A ρ-DOMAIN RATE CONTROLLER FOR MULTIPLEXED VIDEO SEQUENCES

G. Valenzise, M. Tagliasacchi, S. Tubaro, L. Piccarreta

Dipartimento di Elettronica e Informazione - Politecnico di Milano, Italy
Email: {valenzise,tagliasa,tubaro,piccarre}@elet.polimi.it

ABSTRACT

This paper addresses the problem of multiplexing pre-encoded video sequences to be transmitted across a bandwidth constrained channel. At each time instant, the available bit budget needs to be optimally allocated to the sequences. We seek for a solution that minimizes the output distortion variance, in order to keep the quality of the reconstructed sequences as similar as possible, by formulating the problem in the ρ-domain. In addition, by enabling a shared video buffer, we show that it is possible to smooth the overall video quality along time. Experimental results on H.264/AVC compressed video data validate the performance of the proposed algorithm.

***Index Terms*—** Rate control, statistical multiplexing

1. INTRODUCTION

In several applications, from digital TV broadcast to video surveillance, there is the need of transmitting simultaneously several video sequences over a bandwidth-limited channel. Therefore, the available bandwidth has to be distributed across the sequences according to some optimality criterion. This problem has been addressed in the literature under the name of statistical multiplexing [1][2][3], where the goal is to maximize the average output quality, or to achieve constant quality. It can be easily demonstrated that simple equal partitioning of the available bandwidth among the sequences (also called programs in this context) is sub-optimal, in the sense that it does not take advantage of the diversity in video complexity.

Most of the literature on statistical multiplexing refers to encoding of MPEG-2 video sequences. In [2] a rate allocation problem is formulated in order to achieve the same quantization parameter for all the sequences, while fulfilling the overall rate constraint. A similar approach, which also includes buffer management issues, is addressed in [1]. These works formulate the problem in the q-domain, by using simple models to relate the quantization step size q to the rate, i.e. $R(q)$. In addition, they make the implicit assumption that by keeping the quantization parameter equal, constant quality is guaranteed. This is in general not true, as it can be demonstrated by adopting more accurate rate-distortion models. For example, in [4] a ρ-domain model is proposed, where ρ indicates the fraction of zero coefficients in the transform domain, after quantization. It can be shown that there is a very accurate linear relationship between the number of nonzero coefficients and the allocated rate. Leveraging the ρ-domain rate-distortion model, in [5] an optimal rate-allocation for MPEG-4 Visual video objects is proposed, with the goal of minimizing the average distortion.

In this paper we consider an application scenario that consists of jointly transcoding/multiplexing H.264/AVC encoded video streams. During decoding, a limited number of parameters is extracted from each video sequence and collected by the joint rate control module. Based on these parameters, the available rate is optimally allocated among the sequences. Rate allocation is performed in the ρ-domain, with the goal of achieving constant quality among the different programs. Our contribution is novel in two aspects: first, we find a computational efficient solution to the minimum variance distortion problem; second, we show that, with typical video sequences, aiming at constant quality among the programs does not penalize the overall quality. In addition, to reduce the frame-to-frame quality variability, we introduce a global video buffer that can compensate the variability of the allocated bandwidth.

Since the focus of this paper is the rate allocation algorithm, we adopt a simple explicit homogeneous transcoding algorithm that decodes the input sequence up to the pixel domain and re-encodes it, thus avoiding drift. The encoder does not perform rate-distortion optimization, but it inherits mode decisions and motion vectors from the incoming stream. Therefore the encoder simply adjusts the quantization parameter at the macroblock level, in order to attain the desired bit budget. Although this is not optimal, it enables fast transcoding of multiple video streams, and the loss in coding efficiency is limited if the output target rate is of the same order of magnitude as the input rate.

