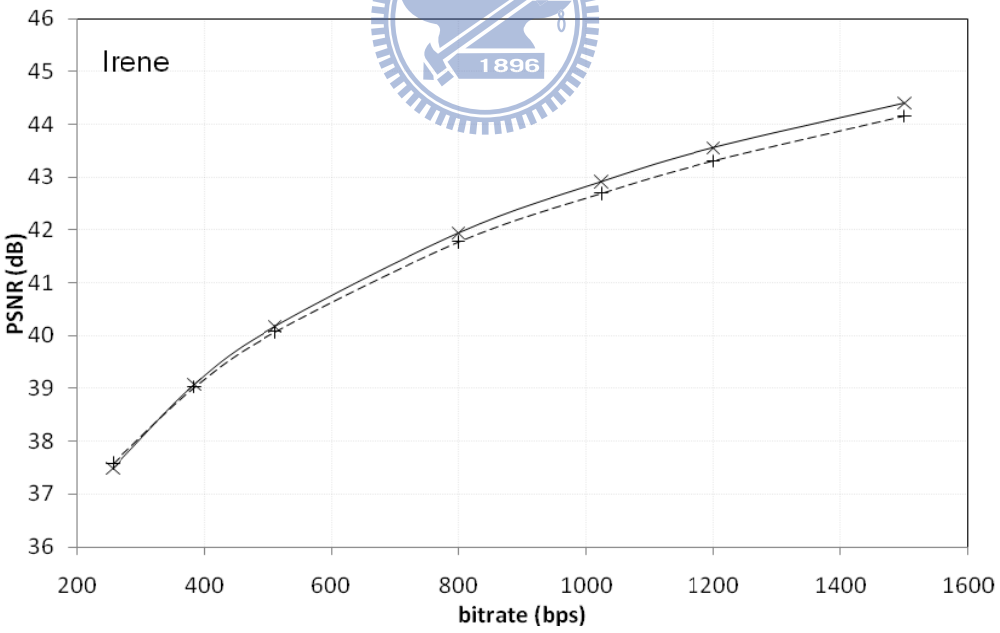
The first scenario is the SNR scalability test. We select 6 test sequences: Container, Irene, Foreman, Tempete, Waterfall, Mobile. All are in the CIF format and 30 fps. In this scenario, C_0 and ω of the MIG cost function are 7 and 0.8, respectively. The operation bitrates are: 256kbps, 384kbps, 512kbps, 800kbps, 1024kbps, 1.2Mbps, and 1.5Mbps. For each test sequence, 7 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. 6-5 shows the PSNR comparison between the two coding methods for the 6 test sequences. Compared with the conventional Lagrangian method, our method shows 0.1 to 0.9 dB PSNR improvements.

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

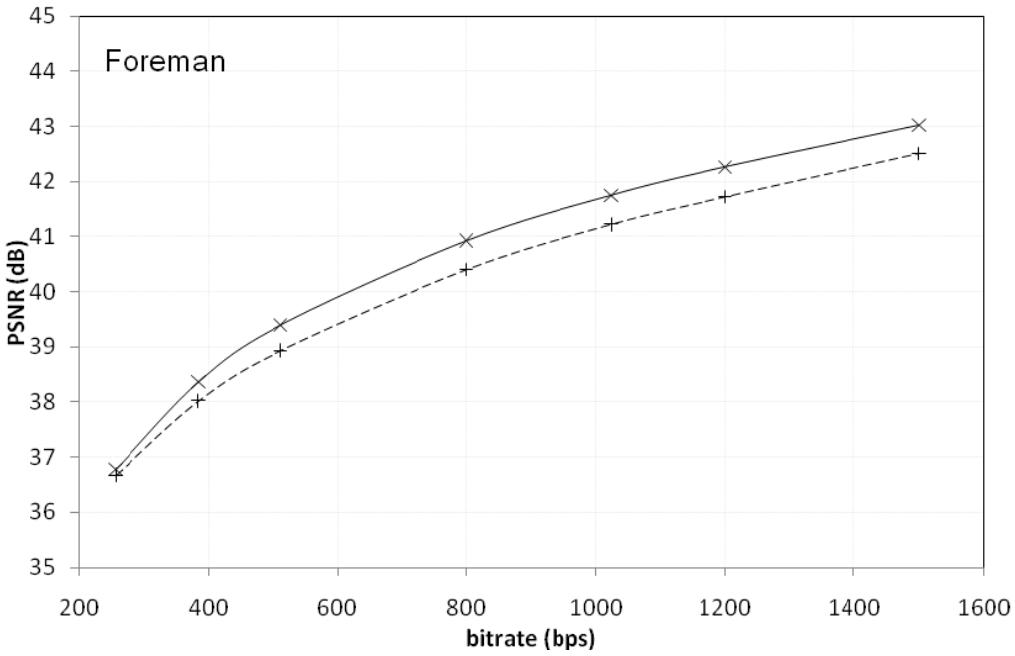
addition to the CIF videos in the first scenario, we also test 5 high-resolution test sequences: City, Crew, Harbour, Soccer, and Ice. All are in the 4CIF format and 60 fps. The operation points include 6 bitrates combined with 3 frame rates. The C_0 value is empirically selected within [7, 10] and ω is 0.8. Table 6-4 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.


