RATE-DISTORTION ANALYSIS OF THE MULTIPLE DESCRIPTION MOTION COMPENSATION VIDEO CODING SCHEME

Shunan Lin[1] *,Anthony Vetro*[2] *and Yao Wang*[1]

[1]Department of ECE, Polytechnic University, Brooklyn, NY, USA,
[2]MERL-Mitsubishi Electric Research Laboratories, Murray Hill, NJ, USA

ABSTRACT

Multiple description motion compensation (MDMC) is a multiple description video coding scheme that has shown good error resilience performance. MDMC enables one to vary coding parameters according to the desired trade-off between coding efficiency and error resilience. To fully utilize this advantage, one needs to establish a set of models, relating the rate, encoder distortion, and the end-to-end distortion after transmission, with the encoder parameters and channel parameters. Using these models, one can find the optimal coding parameters for given channel parameters and rate (or distortion) constraints. In this paper, we formulate and validate the rate and encoder distortion models.

1. INTRODUCTION

The objective of designing a video codec is to maximize the quality of the coded video given a bit rate constraint. In a classic video codec, only quantization distortion, referred to encoder distortion in this paper, is considered, and the design is achieved by choosing an appropriate coding mode and quantization step size. This method, however, is not sufficient for transmission over an error prone environment. In such an environment, the distortion caused by transmission errors, which is referred to as decoder distortion in this paper, must also be taken into account and the design process must consider error resilience issues.

Multiple description coding (MDC) emerged as an attractive coding framework for robust transmission over unreliable networks. There have recently been many MDC schemes proposed (a good review is given in [1]), among which an MDC video codec dubbed multiple description motion compensation (MDMC) [2] has shown good error resilience performance. MDMC can generate two descriptions, each of which includes coded information for alternating frames. For each frame, a linear combination of two motion compensated signals from two previously encoded pictures is used as a prediction signal and the prediction error is coded. When only one description is received, the decoder can only use one of the motion compensated signals for prediction, thus causing a mismatch between the encoder and decoder. This mismatch signal is also coded explicitly at the MDMC encoder, which can help to mitigate transmission error effects.

The cost of increasing error resilience is higher bandwidth consumption. Therefore, there is a trade-off between robustness and redundancy. For MDMC, there are three parameters that control its error resilience and redundancy: Q_0, the quantization parameter of prediction error coding, Q_1, the quantization parameter of mismatch signal coding, and a_1, the linear combination coefficient of motion compensation. The problem of designing an optimal MDMC codec can be formulated as follows. Given the transmission channels' characteristics, such as error characteristic and bandwidth constraint, find the optimal Q_0, Q_1 and a_1, such that the end-to-end distortion is minimized.

It is clear that before solving the above optimization problem, the following questions must be answered: given encoder distortions, what is the effect of transmission error on end-to-end quality; how do the parameters Q_0, Q_1 and a_1 influence the effect of transmission error; and how are the coding parameters chosen to satisfy the bandwidth constraint. In [3], we addressed the first question. To answer the second and third questions, rate-distortion models of the MDMC codec are needed. Although other papers such as [4] have studied the R-D modeling problem for classic video codecs, those R-D models are not sufficient for an MDMC codec since MDMC is more complex and has its own properties. In this paper, we provide these models by establishing the rate-parameter models and distortion-parameter models. The optimal coding parameters can be found using these models.

The rest of this paper is organized as follows. MDMC is briefly reviewed in section 2. In section 3, the proposed models are developed. We provide simulation results to verify the accuracy of those models in section 4 and the paper is concluded in section 5.

This work was done in MERL when the first author was a summer intern there. It is also supported in part by NSF Grant ANI 0081357 and the New York State Center for Advanced Technology in Telecommunication (CATT).

