MERL – A MITSUBISHI ELECTRIC RESEARCH LABORATORY
http://www.merl.com

Constant Quality Constrained Rate Allocation for FGS Coded Video

Xi Min Zhang Anthony Vetro Yun Q. Shi
Huifang Sun

TR-2003-30 February 2003

Abstract

This paper proposes an optimal rate allocation scheme for Fine-Granular Scalability (FGS) coded bitstreams that can achieve constant quality reconstruction of frames under a dynamic rate budget constraint. In doing so, we also aim to minimize the overall distortion at the same time. To achieve this, we propose a novel R-D labeling scheme to characterize the R-D relationship of the source coding process. Specifically, sets of R-D points are extracted during the encoding process and linear interpolation is used to estimate the actual R-D curve of the enhancement layer signal. The extracted R-D information is then used by an enhancement layer transcoder to determine the bits that should be allocated per frame. A sliding window based rate allocation method is proposed to realize constant quality among frames. This scheme is first considered for a single FGS coded source, then extended to operate on multiple sources. With the proposed scheme, the rate allocation can be performed in a single pass, hence the complexity is quite low. Experimental results confirm the effectiveness of the proposed scheme under static and dynamic bandwidth conditions.

Published in IEEE Transactions on Circuits and Systems for Video Technology, Feb 2003.

Constant Quality Constrained Rate Allocation for FGS Coded Video

Xi Min Zhang, *Member, IEEE,* Anthony Vetro*, *Member, IEEE,* Yun Q. Shi, *Senior Member, IEEE*, and Huifang Sun, *Fellow, IEEE*

Contact Author Information: Dr. Anthony Vetro, Mitsubishi Electric Research Labs, 558 Central Ave, Murray Hill, NJ 07974. Tel: +1-908-363-0504 Fax: +1-908-363-0550, Email: avetro@merl.com.

Manuscript received August 2001; revised April 2002. This paper was recommended by Associate Editor J.-R. Ohm.
X.M. Zhang and Y.Q. Shi are with the Department of ECE, New Jersey Institute of Technology, Newark, NJ 07102.
A. Vetro and H. Sun are with are with Mitsubishi Electric Research Laboratories, Murray Hill, NJ 07974.

Abstract

This paper proposes an optimal rate allocation scheme for Fine-Granular Scalability (FGS) coded bitstreams that can achieve constant quality reconstruction of frames under a dynamic rate budget constraint. In doing so, we also aim to minimize the overall distortion at the same time. To achieve this, we propose a novel R-D labeling scheme to characterize the R-D relationship of the source coding process. Specifically, sets of R-D points are extracted during the encoding process and linear interpolation is used to estimate the actual R-D curve of the enhancement layer signal. The extracted R-D information is then used by an enhancement layer transcoder to determine the bits that should be allocated per frame. A sliding window based rate allocation method is proposed to realize constant quality among frames. This scheme is first considered for a single FGS coded source, then extended to operate on multiple sources. With the proposed scheme, the rate allocation can be performed in a single pass, hence the complexity is quite low. Experimental results confirm the effectiveness of the proposed scheme under static and dynamic bandwidth conditions.

Keywords

Fine-Granular Scalability, video coding, rate allocation, constant quality, rate-distortion labeling

I. INTRODUCTION

With advances in networking technology, the Internet has become a primary medium for information transmission. A key concern in the delivery of video content over networks is the ability to adapt the outgoing traffic to meet constraints imposed by users and networks. For the transmission of video over fixed bandwidth channels, the video signal is often encoded at a constant bit-rate (CBR). To account for minor fluctuations in the bits produced at each frame, the output bits of an encoder are sent to a buffer. Subsequently, it is the buffer that releases bits at a constant bit-rate to the channel.

There are many advantages with CBR coded video, however, it does have certain drawbacks. One drawback is that picture quality fluctuates. In the case of video recorded on a DVD, picture quality should be constant and there is no need to impose CBR restrictions. Another drawback of CBR is that it does not provide an efficient means of transmitting video over time-varying heterogeneous networks. Such a network is characterized by varying bandwidth and/or sessions that are established based on available bit-rate (ABR) among many users. In both cases, either to provide constant-quality or improved-quality video, or to fully utilize the link capacity, video coded with variable bit-rate (VBR) is often preferred.

Optimal bit allocation in the rate-distortion (R-D) sense was first addressed by Huang and Schultheiss [1] for transform coded data. This paper focused on the optimal allocation of bits among different quantizers assuming a Gaussian random source. In the context of video coding, rate control methods that

adjust the quantization step size have been proposed [2], [3], where optimal video coding is obtained when all macroblocks have the same R-D characteristic. The above works consider optimal allocation among blocks for single-layer coding schemes, i.e., scalable transmission has not been considered.

Overall, previous optimal rate allocation techniques have provided ways to minimize the overall distortion (rate) subject to rate (distortion) constraints. Using Lagrange multiplier techniques to find the optimal solution is the most common approach. The performance of the above methods heavily depends on the model used. To solve this problem, Lin and Ortega [4] proposed to model R-D characteristics based on measurements of the actual rate-quantizer and distortion-quantizer data. In yet other works, dynamic bandwidth allocation has been studied to smooth the burstiness of compressed video stream [5], [6], however smoothing the variation in quality on a frame-by-frame basis has not been considered in this work.

Users require playback with minimal variation in quality, but dynamic network conditions often make this difficult to achieve with single-layer coding schemes. In a heterogeneous network environment, such as the Internet, variable channel conditions can damage the integrity of the reconstructed video. However, if the server can transmit only the important data at a reduced rate, congestion is prevented and the overall video quality is improved considerably. On the other hand, particular clients may expect high quality video provided that they have enough bandwidth and a capable decoder that can handle the higher data rates. With scalable coding, a multimedia server may store one copy of high quality video bit stream and deliver only part of the bit stream depending on the client demand and channel condition.

Recently, Fine-Granular Scalability (FGS) coding [7] and Fine-Granular Scalability Temporal (FGST) coding [8] have been proposed and adopted as amendments to the the MPEG-4 standard [9]. The scalable coding with fine granularity is a radical departure from traditional scalable coding schemes. With traditional scalable coding techniques, the content would be coded into a base layer and possibly several enhancement layers, where the granularity is only as fine as the number of enhancement layers that are formed. As a result, the resulting rate-distortion curve resembles a step-like function. In contrast, FGS provides an enhancement layer that is continually scalable. This is accomplished through a novel bit-plane coding method of DCT coefficients in the enhancement layer, which allows the enhancement layer bitstream to be truncated at any point. In this way, the quality of the reconstructed frames is proportional to the number of enhancement bits received. The principle of FGS coding is illustrated in Fig. 1. Since the prediction is always based on the base-layer, the bit stream of each frame can be truncated at any point without affecting subsequent frames. However, the coding efficiency is sacrificed to some

extent compared with single-layer coding schemes. In order to obtain a good balance between granularity and coding efficiency, Progressive Fine Granular Scalability (PFGS) coding has been proposed [10]. In contrast to FGS coding, this framework supports prediction from an improved reference frame, which is essentially the base layer with a portion of the enhancement layer added.

An important point to emphasize is that the standard itself does not specify how any form of rate allocation should be done. In the FGS/FGST framework, there are several unique types of rate allocation that one may consider, e.g., the truncation of enhancement layer bits, the optimal allocation of rate between the base and the enhancement layers, as well as the temporal-SNR trade-off proposed in [8]. This paper deals only with the truncation of enhancement layer bits to achieve constant quality and is independent of the original bit allocation between base and enhancement layers. The rate and quality of the base layer is assumed to be a lower bound. Enhancement layer bits cover the range of bit-rates from this lower bound to near lossless quality. Also, once the enhancement layer bitstream has been generated, it is stored at the server and re-used many times. According to e.g., network characteristics, an appropriate number of bits will be allocated to a frame and transmitted. Wang, et al. [11] have studied the problem of rate allocation in the enhancement layer for the PFGS coding scheme. In their paper, an exponential model is used to realize optimal rate allocation. Average PSNR improvements in the range of 0.3 to 0.5db have been reported.