Container
PSNR (dB)
bitrate (bps)
49
48
47
46
45
44
43
42
41
40
39
200
400
600
800
1000
1200
1400
1600

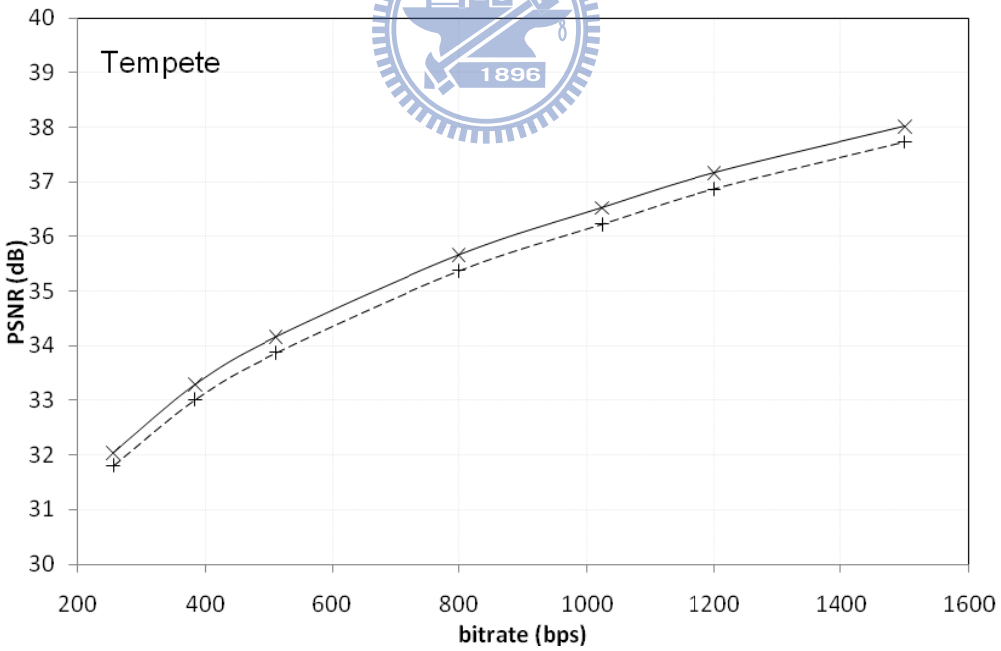
(a)


Irene
PSNR (dB)
bitrate (bps)
46
45
44
43
42
41
40
39
38
37
36
200
400
600
800
1000
1200
1400
1600