2. REVIEW OF THE MDMC CODING SCHEME

At an MDMC encoder, the central prediction for coding $\psi(n)$, the nth frame, is obtained by

$$\hat{\psi}(n) = a_1\tilde{\psi}_e(n-1) + (1-a_1)\tilde{\psi}_e(n-2), \quad (1)$$

where $\tilde{\psi}_e(n-1)$ and $\tilde{\psi}_e(n-2)$ are motion compensated prediction signals constructed from two previously encoded frames $\psi_e(n-1)$ and $\psi_e(n-2)$ respectively. The prediction error, called central prediction error,

$$e_0(n) = \psi(n) - a_1\tilde{\psi}_e(n-1) - (1-a_1)\tilde{\psi}_e(n-2), \quad (2)$$

is quantized by quantizer $Q_0(\cdot)$ to $\tilde{e_0}(n)$. The coded motion vectors and the central prediction error for even frames are included in description one, and those for odd frames in description two. To circumvent the mismatch between the predicted frames used at the encoder and the decoder when some information is lost in frame $n-1$, the mismatch signal

$$e_1(n) = \psi(n) - \tilde{\psi}_e(n-2) - \tilde{e}_0(n) \quad (3)$$

is quantized by another quantizer $Q_1(\cdot)$, which is typically coarser than $Q_0(\cdot)$, and the output $\tilde{e}_1(n)$ is sent along with other information of frame n. At the decoder, if frame $n-1$ is received, frame n is reconstructed using

$$\psi_d(n) = a_1\tilde{\psi}_d(n-1) + (1-a_1)\tilde{\psi}_d(n-2) + \tilde{e}_0(n) \quad (4)$$

where $\tilde{\psi}_d(n-1)$ and $\tilde{\psi}_d(n-2)$ are motion compensated prediction signals constructed from two previously decoded frames, respectively. If $n-1$ is damaged but frame $n-2$ is received, the decoder reconstructs frame n using

$$\psi_d(n) = \tilde{\psi}_d(n-2) + \tilde{e}_0(n) + \tilde{e}_1(n). \quad (5)$$

In addition, the lost frame $\psi(n-1)$ is estimated based on $\tilde{e}_0(n)$ and even frames are reconstructed using

$$\tilde{\psi}_d(n-1) = \Big(\psi_d(n) - (1-a_1)\tilde{\psi}_d(n-2) - \tilde{e}_0(n)\Big)/a_1. \quad (6)$$

3. RATE AND DISTORTION MODELS OF AN MDMC CODEC

The end-to-end quality of a video sequence coded by an MDMC codec is determined by its encoder distortion and decoder distortion. There are two kinds of encoder distortions in an MDMC encoder: the quantization distortion of the central prediction errors, named central distortion D_0, and that of the mismatch signals, named side distortion D_1. We showed in [3] that the decoder distortion is controlled by D_1, a_1 and the error concealment distortion. Therefore, the end-to-end distortion, D, is a function of D_0, D_1 and a_1 for a given channels' error characteristics, which can be expressed as $D = g(a_1, D_0, D_1)$. From equation (2) and (3), D_1 is a function of a_1, Q_0 and Q_1, where Q_0 and Q_1 are the quantization parameters of quantizers $Q_0(\cdot)$ and $Q_1(\cdot)$, respectively, and D_0 is a function of a_1 and Q_0, i.e., $D_1 = d_1(a_1, Q_0, Q_1)$ and $D_0 = d_0(a_1, Q_0)$. Hence, the optimization problem stated in the last section can be formulated as finding the optimal a_1, Q_0 and Q_1, which minimizes $D = g(a_1, d_0(a_1, Q_0), d_1(a_1, Q_0, Q_1))$, while satisfy the bandwidth constraint. The bandwidth constraint can be expressed as $R_h + R_0 + R_1 < R$, where R_h, R_0 and R_1 are the bit rates spent on syntax header and motion vectors, central prediction errors, and mismatch signals, respectively. R_h can be treated as constant except for very low bit rate coding; R_0 is a function of a_1 and Q_0, and R_1 is a function of a_1, Q_0 and Q_1. In other words, $R_0 = r_0(a_1, Q_0)$ and $R_1 = r_1(a_1, Q_0, Q_1)$.

Given all the five models, $g(\cdot)$, $d_0(\cdot)$, $d_1(\cdot)$, $r_0(\cdot)$ and $r_1(\cdot)$, the above optimization problem can be solved numerically. In [3], the function $g(\cdot)$ has already been determined. The remaining four functions will be developed in following.