The rest of the paper is organized as follows. Section 2 provides a brief overview of the ρ-domain linear model used in the rest of the paper. In Section 3 we describe two rate allocation problems that differ in the optimization goal: minimum average distortion and minimum variance distortion. We show that the second task (minimum variance of distortion among sequences) can be turned into a simpler problem if the exponential model presented in [5] is used to approximate the distortion function in the ρ-domain. The problem of temporally smoothing the distortion is addressed in Section 4, by introducing a global video buffer in the system architecture. Experimental results on H.264/AVC video data are presented in Section 5, while in Section 6 we provide references to open issues and future investigations.

2. OVERVIEW OF THE ρ-DOMAIN MODEL

In [5], it is shown that in any typical transform domain system, there is always a linear relationship between the coding bitrate R and the percentage of zeros among the quantized transform coefficients, denoted by ρ, i.e.:

$$R(\rho) = \theta \cdot (1 - \rho) \quad \text{[bps]} \tag{1}$$

where θ is a constant parameter that depends on the source.

It is widely recognized in the literature that the distribution of DCT coefficients of prediction residuals can be recognized as Laplacian [6]. In [5] the exact expression of the distortion D as a function

---

This work has been partially sponsored by EU under Visnet II Network of Excellence

of ρ is provided for the Laplacian distribution and a mean square error (MSE) distortion metrics. Also, an approximation that is mathematically more tractable is given [5]:

$$D(\rho) = \sigma^2 e^{-\alpha(1-\rho)}, \tag{2}$$

where σ^2 denotes the variance of the DCT coefficients and α is a parameter that can be simply computed at the decoder, as explained in [5]. Equation (2) constitutes the basis of the rate-distortion model used in the rest of this paper.

3. RATE ALLOCATION

3.1. Minimum average distortion

Let $R(t)$ denote the total available bit budget at time t. In this section, we consider a strictly constant bit rate (CBR) channel, therefore $R(t) = R \quad \forall t$. Let $\mathbf{R} = [R_1, R_2, \ldots, R_S]$ denote the rate allocated to each of the S sequences. In order to find the optimal rate allocation that minimizes the average distortion of the output, we need to solve the following non-linear constrained optimization problem:

$$\min_{\mathbf{R}} \frac{1}{S} \sum_{i=1}^{S} D_i(R_i), \quad \text{s.t.} \quad \sum_{i=1}^{S} R_i \leq R. \tag{3}$$

Using equation (2), the minimization problem (3) becomes:

$$\min_{\mathbf{R}} \frac{1}{S} \sum_{i=1}^{S} \sigma_i^2 e^{-\alpha_i(1-\rho_i)}, \quad \text{s.t.} \quad \sum_{i=1}^{S} \theta_i (1-\rho_i) \leq R. \tag{4}$$

This problem can be solved with the Lagrange multipliers method, and we obtain the optimum number of bits for each sequence:

$$R_i = \xi \log \frac{\sigma_i^2}{\xi_i} + \frac{\xi_i}{\sum_{i=1}^{S} \xi_i} \left(R - \sum_{i=1}^{S} \xi_i \log \frac{\sigma_i^2}{\xi_i} \right), \tag{5}$$

where $\xi_i = \theta_i / \alpha_i$.

3.2. Minimum variance distortion

The solution of the problem presented in the previous section minimizes the average distortion, but it does not guarantee that the distortion of the individual sequences is the same. In many applications, the goal is to achieve equal quality instead. This is described by the following optimization problem, expressed in the ρ-domain:

$$\min_{\mathbf{R}} \frac{1}{S} \sum_{i=1}^{S} \left(D_i(\rho_i) - \overline{D} \right)^2, \quad \text{s.t.} \quad \sum_{i=1}^{S} \theta_i (1-\rho_i) \leq R, \tag{6}$$

where $\overline{D} = \frac{1}{S} \sum_{i=1}^{S} D_i(\rho_i)$. This problem is difficult to solve in closed form, since $\overline{D}$ depends on the whole set of distortion values D_i of each sequence. To overcome this limitation, we first reformulate problem (6) into a simpler one in order to achieve equal distortion for all the sequences. Then, we evaluate the "goodness" of the obtained distortion value against the minimum average distortion solution.

