IEEE 6th CAS Symp. on Emerging Technologies: Mobile and Wireless Comm.
Shanghai, China, May 31~June 2, 2004

Analysis and Design of Power Constrained Video Encoder

Yongfang Liang[1], Zhihai He[2], Ishfaq Ahmad[1]

[1]Department of Computer Science and Engineering, University of Texas at Arlington, TX, 76019
[2]Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO, 65211

***Abstract*—In wireless communication networks, an important issue that must be addressed is the limited energy supply of a mobile device, especially in wireless video applications. In this paper, we develop a parametric video encoding architecture, which is fully scalable in power consumption, and establish the power-rate-distortion (P-R-D) model of the video encoding system. Both theoretically and experimentally, we show that by using this P-R-D model, the encoding system is able to automatically adjust its complexity control parameters to match the available energy supply of the mobile device while maximizing the picture quality. The P-R-D model provides a theoretical guideline for system design and performance optimization in wireless video communication under power constraint.**

***Keywords*—Energy consumption, complexity scalability, video coding, wireless video communication.**

I. INTRODUCTION

In wireless communication environment, since the wireless networks are to accommodate mobile users with portable and battery-powered equipments, an important issue that must be addressed is the limited energy supply of a mobile device. The problem becomes even more critical with the power-demanding video encoding functionality integrated into the mobile computing platform. As the power goes down, the performance of video encoding is limited by the available processing power as well as, or rather than, the available transmission bandwidth. Moreover, from the power consumption perspective, efficient video compression significantly reduces the size of the video data to be transmitted, which in turn saves a significant amount of energy in data transmission. On the other hand, more efficient video compression often requires higher power consumption. All of this implies that there is a tradeoff among the bandwidth R, power consumption P, and video quality D. To find the best trade-off during system design and performance optimization, we need an analytic framework to explore the P-R-D behavior of the video encoding system. To achieve flexible management and control of power consumption, we also need to develop a video encoding architecture, which is fully scalable in power consumption.

In wireless video communication, data transmission and video encoding are the two dominant power-consuming operations. In order to extend the battery life, research efforts have been made towards decreasing the transmission energy [2][6], and reducing the energy consumed by video encoding, i.e., lowering the complexity of the video encoder [1][5]. However, to our best knowledge, there has been no video encoder available which is fully scalable in power consumption in the literature. In addition, there has been no analytic framework to model the P-R-D behavior of the video encoding system, which can be used as the guideline for system design and performance optimization in wireless communication.

In our work, we develop a fully complexity scalable video encoder. Specifically, we introduce several control parameters into the video encoder to control the power consumption of the major encoding modules. The analysis of the R-D behaviors of these control parameters results in a comprehensive P-R-D model for the video encoding system. Based on the proposed P-R-D model, we develop a quality optimization scheme to allocate the computational power and maximize the video presentation quality.

This paper is organized as follows. In Section 2, we present the analysis of the complexity scalable video encoder. Section 3 describes the integrated P-R-D model and the design of the power-constrained video encoding. Experimental results are given in Section 4. Concluding remarks are provided in Section 5.

II. POWER SCALABLE VIDEO CODING

From the power consumption point of view, higher complexity of the video encoder, more processor cycles are required in the computation. This implies that the power scalability can be translated into complexity scalability eventually. Let C be the computational complexity, measured by the number of processor cycles per second, we have:

$$C = \Phi(P) \tag{1}$$

where $\Phi(P)$ is a function to match the given power supply to the computational complexity. As the hardware implementation is concerned, the dynamic voltage scaling (DVS), a CMOS circuits design

0-7803-7938-1/04 /$20.00 ?2004 IEEE

technology [7], can be used to achieve this goal. Thus, a power scalable video encoder is reduced to a design of a computational complexity scalable video encoder.

A. Encoder Complexity Analysis

Typical video encoders consist of several major encoding modules: motion estimation and compensation (ME), PRECODING, and entropy coding (ENC). Here, the PRECODING modules include the DCT, IDCT, quantization, dequantization and reconstruction modules. Experimental results show that ME is the most computation-intensive module, which consumes about one-third of the processing cycles, and the PRECODING modules collectively consume about 50% of the total processing cycles. Meanwhile, the ENC module only uses a relatively small amount of the total CPU time, especially at low coding bit rates. Therefore, the ME and PRECODING modules are the two important candidates for computational complexity management. By integrating parametric complexity scalable ME and PRECODING schemes, the whole video encoder can be fully complexity scalable, i.e., energy scalable.