3.1. Rate model

When a_1 is fixed, the relationship between R_0 and Q_0 can be expressed using models developed for classic video encoders. So, we use the $R-\rho$ model proposed in [4]. If we assume the motion compensated frames $n-1$ and $n-2$ represent a noisy version of frame n, i.e., $\tilde{\psi}_e(n-i) = \psi(n) + n_i, i = 1, 2$. Assuming n_1 and n_2 are independent, the power spectral density of e_0 is $\Phi_{ee}(\omega_x, \omega_y) = a_1^2\Phi_{n_1n_1}(\cdot) + (1-a_1)^2\Phi_{n_2n_2}(\cdot)$. Using $R_0 = \frac{1}{2}\log(\sigma_e^2/D)$ and $\sigma_e^2 = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\Phi_{ee}d\omega_x d\omega_y$, and assuming the noise signals have white spectrum, the relationship between R_0 and a_1 can be expressed as:

$$R_0 = \frac{1}{2}\log_2(b_1 + b_2a_1 + b_3a_1^2), \quad (7)$$

where b_1, b_2 and b_3 are three constants determined by the quantization distortion and the noise variance, which in turn depends on the video signal statistics, the quality of the reference frames, and Q_0. From the physical meaning of these parameters, we can infer that the absolute value of b_2 is large than b_3 since $\sigma_{n_2}^2$ is usually larger than $\sigma_{n_1}^2$, which is confirmed in our simulations. Furthermore, since a_1 is smaller than 1, we can conclude that $|b_2a_1| \gg |b_3a_1^2|$ is true in most cases. So, the above model can be simplified to

$$R_0 = \frac{1}{2}\log_2(b_1 + b_2 \times a_1). \quad (8)$$

In the above, b_1 and b_2 are constants for a fixed Q_0. Given a fixed value of Q_0, two values of R_0 can be obtained with the $R-\rho$ model for two corresponding values of a_1. Based on the two R_0 values, b_1 and b_2 can be estimated for the given Q_0 value.

As stated in (3), R_1 is a function of a_1, Q_0 and Q_1. As with the $R_0 - Q_0$ model, we also base the relationship between R_1 and Q_1 on the $R - \rho$ model [4] for fixed a_1 and Q_0. From our observation, this relationship is given by

$$R_1 = d_1 + d_2 \times Q_0, \tag{9}$$

in which d_1 and d_2 are two constants for a fixed a_1 and Q_1. We have also found that $\text{abs}(d_2) \ll \text{abs}(d_1)$. Given this, R_1 only varies slightly with changes in Q_0.

To study the relationship between R_1 and a_1, we express e_1 as, $e_1 = a_1(\tilde{\psi}_e(n-1) - \tilde{\psi}_e(n-2)) - q_0$, from equation (3), where q_0 is the the central quantization noise. As with R_0, the relationship between R_1 and a_1 can be approximated by

$$R_1 = \frac{1}{2}\log_2(c_1 + c_2 \times a_1^2), \tag{10}$$

if we assume q_0 is uncorrelated with $n_1 - n_2$. Furthermore, since q_0 is quite small compared with $n_1 - n_2$ and since R_1 varies slowly with Q_0, we can assume c_1 and c_2 are two constants for a fixed Q_1. Given a fixed value of Q_1, two values of R_1 can be obtained with the $R - \rho$ model for two corresponding values of a_1 and any fixed value of Q_0. Based on the two R_1 values, c_1 and c_2 can be estimated directly.

3.2. Distortion Model

In this section, we define the distortion as mean square error (MSE). The quantization distortion of central prediction error, D_0, from equation (2), depends on the predictor a_1 and its quantizer parameter Q_0. First, we will study the influence of Q_0 on D_0 when a_1 is fixed. Then, we consider the influence of a_1.