3.2.1. *Equal distortion bit allocation*

The first problem we deal with is how to allocate the available rate to each video sequence so that the variance of the output distortions is minimized, subject to rate constraint. Hereafter, we assume that the rate-distortion profile is well approximated by the exponential distortion model (2).

Let $\{\tilde{\mathbf{D}}^{(n)}\}$, $n = 1, 2, 3, \ldots$ ($\tilde{\mathbf{D}}^{(n)} = [\tilde{D}_1^{(n)}, \ldots, \tilde{D}_S^{(n)}]^T$) be the sequence of distortions found by solving the sequence of minimization problems $\mathcal{P}^{(1)}, \mathcal{P}^{(2)}, \ldots \mathcal{P}^{(n)}$. The generic problem $\mathcal{P}^{(n)}$ is defined below. The superscript $^{(n)}$ indicates a specific instance of the problem, parametrized by the value of n:

$$\begin{aligned} \mathcal{P}^{(n)}: \quad & \min_{\mathbf{R}} \sum_{i=1}^{S} D_i^n(\rho_i) = \min_{\mathbf{R}} \sum_{i=1}^{S} \sigma_i^{2n} e^{-n\alpha_i(1-\rho_i)}, \\ & \text{s.t.} \quad \sum_{i=1}^{S} \theta_i(1-\rho_i) \leq R. \end{aligned} \tag{7}$$

It is possible to prove the following property (the proof is omitted due to space constraints):

Property 1 *The sequence* $\{\text{var}(\tilde{\mathbf{D}}^{(n)})\}$ *converges to 0 as* $n \to \infty$.

Therefore, it is possible to solve problem (7) for $n \to \infty$ instead of problem (6) to attain minimum variance distortion. We emphasize that the solution is exactly the same as that of (6), not an approximation of it. In fact, we can compute the limit of the sequence of rates $\tilde{R}_i^{(1)}, \tilde{R}_i^{(2)}, \ldots$ which minimize the problems $\mathcal{P}^{(1)}, \mathcal{P}^{(2)}, \ldots \mathcal{P}^{(n)}, \ldots$ as follows:

$$\tilde{R}_i = \xi_i \log \sigma_i^2 + \frac{\xi_i R}{\sum_{i=1}^{S} \xi_i} - \frac{\xi_i \sum_{i=1}^{S} \xi_i \log \sigma_i^2}{\sum_{i=1}^{S} \xi_i}. \tag{8}$$

This result allows us to allocate very efficiently the bit budgets R_i so that all the video programs have the same distortion level

$$\tilde{D} = \exp \left[\frac{\sum_{i=1}^{S} \xi_i \log \sigma_i^2 - R}{\sum_{i=1}^{S} \xi_i} \right] \tag{9}$$

3.2.2. *Evaluation against minimum average distortion*

Once the minimum variance solution $\tilde{D}$ has been found, we investigate how much this value is close to the solution of (3), i.e. the solution of the minimum average distortion problem (3) or, equivalently, the solution of (7) when $n = 1$. Since the average distortion function with the exponential model (2) is a sum of convex functions (thus it is convex too), there exists only one global minimum, D^*, which can be found to be equal to [5]:

$$D^* = \frac{1}{S} \sum_{i=1}^{S} \exp \left[\log \xi_i + \frac{\sum_{i=1}^{S} \xi_i \log \left(\frac{\sigma_i^2}{\xi_i} \right) - R}{\sum_{i=1}^{S} \xi_i} \right] \tag{10}$$