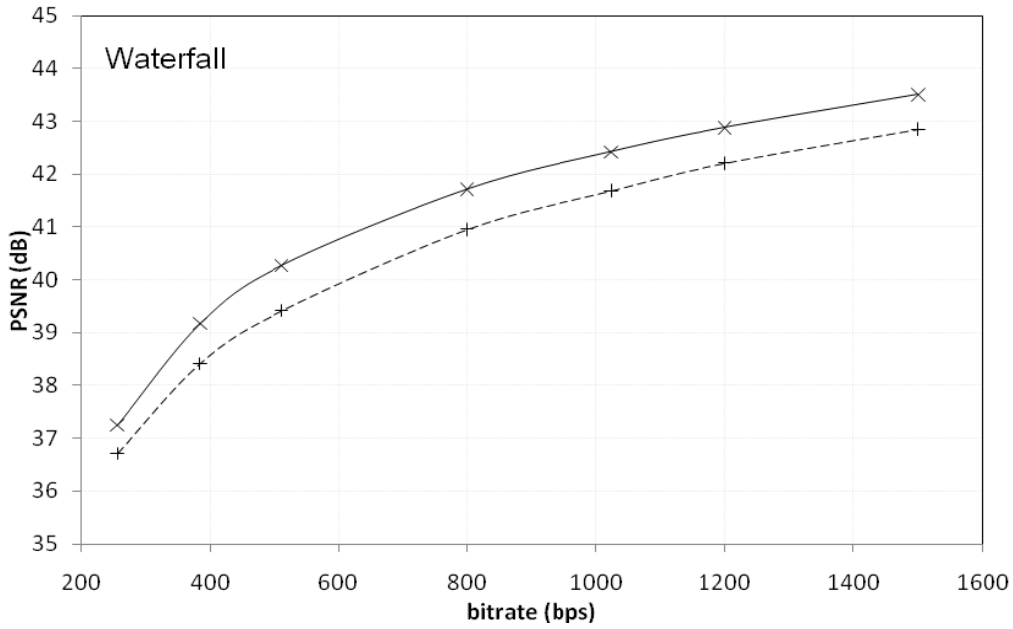
(b)


Foreman
PSNR (dB)
45
44
43
42
41
40
39
38
37
36
35
200
400
600
800
1000
1200
1400
1600
bitrate (bps)

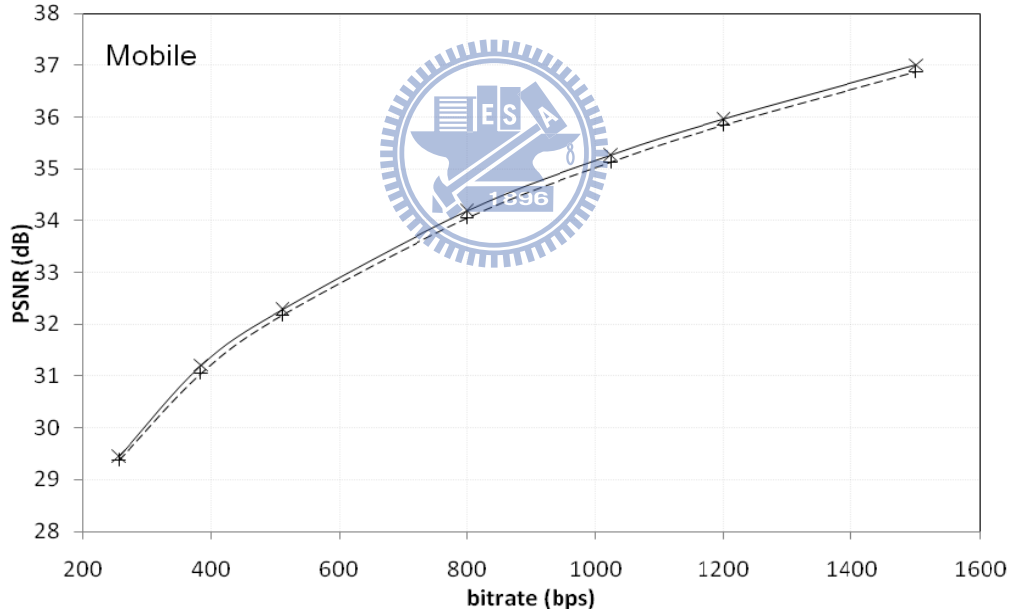
(c)


Tempete
PSNR (dB)
40
39
38
37
36
35
34
33
32
31
30
200
400
600
800
1000
1200
1400
1600
bitrate (bps)

(d)



(e)



(f)

Fig. 6-5. The 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, (f) Mobile. (CIF resolution, 30fps)

Table 6-3

The default parameter settings [36] of MCTF in Vidwav coder.