In this paper, we consider an optimal rate allocation strategy for FGS and FGST coded bitstreams that achieves constant quality reconstruction of frames under a dynamic rate budget constraint. In doing so, we also aim to minimize the overall distortion at the same time. The basic concept of the proposed scheme is illustrated in Fig. 2. The scheme includes an FGS encoder, a rate-distortion extractor and enhancement-layer VBR transcoder. The FGS encoder produces a base layer and an FGS enhancement layer bitstream. The rate-distortion extractor is based on a novel R-D labeling scheme to characterize the R-D relationship of the source coding process. Specifically, a set of actual R-D points are sampled in the encoding process and linear interpolation is used to estimate the real R-D curve of the enhancement layer signal. The enhancement layer VBR transcoder is based on a rate allocation method that is based on a sliding window approach. This approach makes use of the R-D information that is generated by the R-D extractor. The transcoder determines the bits that should be allocated per frame so that constant quality among frames is maintained.

The rest of this paper is organized as follows. In section II, we formulate the optimal rate allocation problem, and R-D characteristics of the video sequence are discussed. We then propose a rate-distortion

(R-D) labeling scheme to extract the R-D relationship of source coding in section III. Specifically, a set of actual R-D points are sampled in the encoding process and linear interpolation is used to estimate the actual R-D curve. A sliding window based rate allocation approach is then proposed in section IV. Experimental results demonstrate that our method can be used to effectively minimize the variation in quality of the reconstructed frames in section V. The conclusions of our work are presented in section VI.

II. Background of Optimal Rate Allocation

It is known that differential sensitivity has a significant impact on the human's visual perception [12]. Therefore, we focus on minimizing the variation in quality. In this section, we first study the problem of optimal rate allocation from the viewpoint of minimizing average distortion. Then, we investigate the relationship between minimizing the average distortion and minimizing the variation in quality. Some important points are clarified for completeness.

Let D_i and R_i be the distortion and rate at each frame i, respectively. To minimize the average distortion, we can equivalently minimize the cost function, $J(\lambda)$,

$$J(\lambda) = \sum_{i=0}^{N-1} D_i(R_i) + \lambda \sum_{i=0}^{N-1} R_i, \quad \text{subject to} \quad \frac{F_s}{N} \sum_{i=0}^{N-1} R_i \leq R_{budget} \tag{1}$$

where N is the total number of frames, λ is a Lagrangian multiplier, F_s is the source frame rate, and R_{budget} denotes the available bandwidth. Given the Gaussian model, $D(R) = a\sigma^2 2^{-2R}$, where R denotes the average bits per pixel, σ^2 is the signal variance, and a is a constant that is dependent on the pdf of the signal and quantizer characteristics, the solution to the above problem is given by,

$$R_i = -\frac{1}{2} log_2 \frac{\lambda}{2a\sigma_i^2} \tag{2}$$

If the given constraint is satisfied with equality, then the optimal rate allocation is,

$$R_i = \frac{R_{budget}}{F_s} + \frac{1}{2} log_2 \frac{\sigma_i^2}{(\prod_{j=0}^{N-1} \sigma_j^2)^{\frac{1}{N}}} \tag{3}$$

In the above result, one constraint is ignored. That is $R_i \geq 0$. For a CIF resolution video sequence, one frame contains 352×288 pixels. Since the unit of rate is bits per pixel (bpp) in the selected model, R_{budget} will be very small at low bit-rates. According to eqn. (3), it is possible to get $R_i < 0$. For example, let the rate budget be 10 kbps and F_s = 30Hz. Then, R_{budget}/F_s = 0.00329 bpp. Without loss of generality, let $N = 10$, $\sigma_1^2 = 20$ and $\sigma_j^2 = 25$ for all $j > 1$. For this set of input parameters, R_1=-0.1449 bpp, which equates to -14.355 kb/frame. For $j > 1$, R_j=1.9667 kb/frame. Since a negative amount of bits have been

allocated, this result cannot be used for bit allocation. By adding the constraint that $R_i \geq 0$, the solution to the problem can be re-written as,

$$R_i = (-\frac{1}{2} log_2 \frac{\lambda}{2a\sigma_i^2})^+ \tag{4}$$

where $(x)^+$ denotes the positive part of x. In this way, the rate constraint given in eqn. (1) would be satisfied with equality. This result is referred to as *water filling* in information theory. It is noted that the solution given by eqn. (4) will give different results from solution given by eqn. (3). Considering the same input parameters as before, the solution given by eqn. (4) is R_i=0 and R_j=0.37 kb/frame for $j > 1$.

To minimize the variation in quality, we can minimize the following function,

$$\xi = \sum_{i=0}^{N-2} |D_i(R_i) - D_{i+1}(R_{i+1})|, \quad \text{subject to} \quad \frac{F_s}{N} \sum_{i=0}^{N-1} R_i \leq R_{budget} \tag{5}$$

Ideally, there is no variation in quality among neighboring frames, i.e., $\xi = 0$. It can be shown that the solution to eqn. (3) for minimizing the overall distortion given the model $D(R) = a\sigma^2 2^{-2R}$ leads to equal quality among frames, i.e., $D_0(R_0) = D_1(R_1) = \cdots = D_{N-1}(R_{N-1})$ [11]. More generally, it can be shown that this result holds for any exponential model, b^{-kx}, in which each frame has the same constant k in the exponent.

We can also consider if the converse is true. That is, whether constant quality also leads to minimum overall distortion. It can be shown that for a monotonically decreasing R-D function, $D(R_i)$, of each frame, the solution that satisfies the problem given by eqn. (5) is unique, given the additional constraint that $D_i(R_i) = D_j(R_j)$, $i \neq j$, $0 \leq j \leq N-1$.

Considering all of the discussion above, we have the following result: *If the R-D relationship $D(R_i)$ of each frame is exponentially decreasing with the same constant in the exponent, the solution to the problem given by eqn. (5) for minimizing the variation in quality is also the solution to the problem given by eqn. (1) for minimizing the overall distortion*.

According to the above result, it is possible for one rate allocation scheme to provide constant quality under an overall rate constraint, while also minimizing the overall distortion in an optimal way. This analysis provides a new perspective with regard to the optimal rate allocation problem for video coding schemes.

III. R-D Extraction of FGS Coded Video

In this section, we propose a scheme to extract the R-D characteristics of FGS coded video. An R-D labeling scheme is used to provide the necessary R-D information about the source coding. This method is mainly used to overcome problems caused by the inaccuracy of closed form models at low bit-rates.

A. Motivation

The performance of the model based methods for rate allocation depends heavily on the accuracy of the selected model. We have found that the commonly used exponential model is not suitable to accurately model the R-D properties of FGS enhancement layer data at low bit-rates. This is consistent with the classic theory on this subject [13].

A comparison of the actual R-D curve for FGS with the exponential R-D model is shown in Figs.3(a) and (b). From these plots we can see that the exponential model is not sufficient to characterize the FGS coded video. The reasons are as follows. In FGS coding, the lower bit rate is obtained by truncating the enhancement layer. Cutting the bits within a bit-plane is equivalent to reducing the quality in only part of the frame. Therefore, after decoding, the quality is not uniform across parts of the frame. Therefore, the exponential model is not accurate to describe the relationship between the rate (obtained by truncation) and the variance of the FGS residual component.

B. R-D Labeling

To overcome the problems described above, we propose to use a set of R-D label parameters to approximate the complete R-D relationship. Piecewise interpolation between data points is used to approximate the curve. It should be noted that the R-D points corresponding to the base layer frames are first extracted. These points provide the starting points from which the R-D curves are formed. The validity of this approach is based on the following two assumptions: (a) the required R-D label parameters are achievable and (b) the R-D labels used have a low overhead. Given this, we present an efficient R-D label extraction scheme for FGS and FGST coding.

In this scheme, the R-D information for the enhancement layer can be obtained either during the encoding process for real-time operation, or from stored bitstreams after the entire video has been encoded [8]. Since the variance of the enhancement layer data is invariant of the DCT, the specified distortion can be obtained either in the DCT domain or in the spatial domain.

Fig. 4(a) is a block diagram of a R-D extractor that determines rate and distortion sample points in the spatial domain. The enhancement layer bitstream is first passed through a bitstream controller. The function of this controller is to first determine the sample rate points. The sample rate points may be linearly spaced or determined according to a pre-specified function. The sample rate points are recorded as a first part of each R-D pair. Based on each rate point, the specified number of bits are used to reconstruct an FGS residual signal. The reconstruction is performed using a bit-plane VLD, bit-plane

shift and IDCT. The reconstructed FGS residual is subtracted from the original FGS residual to yield an error signal. The distortion is then calculated based on the spatial domain error to yield a distortion sample point, which is the second part of each R-D pair. This process is repeated for multiple rate sample points to yield a set of R-D pairs.