B. Complexity Scalable Motion Estimation Design

The most time-consuming part of ME is the SAD (sum of absolute difference) computation. The ME process is just a sequence of SAD computations to find the macroblock (MB) position with the minimum SAD. This suggests that the computational complexity of the ME module can be controlled by the number of SAD computations, which is denoted by λ_{ME} and used as the complexity control parameter. Thus, the computational complexity of ME, denoted by C_{ME} is simply given by:

$$C_{ME} = \lambda_{ME} \times C_{SAD}, \quad (2)$$

where C_{SAD} represents the complexity of one MB SAD computation. It is well known that the moving objects in the video scene contribute most to the overall visual quality. This suggests that to maximize the video quality under energy constraint, we need to allocate the available λ_{ME} SAD computations among the MBs according to their motion characteristics. In our proposed scheme, the SAD computation allocation is performed according to the motion distribution in one scene. The *Motion History Matrix* (*MHM*) approach developed in our previous work [4] is adopted. The MHM shows not only the motion history but also the motion location, thus we can use this information to estimate the motion distribution and distribute the available SAD computations accordingly. Most importantly, our MHM approach has very low complexity overhead and is very cost effective in practice. After the allocation, the number of SAD computations is limited for each MB, and a complexity controllable ME is employed to find the motion vector.

By dynamically allocating the SAD computations throughout the whole frame and applying the parametric motion estimation scheme, the complexity of the ME module can be controlled freely and the overall video quality is improved under the energy constraint. We refer to this scheme as λ_{ME} - scalability.

The simulation results suggest the following relation between λ_{ME} and the frame SAD S_f:

$$S_f(\lambda_{ME}) = \beta_0 + \beta_1 \times e^{-\beta 2x}, \; x = \lambda_{ME} / \lambda^{max}_{ME} \quad (3)$$

where x denotes the normalized complexity parameter, λ_{ME}^{max} is the maximum value of λ_{ME}, and β_0, β_1, β_2 are the model parameters. Simulation with SSD (sum of square difference) that is more applicable in practical applications yields similar results. In this case, the frame SSD S_f becomes the variance of the difference frame. Hereafter, we assume SSD is used for ME.

C. Complexity Scalable PRECODING Design and Dynamic Rate Control

In typical video encoding, DCT is applied to the difference MB after motion compensation, or the original MB if its coding mode is INTRA. The DCT coefficients might become all zeros after quantization, especially at low coding bit rates. We refer to this MB as an all-zero MB (AZMB). Otherwise, it is called a non-zero MB (NZMB). If we can predict an MB to be an AZMB, all the PRECODING operations can be skipped. This unique property of the AZMB is used to design a complexity scalable scheme for the PRECODING modules. A complexity control parameter λ_{PRE} is introduced. After motion estimation and compensation, we sort all the SSD values in an ascending order. Except the last λ_{PRE} MBs, all the other MBs are forced into AZMBs to skip the PRECODING operations. Let C_{NZMB} be the number of processor cycles needed by the PRECODING operations to finish one NZMB, the complexity of PRECODING modules is estimated by:

$$C_{PRE} = \lambda_{PRE} \times C_{NZMB}, \quad (4)$$

We refer to this type of scalability as λ_{PRE} scalability.

Since the DCT coefficients in the AZMBs are all zeros, which do not need any encoding bits, the entire available bit budget can be allocated to the NZMBs. In this work, we adopt the linear rate control (LRC) algorithm developed in our previous work [3] to perform dynamic bit allocation and rate control. This rate control mechanism allows dynamic bits relocation from the AZMB's to the NZMB's, as well as a near-optimal bits allocation among the NZMB's. As mentioned before, the moving objects have unique significance in subjective video quality evaluation. In motion estimation and compensation, these regions often correspond to the blocks with relatively large SSD values. Using the λ_{PRE} scalability and the dynamic rate control scheme, the AZMB bits are reallocated to these

blocks, resulting in improved overall visual quality. Let R be the target coding bit rate in bits per pixel (bpp).

By using the classic R-D distortion analysis, theoretically, the distortion of our complexity scalable video encoder is given by:

$$D(R;y)=2\times B_0\times(1/2\times(1-y)^2+y(1+a_0y)\times 2^{-2yR/y}),\ y=\lambda_{PRE}/M, \quad (5)$$

where y denotes the normalized complexity parameter, M is the number of MBs in one frame, B_0 is the estimation of the average variance, and γ is a model constant. a_0 is a constant, given by:

$$a_0 = 1/e + 1/e^3 - 1, \quad (6)$$

III. Modeling And Design Of Power Constrained Video Encoding

Based on the proposed complexity scalable video encoding structure, we model the P-R-D behavior of the video encoder and develop optimized power-scalable video encoding schemes to maximize the video quality.