	Motion search range (pel)	Motion vector accuracy (pel)		Lagrange parameter	
		CIF	4CIF	CIF	4CIF
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

Table 6-4

The PSNR Comparison between the Proposed MIG cost method and the Conventional Lagrangian Method in Combined Temporal and SNR Scalability Test for 5 Test Sequences (4CIF Resolution, 60fps)

Sequence (4CIF)	GOP size	Mode decision method	750Kbps 15fps	1024Kbps 15fps	1200Kbps 30fps	1500Kbps 30fps	2048Kbps 60fps	3000Kbps 60fps
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
Harbour	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

Conclusions

The interframe wavelet video coding scheme provides a flexible and efficient structure for producing scalable bit streams. However, because of its open-loop structure, its parameter optimization issue becomes a challenging problem. To analytically solve this R-D optimization problem, we construct the wavelet texture model and derive a motion information index.

The ρ-GGD source model is proposed to approximate the probability distribution of the wavelet coefficients and the residual signals in the scalable wavelet video codec. We suggest a fast scheme that constructs the ρ-GGD based on the zero-value probability (ρ) and the source signal variance. Also, we propose 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 for the one-sided ρ-GGD.

We derive the rate-distortion function for the wavelet video coder based on the one-sided ρ-GGD model. The notion of "motion information gain" (MIG) is defined and a mode decision procedure is developed based on this MIG metric. This mode decision procedure is nearly bitrate independent in theory and thus is suitable for solving the multi-operation-point (multiple rates) problem in scalable wavelet video coding. Our simulation results show that the one-sided ρ-GGD based mode decision algorithm provides

an improvement of 0.1 to 0.5 dB in PSNR over the conventional Lagrangian method on both the SNR scalability and the combined SNR and temporal scalability tests.

附錄(Appendix): Differential Entropy of the High-Order Exponential PDF

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

$$p(x) = \gamma \exp(-\beta x^{\alpha}), \quad x \geq 0, \tag{92}$$

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

$$\Phi(X) = -\int_X p(x) \cdot \log\big(p(x)\big) dx, \tag{93}$$

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

$$\begin{aligned} \Phi(X) &= -\int_0^{\infty} \gamma \exp(-\beta x^{\alpha}) \cdot \log\big(\gamma \exp(-\beta x^{\alpha})\big) dx \\ &= \gamma\Big(\beta \int_0^{\infty} x^{\alpha} \exp(-\beta x^{\alpha}) dx - \log \gamma \cdot \int_0^{\infty} \exp(-\beta x^{\alpha}) dx\Big) \end{aligned}. \tag{94}$$

Here we rewrite $\Phi(X)$ as

$$\Phi(X) = \gamma\Big(\beta \cdot A - \log \gamma \cdot B\Big), \tag{95}$$

where

$$\begin{aligned} A &= \int_0^{\infty} x^{\alpha} \cdot \exp(-\beta x^{\alpha}) dx \\ B &= \int_0^{\infty} \exp(-\beta x^{\alpha}) dx \end{aligned}. \tag{96}$$