Due to the orthogonality of DCT transform, calculating D_0 is equivalent to calculating the quantization distortion of the DCT coefficients of e_0. For a given quantization method, the quantization distortion depends on the probability distribution of DCT coefficients. Since AC coefficients are approximately generalized Gaussian distributed [5] and this distribution can be approximated as a Laplacian [6], the distortion of the kth AC coefficient in an inter-coded block, can be expressed as,

$$D_0(k) = 4\eta_k^2 - (6Q^2 + (6\eta_k + 1)Q - 2\eta_k - 1)e^{-2.5Q/\eta_k} - \frac{(2Q^2 + 4(\eta_k + 1)Q)e^{-4.5Q/\eta_k}}{1 - e^{-2Q/\eta_k}} \tag{11}$$

if Q is even, and

$$D_0(k) = 4\eta_k^2 - (6Q^2 + (6\eta_k - 1)Q - 2\eta_k - 1)e^{-(2.5Q-0.5)/\eta_k} - \frac{(2Q^2 + (4\eta_k - 2)Q)e^{-(4.5Q-0.5)/\eta_k}}{1 - e^{-2Q/\eta_k}} \tag{12}$$

if Q is odd.

In the above equations, η_k is equal to $E|e_0(k)|$, the averaged absolute value of kth DCT coefficient, which was determined by the variance and distribution function shape of that coefficient. Also, it is assumed that the quantization method of H.263 is used and Q is the corresponding quantizer parameter. Unlike AC coefficients, DC coefficients are better approximated by a uniform distribution. So, the distortion of DC coefficients is

$$D_0(0) = 25Q^2/12. \tag{13}$$

The total distortion D_0 of one frame can be calculated as

$$D_0 = \frac{1}{64}\sum_{k=0}^{63} D_0(k), \tag{14}$$

For a fixed a_1, the η_k of e_0 can be calculated and D_0 can then be estimated using the above models. For different a_1's, it turns out that the signals' probability distributions, and hence η_k's, remain almost unchanged. Therefore, D_0 is almost unchanged with variation in a_1.

From equation (3), the quantization distortion of the mismatch signal, D_1, depends not only on its quantizer parameter, Q_1, but also on a_1 and Q_0. We find the distribution of the mismatch signal's AC coefficients can still be approximated as a generalized Gaussian and the DC coefficients as a uniform distribution. So, given a fixed a_1 and Q_0, equations (11,12,13 and 14) can also be used to calculate D_1, in which η_k is equal to $E|e_1(k)|$.

When using different a_1 or Q_0, the parameters of the generalized Gaussian distribution functions are different, and consequently, D_1 is different. However, the relationship between a_1 (or Q_0) and those parameters is difficult to model. Fortunately, it has been found that we can calculate D_1 with arbitrary a_1 (or Q_0) by interpolation, where we assume two D_1 values corresponding to two a_1 values (or Q_0 values) are known. In other words, given $a_1 = x$ (or $Q_0 = x$), the distortion, D_1, can be calculated using

$$D_1(x) = D_1(a) + \frac{x-a}{b-a}(D_1(b) - D_1(a)), \tag{15}$$

where a and b are the values of two a_1's (or Q_0's) for two known D_1's.

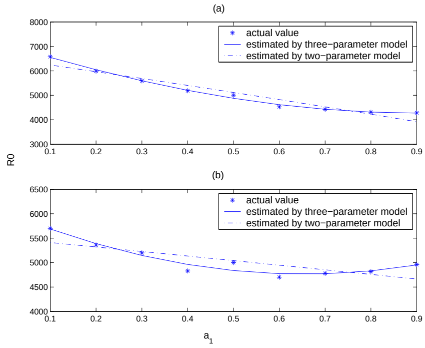

Fig. 1. verification of $R_0 - a_1$ models(equation(8 and 7)): (a) frame 24; (b) frame 12.

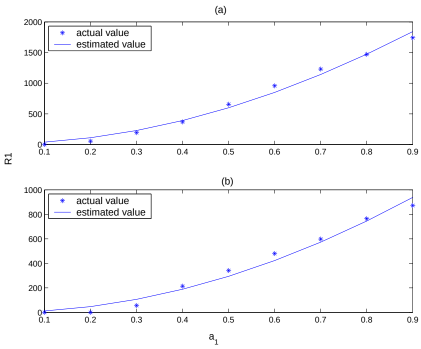

Fig. 2. verification of $R_1 - a_1$ model (equation(10)): (a) frame 24; (b) frame 12.