Fig. 4(b) is a block diagram of an alternate R-D extractor that determines rate and distortion sample points in the DCT domain. The process is similar to the process shown in Figure 4(a), except that no IDCT is taken to yield a reconstructed FGS residual in the DCT domain. This reconstructed FGS signal is subtracted from the FGS residual in the DCT domain and the DCT error is used to determine the corresponding distortion sample points.

Although a model that can accurately represent the R-D relationship of video coding has not been found yet, the exponential decreasing property of R-D relationship has been adopted. To capture this property, piecewise exponential interpolation can be used to estimate the R-D relationship. In the following, we compare the performance of piecewise exponential interpolation and piecewise linear interpolation provided. We will show that piecewise linear interpolation is better than piecewise exponential interpolation with the probable selected sampling position and sampling interval. Only a small amount of side information is sufficient to accurately approximate the actual R-D curve.

Consider any two neighbor sampling point $D(R_m)$ and $D(R_n)$ along the R-D curve, where $R_m < R_n$. Let $\Delta R = R_n - R_m$ denote the difference between the two rates. The piecewise linear model is given by,

$$D_L(R) = D(R_m) - \frac{D(R_m) - D(R_n)}{\Delta R}(R - R_m) \tag{6}$$

where $D_L(R)$ is the distortion at point R, and $R_m \leq R \leq R_n$. Alternatively, using the piecewise exponential model to approximate the actual characteristics, the distortion is given by,

$$\begin{aligned} D_E(R) &= \sigma^2 e^{-\eta(R_m + R - R_m)} \\ &= D(R_m) e^{-\eta(R - R_m)} \end{aligned} \tag{7}$$

where η is obtained from two neighboring sampling points. Given that $D(R_n) = D(R_m)e^{-\eta \Delta R}$, we have,

$$\eta = \frac{1}{\Delta R} log \frac{D(R_m)}{D(R_n)} \tag{8}$$

In order to compare the performance of exponential and linear interpolation, we encode the Foreman and Carphone sequence at CIF resolution using FGS coding. The encoding frame rate is fixed at 30 fps. We first let ΔR=16 kb/frame and uniformly sample along the R-D curve. It is found that linear

TABLE I

COMPARISON OF MODEL ACCURACY USING LINEAR AND EXPONENTIAL INTERPOLATION. MD DENOTES MAXIMUM DIFFERENCE, AD DENOTES AVERAGE DIFFERENCE, AND KBPF DENOTES KBITS PER FRAME.

Foreman						
Rates	R=5 kbpf	R=10 kbpf	R=20 kbpf	R=40 kbpf	R=80 kbpf	R=120 kbpf
MD_{lin}	2.93%	1.72%	2.81%	2.3%	2.2%	1.1%
MD_{exp}	3.73%	4.61%	8.01%	5.9%	16.2%	12.2%
AD_{lin}	0.72%	0.37%	1.23%	0.37%	0.94%	0.28%
AD_{exp}	1.38%	2.90%	5.88%	4.2%	9.4%	10.3%
Carphone						
Rates	R=5 kbpf	R=10 kbpf	R=20 kbpf	R=40 kbpf	R=80 kbpf	R=120 kbpf
MD_{lin}	2.02%	2.97%	2.72%	2.06%	1.33%	2.15%
MD_{exp}	3.71%	7.56%	9.22%	10.52%	8.43%	15.03%
AD_{lin}	0.76%	1.81%	0.37%	1.04%	0.74%	0.69%
AD_{exp}	1.86%	5.72%	3.94%	9.03%	6.06%	11.85%

interpolation is better in some intervals, while exponential interpolation is better in others. We then sample the R-D curve at the end of each bit-plane. Six different points (bit-rates) are selected at each frame to calculate the maximum and average differences between the approximated and actual values. The results for the first 100 frames of each sequence are summarized in Table I.

The results in Table I demonstrate that the R-D relationship within each bit-plane is more linear than exponential. These results are consistent with both sequences tested and over a wide range of bit-rates. We believe that the reason for this phenomenon is due to the nature of the bit-plane coding method. In bit-plane coding, each quantized DCT coefficient is considered as a binary number of several bits instead of a decimal integer of a certain value [7]. For each 8×8 DCT block, the 64 absolute values are zigzag ordered into an array. A bit-plane of the block is defined as an array of 64 bits, where each element in the array corresponds to a particular bit position of the DCT coefficients. Thus, every bit in a bit-plane has the same impact on the distortion. With this type of coding scheme, it is reasonable to think the relationship between the number of bits and the distortion is linear within each bit-plane.

To summarize, we have found that linear interpolation is better than exponential interpolation when two neighboring points are located within same bit-plane. On the other hand, exponential interpolation

is better than linear interpolation when two neighboring points are located in different bit-planes. Since the actual R-D lines have different gradients for different bit-planes, using linear interpolation to cross two bit-planes leads to a larger deviation than if exponential interpolation is used. In typical FGS coded frames, 7 bit-planes are used. Therefore, 8 sampling points are enough to approximate the R-D curve of each frame, which is a minor overhead.

It is noted that the sampling of R-D data followed by interpolation is seemingly similar to the technique of Lin and Ortega [4] to determine optimal quantization in video encoding. The key difference between their method and our proposed method is that we operate on bit-plane coded data that is produced by a FGS enhancement layer encoder and the method of Lin and Ortega operate on base layer data. To produce corresponding R-D points for base-layer data involves coding the input video with various quantizers and is very computationally demanding and not suitable for real-time. In contrast, the enhancement layer is coded using a bit-plane coding method that produces an embedded bitstream from which R-D points can directly be extracted.

IV. Constant-Quality Rate Allocation

The proposed R-D labeling scheme described in the previous section provides sufficient R-D information that can be used for rate allocation. Given this data, a similar cost function such as that given by eqn. (1) may then be used by placing constraints to minimize the variation in quality among frames. In the literature, an exhaustive search is typically used to find the optimal solution.

As we have shown in section II, the solution that leads to constant quality is also the solution for minimizing the overall distortion provided the R-D relationship $D(R_i)$ of each frame is exponentially decreasing with the same constant in the exponent. This observation motivates us to propose a practical rate allocation method. In our method, instead of placing constraints to minimize the variation, we apply constant quality as the constraint. Based on this constraint and a suitable initial estimation of the constant distortion, the optimal solution can be obtained in one pass. In the following, we first introduce the constant quality constrained rate allocation approach. A sliding window technique is then used to adapt to the channel variation in time.

According to the piecewise linear interpolation scheme as described by eqn. (6), the rate allocation can be calculated by,

$$\begin{cases} \sum_{i=0}^{N-1} R_i = N \times R_{budget}/F_s \\ D_{m_i} - (R_i - R_{m_i})\frac{\Delta D_i}{\Delta R_t} = D_{m_{i+1}} - (R_{i+1} - R_{m_{i+1}})\frac{\Delta D_{i+1}}{\Delta R_{i+1}}, \quad 0 \le i \le N-2 \end{cases} \tag{9}$$

where R_{budget} is the available bandwidth, N is the total number of frames, F_s denotes the source frame-rate, R_i is the optimal rate that should be allocated to frame i to achieve the constant distortion D. Let $\{R_{m_i}, D_{m_i}\}$ and $\{R_{n_i}, D_{n_i}\}$ be two adjacent R-D points such that $D_{m_i} \geq D \geq D_{n_i}$ and $R_{m_i} \leq R_i \leq R_{n_i}$. In the above equations, $\Delta R_i = R_{n_i} - R_{m_i}$ and $\Delta D_i = D_{m_i} - D_{n_i}$ represent the difference in rate and distortion at adjacent R-D points, respectively. The above yields a set of N equations with N unknowns and can be solved by using known methods.

To solve the above equations, we should determine the correct interval first, namely, two correct adjacent R-D points. The most straightforward method is exhaustive search. However, it is time consuming. In our method, we first estimate the initial value of the constant distortion, D, then determine the two adjacent R-D points $\{R_{m_i}, D_{m_i}\}$ and $\{R_{n_i}, D_{n_i}\}$, such that $D_{m_i} \geq D \geq D_{n_i}$. We calculated this initial estimation of D by using the extracted side information, i.e., $D = \frac{\sum_{i=0}^{N-1} D_i}{N}$, where D_i is the distortion associated with uniform bit allocation. This initial estimation is empirically found to be effective. Using the rate associated with uniform bit allocation provides a simple way to approximate the neighborhood within which an optimal rate is located.