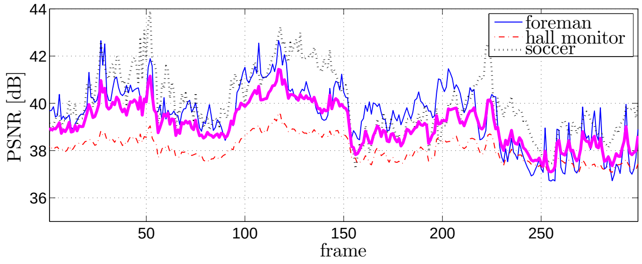
To evaluate the performance of the minimum variance distortion, we analyze the difference:

$$\tilde{D} - D^* = \tilde{D} \cdot [1 - \mathcal{E}], \tag{11}$$

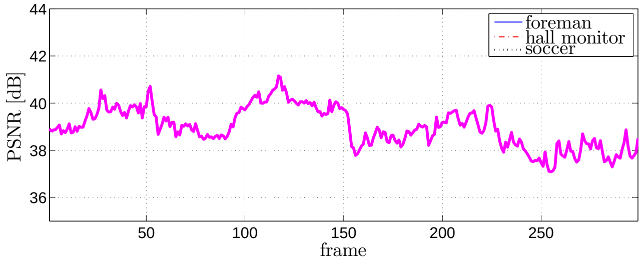
where

$$\mathcal{E} = \frac{1}{S} \exp \left(\frac{-\sum_{i=1}^{S} \xi_i \log \xi_i}{\sum_{i=1}^{S} \xi_i} \right) \sum_{i=1}^{S} \xi_i \tag{12}$$

is a loss factor which should be one to guarantee that $\tilde{D}$ is equal to D^*. The following property fixes the bounds on $\tilde{D}$:



(a) Minimum average distortion.



(b) Minimum distortion variance.

Fig. 1. Distortion profiles using: a) minimum average distortion; b) minimum variance distortion optimization, using the exponential model. CBR, $R = 1.2$ bps.

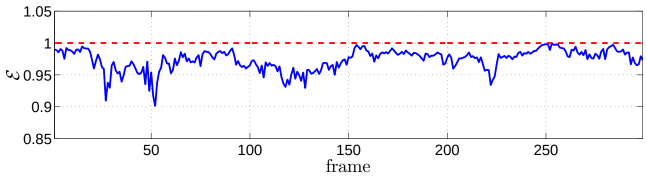


Fig. 2. Quality loss factor $\mathcal{E}$.

Property 2

$$D^* \leq \tilde{D} \leq S \cdot D^*, \tag{13}$$

In fact, let $\zeta_i = \frac{\xi_i}{\sum_{i=1}^{S} \xi_i}$, $0 \leq \zeta_i \leq 1$, be the normalized values of ξ_i. The loss factor $\mathcal{E}$ can be rewritten as:

$$\mathcal{E} = \frac{1}{S} e^{H(\zeta)}, \tag{14}$$

where we define $H(\zeta) = -\sum_{i=1}^{S} \zeta_i \log \zeta_i$ as the entropy function of a source having the set ζ_i, $i = 1 \ldots S$ as probability mass function of its symbols. From information theory, we know that

$$0 \leq H(\zeta) \leq \log(S). \tag{15}$$

Therefore $\mathcal{E}$ is bounded as $\frac{1}{S} \leq \mathcal{E} \leq 1$, and property 2 holds.

In proving Property 2, we have found that the performance loss in terms of average MSE distortion, due to the constraint that all video sequences must have the same visual quality, depends on how the video programs differ to each other. More precisely, the connection between the characteristics of the video sequences and the quality loss is given by the empirical entropy of the normalized parameters ζ_i: when the ζ_i are the same for all sequences, then $\mathcal{E} = 1$ and $\tilde{D}$ reaches the lower bound D^*.

Figure 1(a) and Figure 1(b) show the solutions in terms of both rate and distortion of problem (3) and (7) respectively. The model parameters have been estimated from real video sequences encoded using H.264/AVC at fixed $QP = 20$. From Figure 1(b) we notice that, at least in the model-based simulations, all sequences attain the same distortion at each time instant. These results also show that the minimum variance distortion $\tilde{D}$ is very close to minimum average distortion D^*, and the loss factor $\mathcal{E}$ is typically greater than 0.90 (i.e. less than 0.45 dB), as shown in Figure 2.