A. Power-Rate-Distortion Model

According to our proposed λ_{ME}-scalability and λ_{PRE}-scalability schemes, the computational complexity of our video encoder is given by:

$$C(R;\lambda_{ME},\lambda_{PRE},\lambda_F) = \lambda_F\times(C_{ME}+C_{PRE}+S\times R\times C_{BIT}) = z\times(C_1x+C_2y+C_3R) \quad (7)$$

where C_{BIT} is the per bit ENC complexity, and S is the size of the frame. $C1$, $C2$, and $C3$ are constants that can be obtained either by theoretical cycle estimation or from simulation statistics.

Using (3) as an approximation of B_0 in (5), the distortion of our video encoder is given by:

$$D(R;x,y) = 2\times(\beta_0+\beta_1\times\exp(-\beta_2x))\times(1/2\times(1-y)^2+y(1+a_0y)\times 2^{-2yR/y}), \quad (8)$$

Equation (8) measures the spatial video quality for a single frame. On the other hand, the frame rate λ_F plays a very important role in video quality evaluation, affecting the temporal video quality. The temporal quality can be estimated from (8) when the parameters R, x and y are all zeros. In our work, we measure the video presentation quality considering different weights for both the spatial and temporal quality, which is given by:

$$D_v = \omega_s\times(\beta_0+\beta_1)+\omega_t\times D(R;x,y), \quad (9)$$

We choose the perceptual weights as follows,

$$\omega_s=(1-z)^2,\ \omega_t=1-\omega_s, \quad (10)$$

where z is the normalized frame rate and $z=\lambda_F/f_{max}$.

From (1) and (7), we have:

$$\Phi(P) = z\times(C_1x+C_2y+C_3R), \quad (11)$$

Thus, for a given power supply level P, to achieve the best video quality, the solution can be obtained by solving the following minimization problem:

$$\min_{\{x,y,z\}} D_v(R;x,y,z),\ s.t.\ \Phi(P)=z\times(C_1x+C_2y+C_3R), \quad (12)$$

B. R-D Optimized Power Constrained Video Encoding Design

Based on the proposed P-R-D model we can find the best configuration of the complexity parameters to maximize the video quality. The optimal complexity parameters (x,y,z) can be obtained using binary search at an adjustment period, say 5 seconds. Our R-D optimized power-scalable video encoder system operates as follows:

Step 1: Model parameters estimation: the parameters in the P-R-D model β_0, β_1, β_2, and γ are estimated from the R-D statistics of previous frames.

Step 2: Optimization: Find the optimal complexity parameter set $\{x,y,z\}$ by solving (12) when the power control module is triggered, say every 5 seconds.

Step 3: Encoder complexity control: Set the encoding frame to $\lambda_F=z\times f_{max}$. Distribute the allocated available computations $\lambda_{ME}=x\times\lambda_{ME}^{max}$ among the MBs and use the complexity controllable ME scheme to get the motion vector. Apply the λ_{PRE} scalability scheme and dynamic rate control after motion compensation.

C. Power-Bit Allocation for Multiple Encoders

Having analyzed the P-R-D behavior of one single video encoder, now we discuss for a given total power constraint and total bit rate budget, how to adaptively perform the power-bit allocation for each individual encoder so as to minimize the total distortion under the power constraint.

The overall video quality of multiple encoders can be judged by using different weight factor ω_i for each video encoder.

$$D=\Sigma\,\omega_i\times D_v^i(R^i;x^i,y^i,z^i), \quad (13)$$

where D_v^i is the individual distortion for the i^{th} encoder. The total power consumption is composed of the individual power consumption (C^i, given by (7)):

$$\Phi(P)=\Sigma\,C^i(R^i,x^i,y^i,z^i), \quad (14)$$

Now for power-bit allocation across multiple N video encoders, the objective is to minimize the total video distortion under the total power (P) and tbit rate constraints (R):

$$\min_{\{R^i,x^i,y^i,z^i\}}\sum_{i=1}^{N}\omega_iD_v^i(R^i;x^i,y^i,z^i),$$

$$s.t.\ \Sigma R^i<R,\ and\ \sum_{i=1}^{N}C^i(R^i;x^i,y^i,z^i)=\Phi(P) \quad (15)$$

Optimization methods such as Lagrange multiplier and penalty function methods can be used to solve the above optimization problem. Further simplification and analysis of the P-R-D model lead to the reduction of the computational load of solving this problem.

IV. Experimental Results

To evaluate the performance of our proposed video encoding system, we implement the P-R-D model into the public domain H.263+ encoder. In our simulations, λ_{ME}^{max} is 50 and f_{max} is 30fps. The video distortion is measured by the mean square error (MSE).