A. Sliding Window Approach for Single FGS Source

Since the channel condition is changing with time, the available bandwidth for each frame is varying. Under this condition, we use a sliding-window resource allocation scheme. Let the rate budget for a window of M frames beginning with frame a be denoted by, W_a , and be given by,

$$W_a = \begin{cases} M \times R_{budget}/F_s; & a = 0 \\ W_{a-1} - R_{a-1} + R_{budget}/F_s; & 1 \leq a \leq M-1 \end{cases} \tag{10}$$

where R_{budget} is the available bandwidth at time a and F_s denotes the source frame-rate. For each frame, the rate budget is computed and the rate allocation for the current frame is found based on the set of equations given below.

$$\begin{cases} \sum_{i=a}^{a+M-1} R_i = W_a \\ D_{m_i} - (R_i - R_{m_i})\frac{\Delta D_i}{\Delta R_t} = D_{m_{i+1}} - (R_{i+1} - R_{m_{i+1}})\frac{\Delta D_{i+1}}{\Delta R_{i+1}}, \quad a \leq i \leq a+M-2 \end{cases} \tag{11}$$

If the solution to the above equations is negative for frame i, we let $R_i = 0$ and recompute the solution. Since the rate allocation to each window is changing on a per frame basis, we only need to solve the above function for the current frame R_i.

Although the computational complexity of above method is very low, we can further reduce the computational cost by computing the rate for every set of M frames, rather than on a per frame basis. In this way, the sliding window would move by M instead of 1 frame. At each pass, the rate allocated to each frame in the window would be assigned. This would work best for slowly varying channel conditions. Also, the smoothness can be improved by increasing the size of the sliding window. This can be observed from the calculation of the rate budget for the moving window as given by eqn. (10). Each time the available bandwidth changes, the rate budget for the window is updated accordingly. Then, the influence of this variation is distributed to each of M frames within the current window. Statistically, each frame absorbs $1/M$ of the total variation. Thus, fluctuations between frames are expected to reduce to $1/M$ of the value without window.

With an increase of window size, the R-D information of more frames have to be known before the transmission. If the R-D information has been obtained off-line and stored, a rate control processor has instant access to this data. Since the computation complexity of the proposed method is very low $O(M)$, the computation delay can be ignored. Under the stable channel condition, it is desirable to select a larger window to smooth the fluctuations caused by varying scene complexity. On the other hand, if the channel condition is unstable, we should pay for smoothness at the expense of an initial delay. In this case, a buffer is used to temporarily store the current M frames and adjust the bit-rate allocation among them. In a real application scenario, the window size can adaptively be determined based on the maximum variation among frames, the sensitivity to the initial delay and the target smoothness. The optimal solution will be a balance of these factors.

B. Sliding Window Approach for Multiple FGS Sources

In modern communication system, the server is usually connected to a wide-band network. The downstream channel is typically a CBR channel with high bandwidth. In the transmission of multiple sources over this high bandwidth network, the individual bitstreams are multiplexed and must satisfy a constant aggregate bit-rate. This problem is commonly referred to as the StatMux problem and has been studied in [14] for multiple encoding of MPEG-2 sources and in [15] for the transcoding of multiple MPEG-2 bitstreams.

In these works, the main objective was to utilize the high bandwidth link and maintain constant quality among the multiple sources, where each source is VBR coded and the sum of multiple VBR sources yields a constant aggregate bit-rate. The method that we describe is an extension of this previous work to scalable FGS coded bitstreams. The objective is still the same, but rather than considering rate allocation

of a multiple base-layers, we consider the allocation of bits in the enhancement layer.

Fig. 5 shows a block diagram of our enhancement-layer statistical multiplexer. Each video program is subject the FGS encoder with extractor discussed earlier. This will output a base-layer bitstream, an enhancement layer bitstream and corresponding R-D information. For non-real-time applications, all the output is stored in a storage device. R-D information is sent to a rate control processor that provides rate allocation to each enhancement layer VBR transcoder. For real-time operation, the storage would be bypassed and base and enhancement layer bitstreams would be passed directly to the enhancement layer VBR transcoder. The reduced rate bitstreams are multiplexed, buffered and transmitted over a CBR channel. The buffer provides feedback to the rate control processor on the buffer fullness.

In the following, we describe the operation of the rate control processor by extending the formulation described above for a single source to multiple sources. Similar assumptions are made. Namely, the minimum quality variance across the multiple sources lead to the minimum overall distortion. Let the rate budget for a 2D window of M frames and K sources beginning with frame b be denoted by, W_b, and be given by,

$$W_b = \begin{cases} M \times R_{budget}/F_s; & b = 0 \\ W_{b-1} - \sum_{j=0}^{K-1} R_{j,b-1} + R_{budget}/F_s; & 1 \le b \le M-1 \end{cases} \tag{12}$$

where R_{budget} is now the total bit budget for K sources and $R_{j,i}$ denote the bits used for source j at frame i. For each frame, the rate budget is computed and the rate allocation for the current frame is found using the set of equations given below,

$$\begin{cases} \sum_{j=0}^{K-1} \sum_{i=b}^{b+M-1} R_{j,i} = W_b; \\ D_{m_{j,i}} - (R_{j,i} - R_{m_{j,i}}) \frac{\Delta D_{j,i}}{\Delta R_{j,i}} = D_{m_{j+1,i}} - (R_{j+1,i} - R_{m_{j+1,i}}) \frac{\Delta D_{j+1,i}}{\Delta R_{j+1,i}}; \\ \qquad i = b, \quad 1 \le j \le K-2 \\ D_{m_{j,i}} - (R_{j,i} - R_{m_{j,i}}) \frac{\Delta D_{j,i}}{\Delta R_{j,i}} = D_{m_{j,i+1}} - (R_{j,i+1} - R_{m_{j,i+1}}) \frac{\Delta D_{j,i+1}}{\Delta R_{j,i+1}}, \\ \qquad b \le i \le b+M-2, \quad 0 \le j \le K-1 \end{cases} \tag{13}$$

where $\Delta R_{j,i} = R_{n_{j,i}} - R_{m_{j,i}}$ and $\Delta D_{j,i} = D_{m_{j,i}} - D_{n_{j,i}}$ represent the difference in rate and distortion at adjacent R-D points of source j, respectively. The above yields a M×K equations with M×K unknowns and can be solved for in a straightforward manner as the single source.

V. EXPERIMENTAL RESULTS

To validate the effectiveness of the methods that we have described, we encode the Foreman sequence (CIF resolution) using FGS and FGST coding. The encoding frame rate for the base layer is fixed at 10 fps for both FGS and FGST coding. Three rate allocation methods are tested: uniform bit allocation, Gaussian model based optimal bit allocation and the proposed method. For both Gaussian model based method and proposed method, we select N=3. Fig. 6 shows the distortion for each frame corresponding to different rates, where each group of the first three consecutive frames in the sequence are compared and each bar denotes the distortion of the corresponding frame. Among the three bars, the gray bar denotes the first frame (I frame). For example, the first three bars in each figure demonstrate the distortions of the three frames correspond to rate 480kbps, which allows 16 kbits per fgs-vop and fgst-vop. It is evident from these plots that the proposed method can achieve constant quality across frames for a wide range of bit-rates.

Figs. 7(a) and (b) compare our sliding window approach to uniform bit allocation. The base layer is encoded with two sets of quantization parameters and the enhancement layer is allocated a rate of 600 kbps, which allows 20 kbits per fgs-vop and fgst-vop with uniform bit allocation. The distortion for each method is plotted over 100 consecutive frames and the results indicate that the quality becomes constant after only a few frames with the proposed approach, while the quality obtained by the uniform bit allocation method contains significant variation. Moreover, the average MSE distortion is decreased from 35.14 to 34.91 in (a), and decreased from 46.31 to 45.50 in (b).

To test the performance of our sliding window approach on multiple source rate allocation, we encode the Foreman sequence, Coastguard sequence, Carphone sequence and Mobile sequence (all CIF resolution) using FGS and FGST coding. The encoding frame rate for the base layer is fixed at 10 fps for both FGS and FGST coding. Figs. 8(a) and (b) compare our sliding window approach to uniform bit allocation. In our experiment, we encoded the sequences with quantization parameters (I:30,P:30,B:30) for the base-layer and FGS layer coding, and (I:28,P:28,B:28) for FGST layer coding. The enhancement layer is allocated a rate of 1320 kbps, which allows 44 kbits per fgs-vop and fgst-vop with uniform bit allocation. The distortion for each method is plotted over 100 consecutive frames. The results using uniform bit allocation indicate that the sequences have significant quality difference. This is due to the different complexity associated with each sequences. Besides the inter-difference (among the sequences), the intra-fluctuations among the frames within the same sequence cannot be avoided by uniform rate allocation. On the other hand, almost constant quality is obtained with the proposed approach. The average

distortion is decreased from 30.70 to 27.66.