4. TEMPORAL SMOOTHING

To obtain a visually pleasing video presentation, not only does each video frame of each sequence need to be encoded at the optimal quality level, but also the frame-to-frame perceptual quality changes need to be smooth enough so that temporal artifacts are minimized. Note that this task conflicts with the CBR channel requirements, since smooth quality change from frame to frame gives rise to large bit rate fluctuations, which inexorably infringe the total rate constraint. In order to introduce quality smoothing in our minimum distortion variance transcoder, we need therefore to add an *encoder buffer* into the system. In this paper we limit to describe the case of a global encoder buffer, shared by all bitstreams.

In [7] it is proved that using a geometric averaging filter, it is possible to smooth the optimal minimum distortion profile while achieving the target bit rate on average. Let $\tilde{D}_{\text{CBR}}(t)$ be the minimum variance CBR distortion at frame t; we define the smoothed distortion target at time t as

$$D_S(t) = \prod_{k=0}^{M-1} \left[\tilde{D}_{\text{CBR}}(t-k)\right]^{\frac{1}{M}}, \tag{16}$$

where M is the length of the averaging window (e.g. $M = 15$ frames). To maintain a temporally-smooth average distortion, we need: 1) to compute the CBR distortion profile; 2) to smooth it using (16); 3) to set $D_S(t)$ as target distortion and find the rates $R_i(t)$ which meet $D_S(t)$ in a minimum distortion variance sense, for each frame t. Quality smoothing requires therefore that we relax or tighten the rate constraint according to the current buffer level. Let B_{max} be the size of the buffer (which conditions the latency experienced at the decoder); B_0 is the desired buffer level (e.g. $B_0 = 0.5 \cdot B_{\text{max}}$); $b(t)$ denotes instead the buffer fullness at time t; finally let C be the channel rate, i.e. the rate at which the buffer is drained. In our scenario $C = R$. The buffer state evolves according to the difference equation:

$$b(t) = b(t-1) + \sum_{i=1}^{S} R_i(t) - C. \tag{17}$$

The key idea of the smoothing algorithm is to relax the rate constraint when the buffer level is under the target B_0, so that the smoothed distortion profile can be tracked by the rate control algorithm. If the buffer fills up over the desired level B_0, then the rate constraint is re-enabled in such a way that the buffer level is reset to the desired target.

If the buffer level $b(t)$ is lower than B_0, we relax the rate constraint and allocate the bit budget according to the following straightforward minimization problem:

$$\min_{\mathbf{R}} \sum_{i=1}^{S} \left(D_i(t) - D_S(t)\right)^2. \tag{18}$$

The rates $R_i(t)$ for each sequence are then:

$$R_i(t) = \xi_i(t) \left(\log \sigma_i^2(t) - \log D_S(t)\right). \tag{19}$$

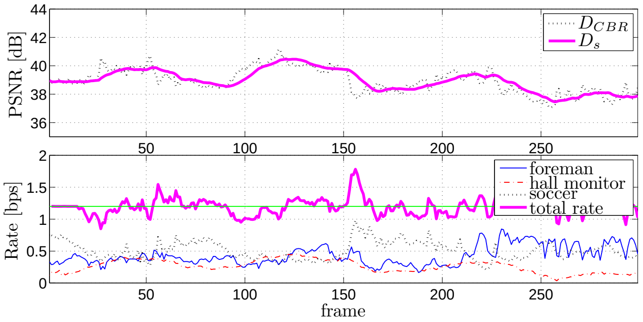


Fig. 3. Quality smoothing with 3 CIF video sequences, using the exponential distortion model.