Let us derive A and B first, and then substitute the results into $\Phi(X)$ in (95). We use a new variable $t = -\beta x^{\alpha}$ to replace the variable x in A. Thus, A is derived as

$$
\begin{aligned}
A &= \int_0^\infty \beta^{-1} t \exp(-t) \alpha^{-1} \beta^{-1/\alpha} t^{1/\alpha - 1} dt \\
&= \alpha^{-1} \beta^{-(1/\alpha+1)} \int_0^\infty \exp(-t) t^{(1/\alpha+1)-1} dt \\
&= \alpha^{-1} \beta^{-(1/\alpha+1)} \cdot \Gamma(\alpha^{-1} + 1)
\end{aligned}
\quad , \tag{97}
$$

where $\Gamma(\cdot)$ is the standard Gamma function. With the similar procedure, B in (96) is derived as

$$
\begin{aligned}
B &= \alpha^{-1} \beta^{-1/\alpha} \int_0^\infty \exp(-t) \cdot t^{1/\alpha - 1} dt \\
&= \alpha^{-1} \beta^{-1/\alpha} \cdot \Gamma(\alpha^{-1})
\end{aligned}
\quad . \tag{98}
$$

By using (97) and (98), $\Phi(X)$ can be rewritten as

$$
\begin{aligned}
\Phi(X) &= \gamma \left(\beta \alpha^{-1} \beta^{-(1/\alpha+1)} \Gamma(\alpha^{-1} + 1) - \log \gamma \cdot \alpha^{-1} \beta^{-1/\alpha} \Gamma(\alpha^{-1}) \right) \\
&= \gamma \alpha^{-1} \beta^{-1/\alpha} \left(\Gamma(\alpha^{-1} + 1) - \log \gamma \cdot \Gamma(\alpha^{-1}) \right) \\
&= \gamma \alpha^{-1} \beta^{-1/\alpha} \Gamma(\alpha^{-1}) \left(\alpha^{-1} - \log \gamma \cdot \right) \quad \text{(nat)}
\end{aligned}
\quad . \tag{99}
$$

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

參考文獻(References)

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

[2] R. Leonardi, T. Oelbaum, and J.-R. Ohm, ”Status report on wavelet video coding exploration,” ISO/IEC JTC1/SC29/WG11 MPEG, N8043, April 2006.

[3] J. M. Shapiro, “An embedded wavelet hierarchical image coder,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 657-660, San Francisco, CA, Mar. 1992.

[4] 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, pp. 243-250, June 1996.

[5] D. Taubman, “High performance scalable image compression with EBCOT,” *IEEE Trans. Image Processing*, vol. 9 no. 7, pp. 1158-1170, July. 2000.

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

[7] J.-R. Ohm, “Three-dimensional Subband Coding with Motion Compensation,” *IEEE Trans Image Processing.*, vol. 3, no. 5, pp. 559-571, Sep. 1994.

[8] S.-T. Hsiang and J. W. Woods, “Embedded Video Coding Using Invertible Motion Compensated 3-D Subband Wavelet Filter Bank,” *Signal Processing: Image Communications*, vol. 16, pp.705-724, May 2001.

[9] A. Secker and D. Taubman, “Lifting-Based Invertible Motion Adaptive Transform (LIMAT) Framework for Highly Scalable Video Compression,” *IEEE Trans Image Processing.*, vol. 12, no. 12, Dec. 2003.

[10] J. Xu, R. Xiong, B. Feng, G. Sullivan, M. C. Lee, F. Wu and S. Li, "3D Subband Video Coding Using Barbell Lifting," *ISO/IEC JTC1/SC29/WG11 MPEG*, M10569, 2004.

[11] J. Xu, Z. Xiong, S. Li, and Y.-Q. Zhang, "3-D embedded subband coding with optimal truncation (3-D ESCOT)," *J. Applied Computational Harmonic Analysis*, vol. 10, pp.290-315, May 2001.

[12] W.-H Peng, C.-Y. Tsai, T. Chiang and H.-M. Hang, *Knowledge-Based Intelligent Information and Engineering Systems* Berlin, Germany: Springer, ch. 3, Advances of MPEG scalable video coding standards, 2005, vol. 3684, pp.889-895.

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

[14] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences," *IEEE J. Sel. Areas Commun.*, vol. SAC-5, pp. 1140-1154, Aug. 1987.

[15] B. Girod, "Rate-constrained motion estimation," in *Proc. Int. Symp. Visual Commun. Image Processing*, Nov. 1994, pp. 1026-1034.