To test the performance of our sliding window approach corresponding to different window size under changing channel condition, we encode the Coastguard sequence using FGS and FGST coding. All the parameters are same as above. In this experiment, we concatenate the same sequence three times to generate a longer sequence that is 10s long. Fig. 9 shows the channel condition used in our simulation. The bandwidth begins at 1920 kbps for the first 3.33s, drops to 1440 kbps for next 3.33s and then recovers back to 1920 kbps in for the final 3.33s. It should be noted that these values and changes in the bandwidth may be estimated from other channel conditions, such as a packet loss ratio or the occurrence of burst errors [16]. Fig. 10 shows the result of our sliding window approach for a window size equal to 0 frames (uniform bit allocation), 20 frames (0.67s), 60 frames (2s) and 160 frames (5.33s). The results indicate that the quality becomes smoother with the increase of the window size.

We also test the performance of our approach on multiple sources under changing channel condition. The Foreman sequence, Coastguard sequence and Carphone sequence (all CIF resolution) are encoded by FGS and FGST coding. All the parameters and the channel are same as above. We again obtain the 10s sequences by concatenating the same sequences three times. Figs. 11(a) and (b) compare our sliding window approach to uniform bit allocation. The distortion for each method is plotted over 300 consecutive frames with the window size of 10s. The results indicate that the constant quality is obtained with the proposed approach, while the quality obtained by the uniform bit allocation method contains significant variation.

VI. Concluding Remarks

In this paper, we have studied the problem of rate allocation for FGS coded video. Specifically, we have proposed techniques that are able to minimize the variation in quality. An R-D labeling scheme was proposed to characterize the R-D relationship of the source coding process. A set of actual R-D points are extracted during the encoding process and piecewise interpolation is used to estimate the actual R-D curve of the enhancement layer signal. The main contribution of this paper is the sliding window based rate allocation method. With an initial estimate, the optimal rate allocation can be obtained in one pass with very low computation. The impact of the window size on the variation in the quality has been discussed in detail and supporting experimental results have been provided. Overall, the proposed framework is able to achieve constant quality reconstruction with both static and dynamic channel conditions. Simulation results show that smooth transitions in quality result under dynamic channel conditions can be obtained. Also, we have demonstrated the effectiveness of this framework for both single and multiple sources.

ACKNOWLEDGMENTS

This work is supported in part by New Jersey Commission of Science and Technology via NJCWT, New Jersey Commission of High Education via NJ-I-TOWER, and NSF via IUCRC for Next Generation Video.

REFERENCES

[1] J. Huang and P. Schultheiss, "Block quantization of correlated Gaussian random variables," *IEEE Transaction on Communication Systems*, vol. 11, pp. 289–296, 9 1963.

[2] W. Ding and B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 6, no. 1, pp. 12–19, 1996.

[3] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Transaction on Image Processing*, vol. 3, no. 5, pp. 533–545, 1994.

[4] L.-J. Lin and A. Ortega, "Bit-Rate Control Using Piecewise Approximated Rate-Distortion Characteristics," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 8, no. 4, pp. 446–459, 1998.

[5] L. Zhang and H. Fu, "A novel scheme of transporting pre-stored MPEG video to support video-on-demand (VoD) services," *Computer Communications*, vol. 23, pp. 133–148, 2000.

[6] M. Hamdi, J. W. Robers, and P. Rolin, "Rate control for VBR video coders in Broad-Band networks," *IEEE Journal on Select Areas of Communication*, vol. 16, no. 6, pp. 1040–1051, 1997.

[7] W. Li, "Overview of Fine Granularity Scalability in MPEG-4 video standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 301–317, 2001.

[8] M. van der Schaar and H. Radha, "A hybrid Temporal-SNR Fine-Granular Scalability for Internet video," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 318–331, 2001.

[9] ISO/IEC 14496-2:1999/FDAM 4, *Information technology-Coding of audio-visual objects-Part 2:Visual Amendment 4: Streaming video profile*. N3904, Jan. 2001.

[10] F. Wu, S. Li, and Y.-Q. Zhang, "A framework for Efficient Progressive Fine Granularity Scalable Video Coding," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 332–344, 2001.

[11] Q. Wang, F. Wu, S. Li, Z. Xiong, Y.-Q. Zhang, and Y. Zhong, "A new rate allocation scheme for progressive fine granular scalable coding," *The IEEE International Symposium on Circuits and Systems*, 2001.

[12] Y.-Q. Shi and H. Sun, *Image and Video Compression for Multimedia Engineering - Fundamentals, Algorithms, and Standards*. Boca Raton, FL: CRC Press LLC, 1999.

[13] N. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice Hall, 1984.

[14] L. Wang and A. Vincent, "Joint rate control for multi-program video coding," *IEEE Trans. on Consumer Electronics*, vol. 42, no. 3, pp. 300–305, 1996.

[15] H. Sorial, W. Lynch, and A. Vincent, "Joint transcoding of multiple MPEG video bitstreams," *Proc. Int'l Symp. Circuits and Syst.*, 1999.

[16] D. Wu, Y. T. Hou, W. Zhu, Y.-Q. Zhang, and J. M. Peha, "Streaming Video over the Internet: Approaches and Directions," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 282–300, 2001.

LIST OF FIGURES

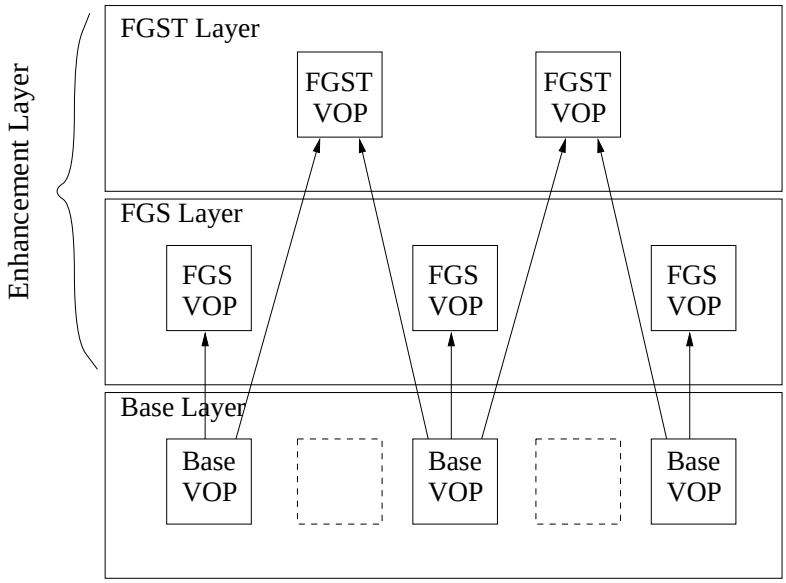


Fig. 1. Illustration of basic structure used for FGS and FGST coding.