To test the accuracy of the P-R-D model, we run the video encoder over the "Foreman" QCIF sequence at 128kbps and 15 fps for different complexity control parameters (x,y) and measure the corresponding distortion. Fig. 1 shows the comparison of the actual results with the estimated results by the model. We can see that the proposed model is quite accurate. Simulation over other test video sequences yields similar results. Fig. 2 (a)~(c) shows the picture distortion, and the optimal control parameters $\{x,y,z\}$ as functions of the percentage of power consumption. We can see that using the proposed P-R-D model and the optimized power scalable video encoding schemes, our video encoder can adjust its computational complexity and energy consumption according to the energy supply level. From the subjective video quality point of view, as the encoder scales down the power consumption, the video quality degrades from the highly quality motion video to still image.

V. Concluding Remarks

In this work, we have developed a parametric video encoding architecture that is fully scalable in computational complexity. We have also developed a P-R-D analytic framework for video encoding under power constraint. Coupled with the quality optimization scheme, the video encoder is able to determine the best configuration of its complexity control parameters to match the available power supply level of the mobile device. Our work establishes a theoretical basis and provides a guideline in system design and performance optimization for wireless video communication under power constraint.

References

[1] S. M. Akramullah, I. Ahmad, and M. L. Liou, "Optimization of H.263 video encoding using a single processorcomputer: performance tradeoffs and benchmarking," IEEE Trans. On Circuits and System for Video Technology,vol. 11, pp. 901 - 915, August 2001.

[2] Y. Eisenberg, T.N. Pappas, R. Berry, A.K. Katsaggelos, "Minimizing transmission energy in wireless video communications," *ICIP*, pp. 958-961, Oct. 2001.

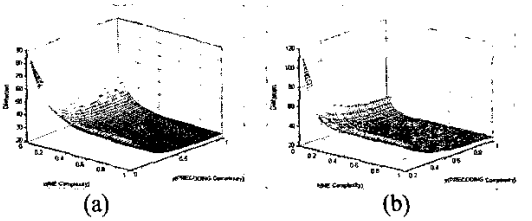
[3] Zhihai He and Sanjit K. Mitra, "A linear source model and a unified rate control algorithm for DCT video coding," *IEEE Transactions on Circuits and System on Video Technology*, vol. 12, pp. 970 - 982, November 2002.

[4] Yongfang Liang, Ishfaq Ahmad, Jiancong Luo, Yu Sun, "Fast Motion Estimation Using Hierarchical Motion Intensity Structure," in *ICME 2004*, Taipei, Taiwan, June 2004, in Press.

[5] X. Lu, Y. Wang, E. Erkip, "Power Efficient H.263 Video Transmission over Wireless Channels," *Proc. Of IEEE ICIP*, 2002.

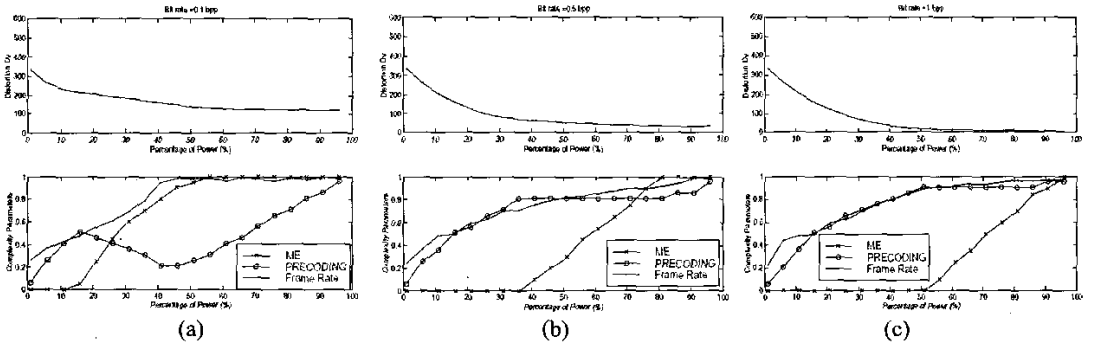
[6] C. E. Luna, Y. Eisenberg, T. N. Pappas, R. Berry, A. K. Katsaggelos, "Transmission energy minimization in wireless video streaming applications," *Proceedings of International Symposium on Telecommunications*, pp. 106-110, 2001.

[7] Rex Min, Travis Furrer, and Anantha Chandrakasan, "Dynamic Voltage Scaling Techniques for Distributed Microsensor Networks," *IEEE Computer Society Workshop on VLSI (WVLSI '00)*, pp. 43-46, April 2000.



(a) (b)

Fig. 1. Complexity distortion surface $D(x,y)$, (a) Estimated results by the model, (b) Actual results.



(a) (b) (c)

Fig. 2. R-D optimized power control for the "Football" CIF video at (a) R = 0.1 bpp, (b) 0.5bpp, (c) 1.0 bpp.