[16] M. C. Chen, A. N. Willson, Jr., "Rate-distortion optimal motion estimation algorithms for motion-compensated transform video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 2, pp. 147-157, April 1998.

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

[18] T. Berger, *Rate Distortion Theory*, Englewood Cliffs, NJ: PrenticeHall, 1984.

[19] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for lowe-delay communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 172-185, Feb. 1999.

[20] Z. He and S. K. Mitra, "Optimum bit allocation and accurate rate control for video coding via rho-domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 10, pp. 840-849, Oct. 2002.

[21] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Processing*, vol. 9, pp. 1661-1666, Oct. 2000.

[22] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Mag.*, vol. 15, pp. 74-90, Nov. 1998.

[23] T. Wiegand and B. Girod, "Lagrangian multiplier selection in hybrid video coder control," in *Proc. ICIP 2001*, Thessaloniki, Greece, Oct. 2001

[24] T. Wiegand *et al.*, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp.688-703, Jul. 2003.

[25] R. Xiong, X. Ji, D. Zhang, and J. Xu, "Vidwav wavelet video coding specifications," ISO/IEC JTC1/SC29/WG11 MPEG, M12339, 2005.

[26] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation,", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 2, pp. 674-693, July 1989.

[27] R. W. Buccigrossi and E. P. Simoncelli, "Image Compression via Joint Statistical Characterization in the Wavelet Domain," *IEEE Trans Image Processing.*, vol. 8, no. 12, pp. 1688-1701, Dec. 1999.

[28] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Annals. of Mathematical Statistics*, vol. 22, pp. 79-86, 1951

[29] M. Antonini and *et al.*, "Image Coding Using Wavelet Transform," *IEEE Trans Image Processing.*, vol. 11, no. 2, pp. 205-221, April 1992.

[30] S. S. Tsai and H.-M. Hang, “Motion information scalability for MC-EZBC”, *Signal Processing: Image Communication,* Vol. 19, no.7, pp.675-684, Aug. 2004.

[31] A. Secker and D. Taubman, “Highly scalable video compression with scalable motion coding,” *IEEE Trans. Image Processing*, vol. 13, no. 8, pp.1029-1041, Aug. 2004.

[32] C. -Y. Tsai and H. -M. Hang, “ρ-GGD source modeling for wavelet coefficients in image/video coding,” in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME),* pp.601-604, Hannover, Germany, June 2008.

[33] T. Rusert, K. Hanke, and J. Ohm, “Transition filtering and optimized quantization in interframe wavelet video coding,” in *Proc. SPIE Visual Communications and Image Processing (VCIP)*, vol. 5150, pp. 682-693, 2003.

[34] M. S. Crouse, R. D. Nowak, R. G. Baraniuk, “Wavelet-based statistical signal processing using hidden markov models,” *IEEE* Trans. *Signal Processing*, vol. 46, no. 4, pp. 886-902, April 1998.

[35] J. Xu, R. Xiong, B. Feng, G. Sullivan, M. C. Lee, F. Wu and S. Li, “3D subband video coding using barbell lifting,” *ISO/IEC JTC1/SC29/WG11 MPEG*, M10569, 2004.

[36] R. Xiong, J. Xu and F. Wu, “Coding performance comparison between MSRA wavelet video coding and JSVM,” *ISO/IEC JTC1/SC29/WG11 MPEG*, M11975, 2005.

[37] R. Xiong, J. Xu, and F. Wu, “Optimal subband rate allocation for spatial scalability in 3D wavelet video coding with motion aligned temporal filtering,” *Proc. of VCIP 2005*, pp.381-392, Beijing, China, July 2005.

[38] C.-Y. Tsai and H.-M. Hang, “A rate-distortion analysis on motion prediction efficiency and mode decision for scalable wavelet video coding,” Jounal of Visual Communications and Image Representation, accepted , Aug. 2010

[39] C.-Y. Tsai and H.-M. Hang, “One-sided ρ-GGD source modeling and rate-distortion optimization in scalable wavelet video coder,” IEEE Trans. Circuits Syst. Video Technol. accepted, Sept. , 2010