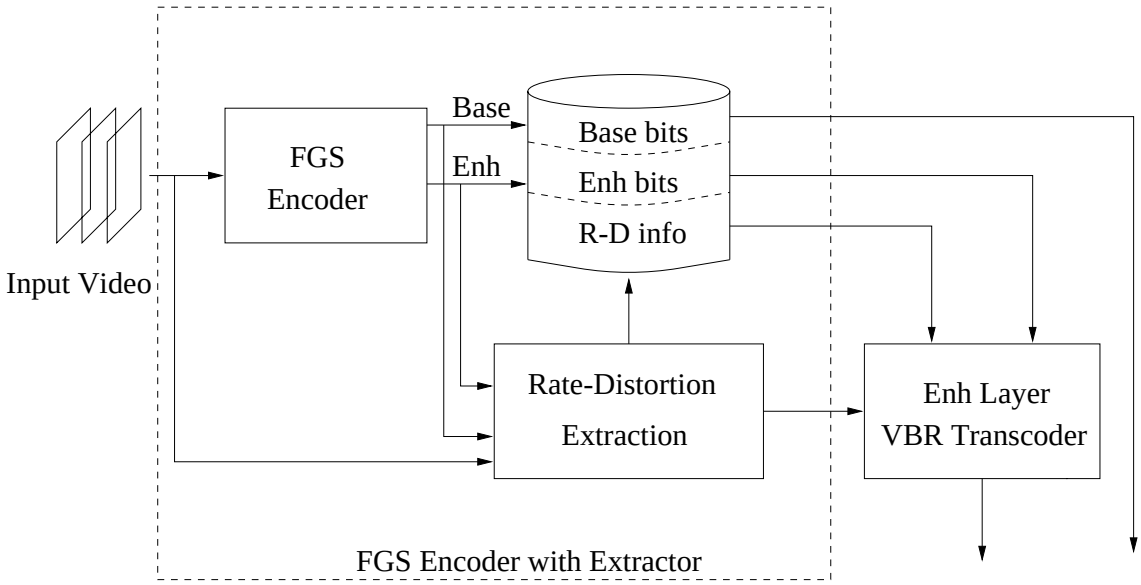
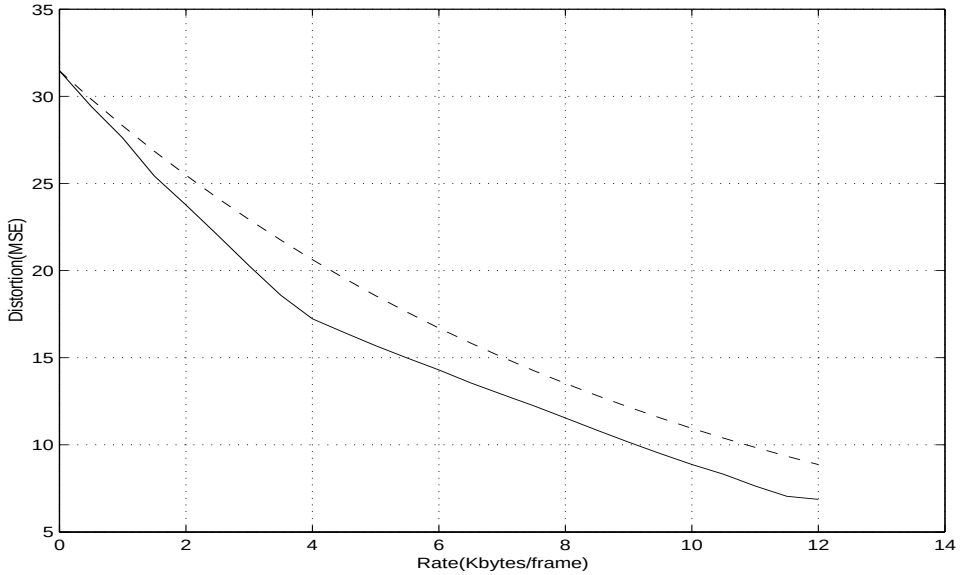
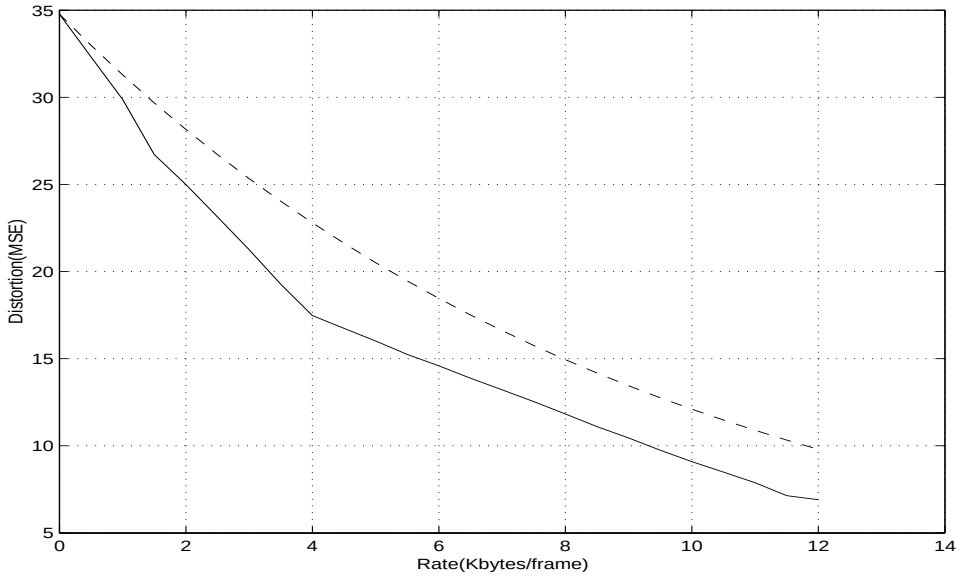


Fig. 2. Overview of proposed rate allocation framework for achieving constant quality.



(a)



(b)

Fig. 3. Comparison of the R-D curves corresponding to the enhancement layer bit-rates of Foreman at CIF resolution: actual curve (solid) vs estimated curve using the exponential model (dashed). (a) First B-frame in FGST layer, (b) Second B-frame in FGST layer.

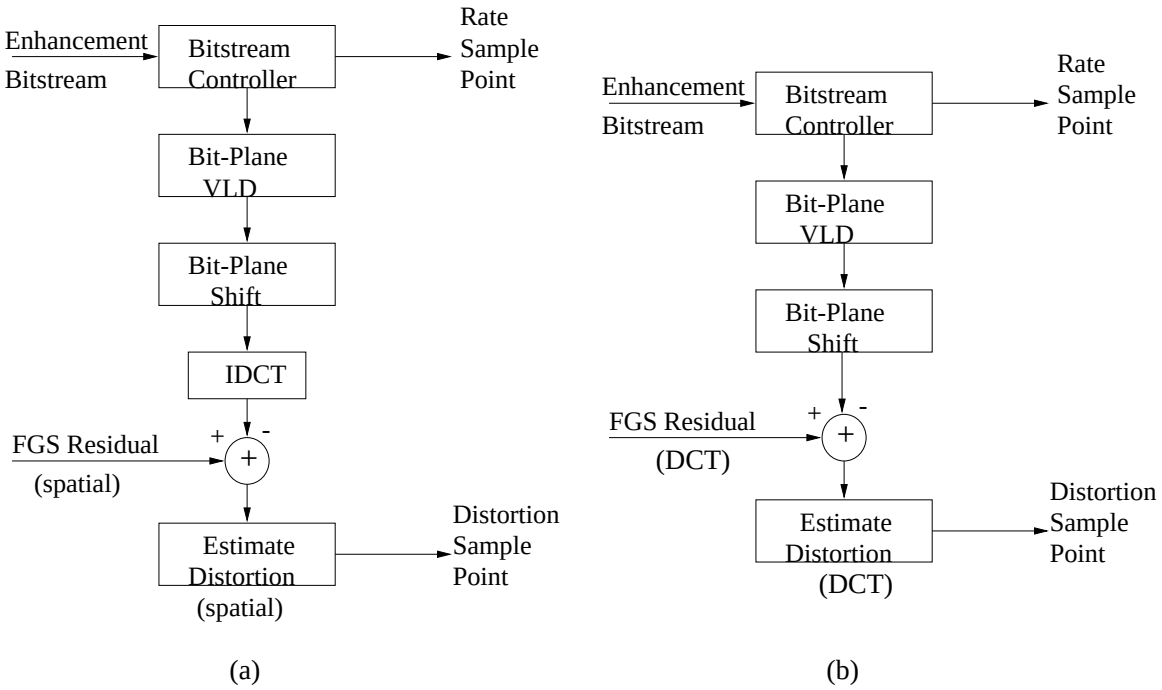


Fig. 4. R-D extraction for FGS enhancement layer bitstreams: (a) spatial domain approach, (b) DCT-domain approach.