4. SIMULATION RESULTS

To confirm the accuracy of the models discussed in above section, we conduct simulations using our MDMC codec, which is implemented on top of the public domain H.263+ codec[7]. To apply the above models, for each frame, after motion estimation, the encoder collects necessary statistics by performing motion compensation several times. Based on this data, the encoder then estimates the necessary parameters for the model. The additional complexity is only slightly higher than that in [4].

To verify our $R_0 - a_1$ and $R_1 - a_1$ models, QCIF "Foreman" is encoded at 10 fps using the MDMC codec. We fix $Q_0 = 8$ and $Q_1 = 14$ for all frames and $a_1 = 0.9$ for all but the two test frames, frame 12 and frame 24. For these frames, we vary a_1 and record the bit rates used for coding the Y component. Then, we fit our models to the actual data and use those models to estimate the bit rate of the tested frame. Figure 1 and 2 shows the actual and estimated bit rates in this experiment. In the figures, the actual number reflects the total bits spent on coding the Y component in one frame and the estimated number is calculated using our models. From the two figures shown, we can see those models are quite accurate. We do note that the two-parameter $R_0 - a_1$ model (equation (8)) may not fit the actual data as well as the three-parameter model.

In order to verify the $D - Q$ model, a simulation similar to the previous experiment is conducted, where instead of varying a_1 in the two tested frames, we vary Q_1 from 12 to 30. Figure 3 plots the actual and estimated values of D_1 for the two test frames. From this figure, we can see the $D - Q$ model can work well enough. Here, only $D_1 - Q_1$ verification is plotted. However, the $D - Q$ model works even better for estimating D_0 in our other simulations. The reason is that the assumption of a generalized Gaussian distribution is more realistic for e_0. In our other simulations, $D_1 - a_1$ and $D_1 - Q_0$ relationship (15) and $R_1 - Q_0$ model (9) are also accurate. However, the results are not shown here due to the limited space.

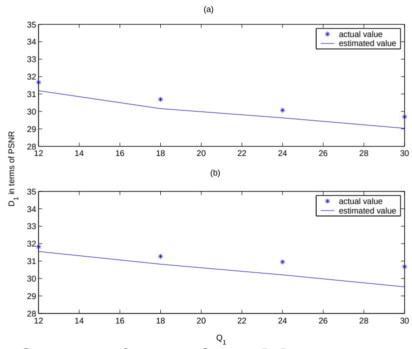

Fig. 3. verification of $D - Q$ model(equations (11, 12, 13 and 14): (a) frame 24; (b) frame 12.

5. CONCLUSION

This paper provides several models to estimate the rates and distortions of MDMC video coding scheme. Using these models together with the decoder distortion model introduced in [3], one can find the optimal parameters directly when coding a video sequence for transmission over error prone channels. The algorithm for dynamically selecting the optimal parameters and the corresponding gains that can be achieved will be reported in a later publication. Preliminary results show that more than 1 dB gains can be achieved compared to a fixed suboptimal parameters selection.

6. REFERENCES

[1] V. K. Goyal, "Multiple Description Coding: Compression Meets the Network ," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74–93, Sep. 2001.

[2] Y. Wang and S. Lin, "Error resilient video coding using multiple description motion compensation," *IEEE Trans. Circuit and System for Video Technology*, vol. 12, no. 6, pp. 438-453, Jun., 2002.

[3] S. Lin and Y. Wang, "Analysis and Improvement of Multiple Description Motion Compensation Video Coding for Lossy Packet Networks," *IEEE Int'l Conf. Image Processing* , Rochester, NY, Sep., 2002.

[4] Z. He and S. K. Mitra, "A unified rate-distortion analysis framework for transform coding," *IEEE Trans. Circuit and System for Video Technology*, vol. 11, no.12, pp. 1221-1237, Dec., 2002.

[5] F. Müller, "Distribution shape of two-dimentional DCT coefficients of natural images," *Electron. Letter*, vol. 29, no.22, pp. 1235-1236, Oct., 2002.

[6] J. Minguillon and J. Pujol, "Uniform quantization error for Laplacian source with applications to JPEG standard," *Mathmatcis of Data/Image Coding, Compressing and Encryption, Proc. of SPIE*, vol. 3456, 1998.

[7] University of British Columbia, "H.263+ codec," previously availabe at ftp://dspftp.ece.ubc.ca.