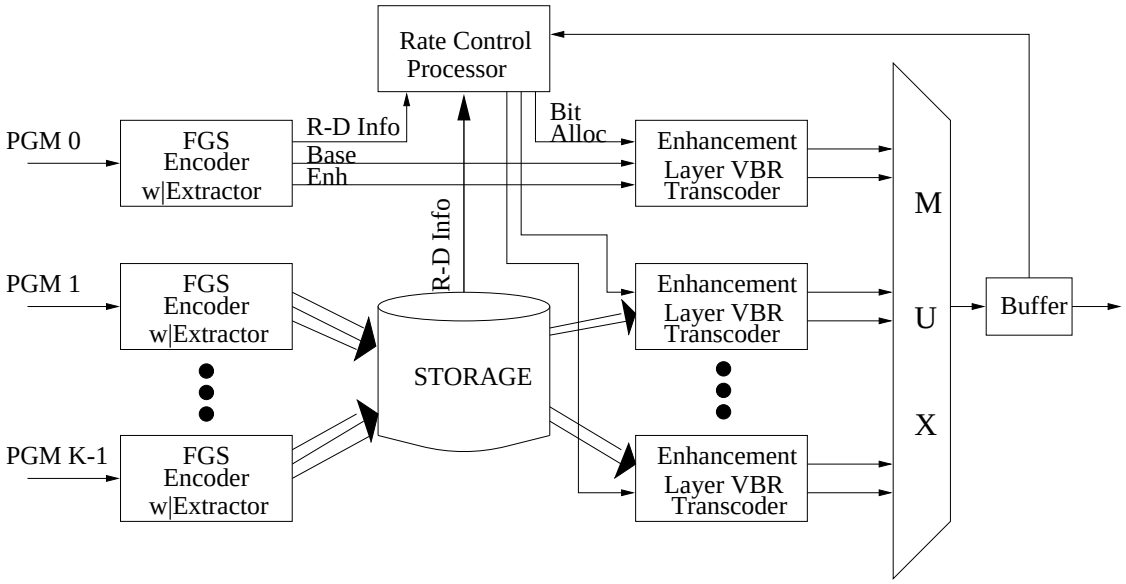
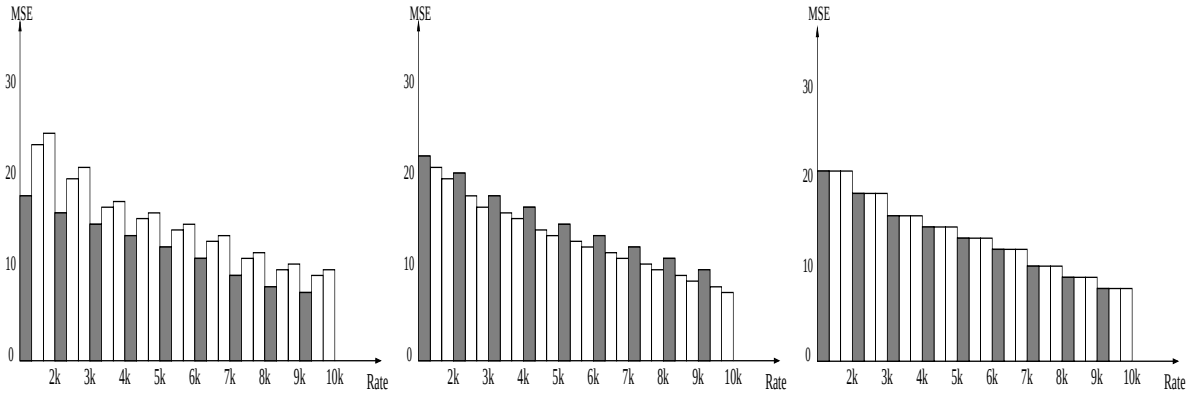
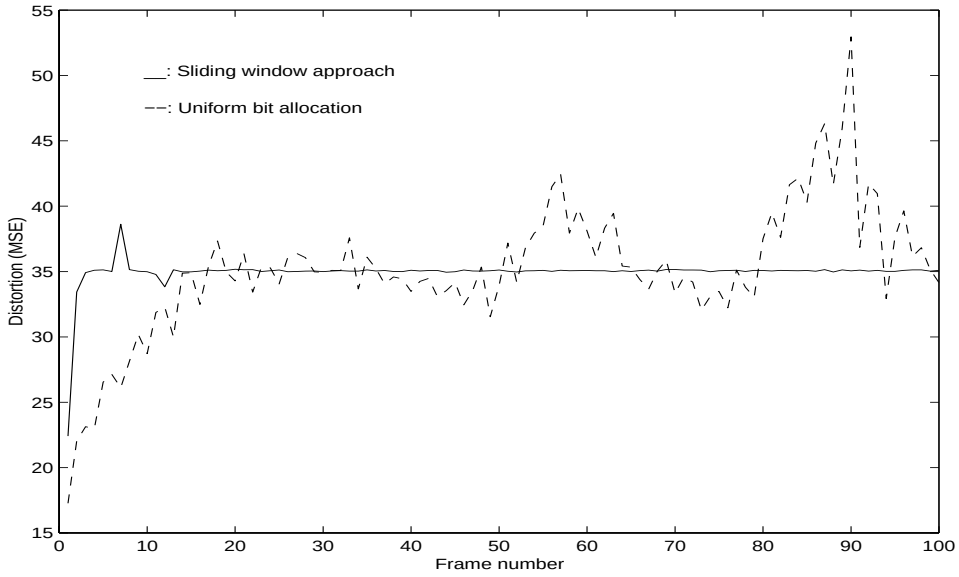


Fig. 5. Block diagram of statistical multiplexer for multiple FGS coded sources.

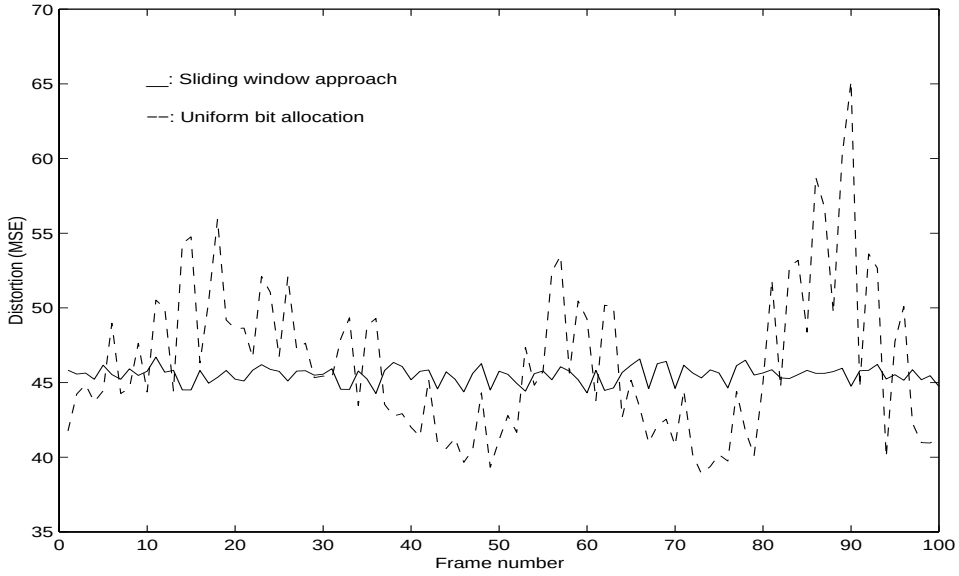


(a) (b) (c)

Fig. 6. Simulation results comparing fluctuation among frames across bit-rates with three methods. Each group of three frames are compared and the gray bar denotes the first frame. (a) Uniform bit allocation, (b) Gaussian model based bit allocation, (c) Piecewise interpolation based bit allocation.

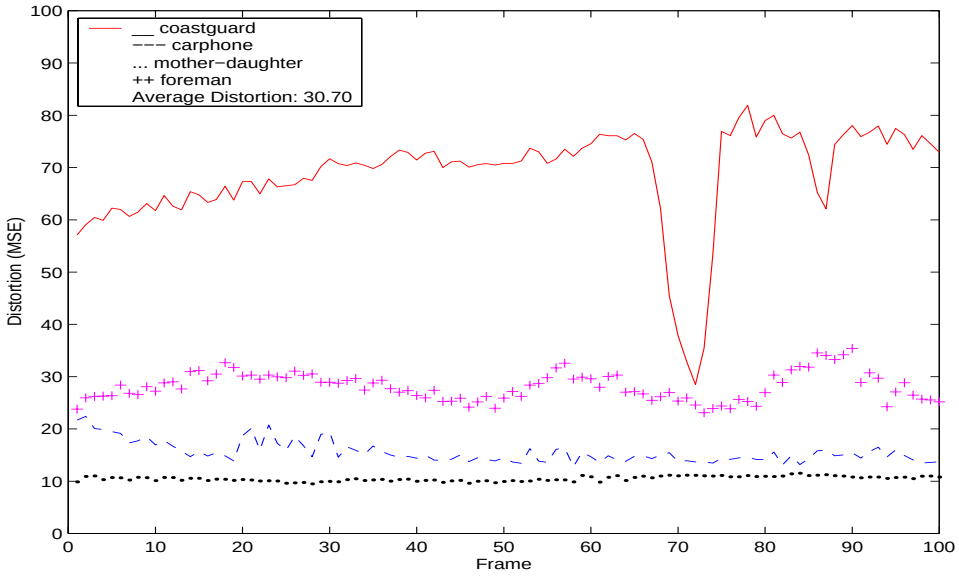


(a)

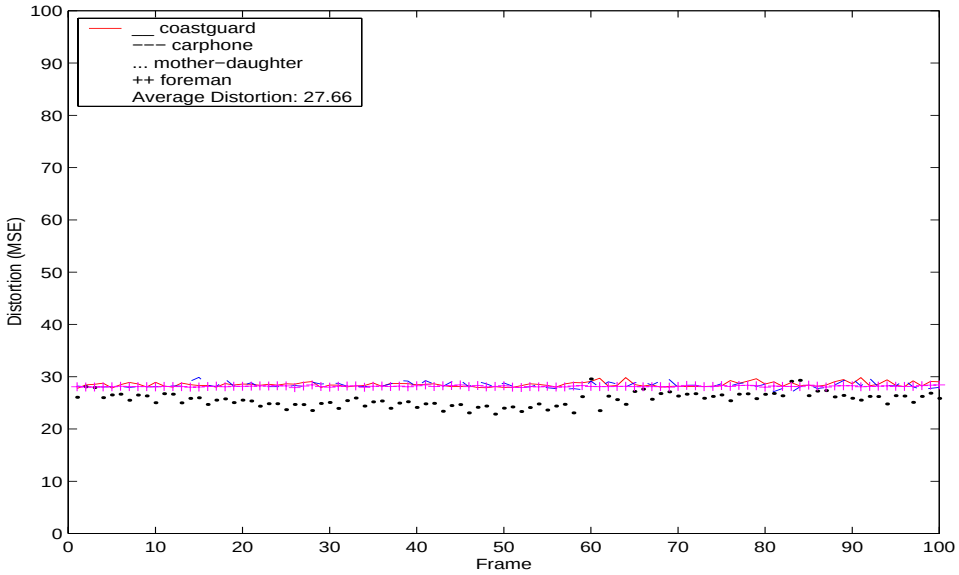


(b)

Fig. 7. Simulation results comparing the distortion per frame using uniform bit allocation and the sliding window based method. (a) Quality variation with quantization, I:10, P:22, B:15, (b) Quality variation with quantization, I:30, P:30, B:30.



(a)



(b)

Fig. 8. Comparison of quality variation with multiple sources transmitted and static channel bandwidth. (a) uniform bit allocation, (b) sliding window approach.

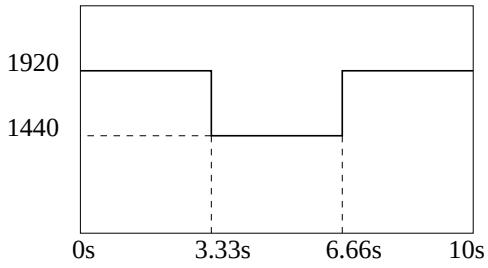


Fig. 9. Dynamic channel condition illustrating available bandwidth (kbps).

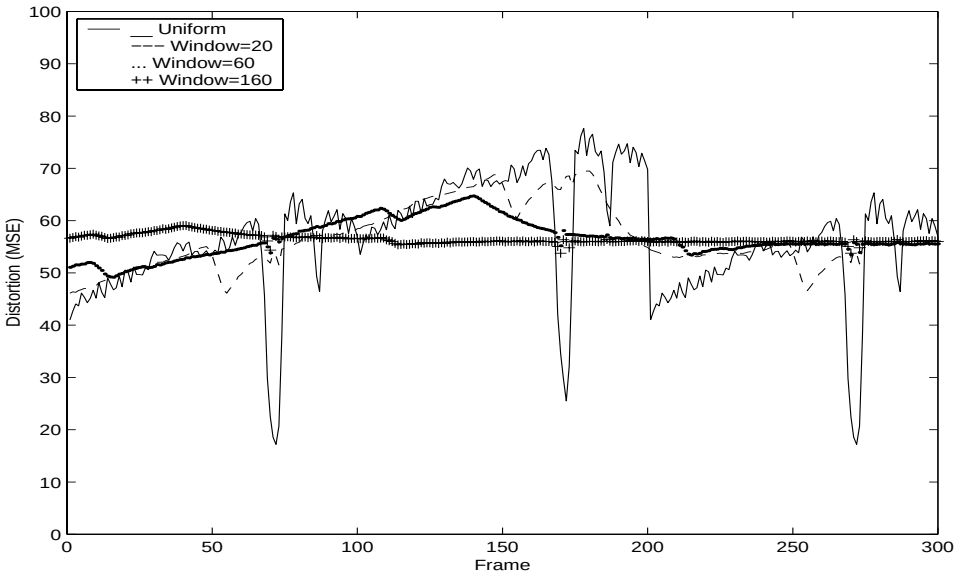
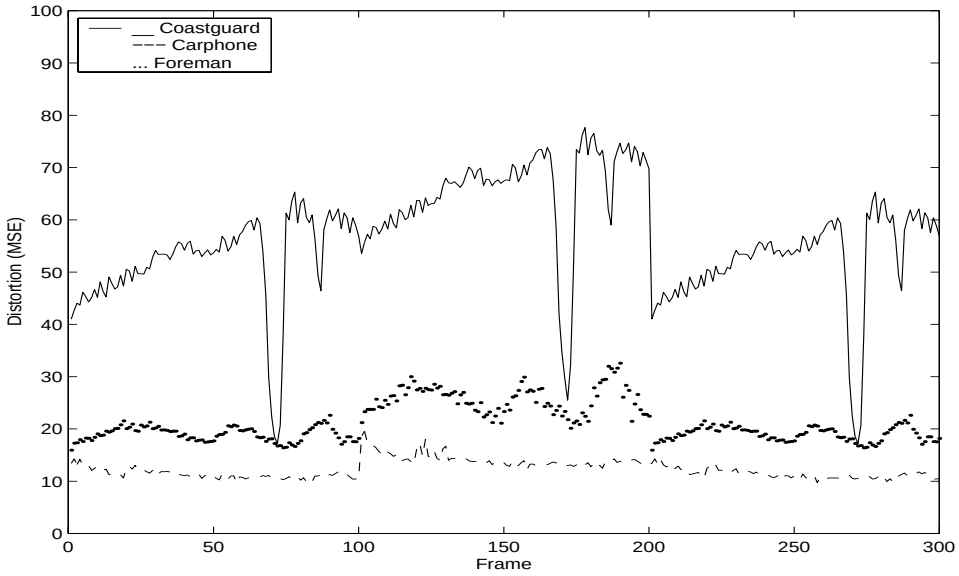
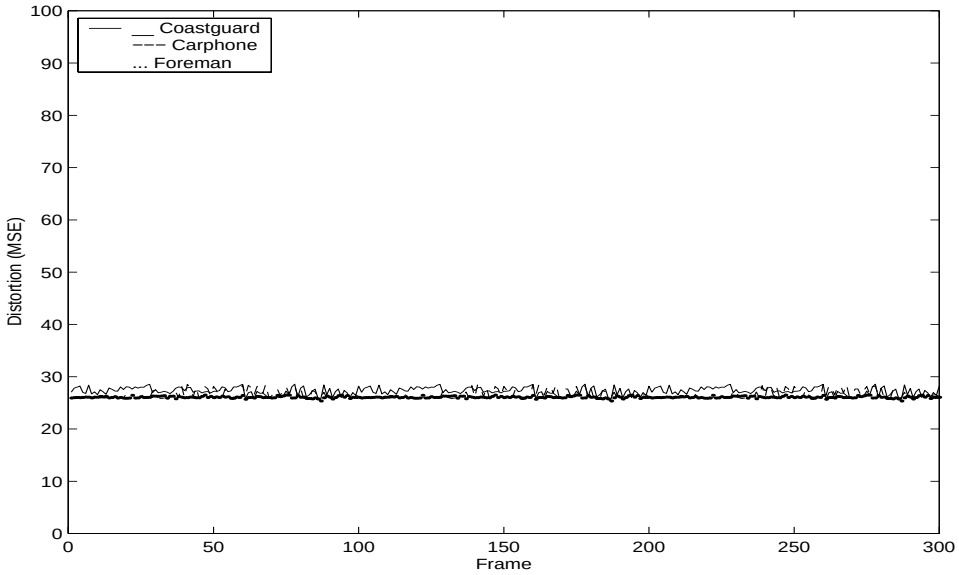


Fig. 10. Simulation results to illustrate impact of window size. Larger window size results in less quality variation.



(a)



(b)

Fig. 11. Comparison of quality variation with multiple source transmission and dynamic channel bandwidth. (a) uniform bit allocation, (b) sliding window approach.