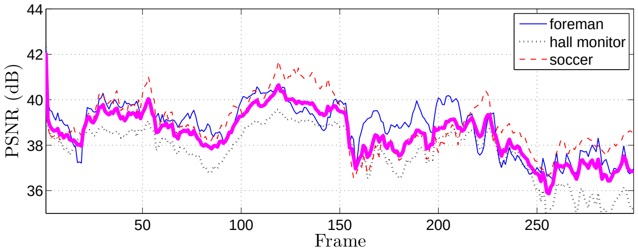
When the number of bits in the buffer exceeds B_0, the target bit rate R of the CBR distortion profile is reduced to prevent buffer overflow. Let $B_{\text{res}} = b(t) - B_0$; if $B_{\text{res}} > 0$, the encoder needs to reduce the output bits by B_{res} within the next K frames (e.g. K can be set to $0.5M$, as suggested in [7]). Therefore, the new CBR target becomes:

$$R' = R - \frac{B_{\text{res}}}{K}. \tag{20}$$

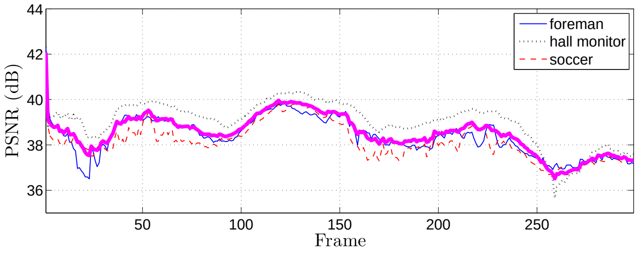
The value of distortion $\tilde{D}_{\text{CBR}}(t)$ is smoothed with (16), and the target distortion $D_S(t)$ is used to find the rates for each sequence with (19). The results of model simulations for quality smoothing is shown in Figure 3. Note that, on average, the total bit rate for the three sequences is equal to the bit rate of the CBR problem (in this example, 1.2 bits per pixel).

5. EXPERIMENTAL RESULTS

We have tested the smoothed minimum distortion variance algorithm against the minimum average distortion bit allocation on *Foreman*, *Hall monitor* and *Soccer* CIF sequences encoded with H.264/AVC baseline profile at 30 fps with a fixed $QP = 20$. The first frame was encoded in INTRA mode, while the rest of the sequence is composed by P slices only. The test procedure consists of extracting the parameters $\{\alpha_i, \theta_i, \sigma_i^2\}$ from the incoming bitstreams. As explained in [5], these values can be readily obtained form the histogram of the DCT coefficients by fitting a Laplacian distribution model. Rate allocation is then performed at each time instant according to the algorithms described in Section 3 (for minimum average distortion and minimum variance distortion) and in Section 4 (for the temporal smoothing case). The resulting sets of rates $[R_1(t), R_2(t), R_3(t)]^T$ are used in input to the encoder part of the transcoder. A rate allocation algorithm is adopted at the macroblock level by adaptively adjusting the quantization parameter QP in order to match the desired target rate $R_i(t)$. Due space constraints, the actual rate allocation algorithm used in our simulations is not described here in detail. Nevertheless, any rate control algorithm, e.g. the one described in [4] can be used for this purpose. The results are presented in Figure 4. The mean standard deviation in PSNR between the three programs in the case of minimum average distortion is 0.85 dB, while it is 0.51 dB in the temporal smooth case: thus, not only the visual quality has been smoothed across time (as it can be clearly noticed from the graph), but also the quality variance between the video sequences has been significantly reduced.



(a) Minimum average distortion.



(b) Minimum distortion variance plus smoothing.

Fig. 4. Minimum average distortion vs. smoothed minimum distortion variance optimization applied to the transcoding of 3 H.264/AVC video sequences.

6. CONCLUSIONS

In this paper we have shown that, using an exponential rate-distortion model in the ρ-domain, it is possible to encode different video sequences minimizing the inter-sequence quality variance. Adding a global buffer at the encoder allows to smooth the overall distortion from frame to frame. Future work will investigate the use of an encoder buffer for each sequence, in order to smooth the distortion profiles of each program. We also want to characterize the quality loss due to minimization of variance from a statistical point of view.

7. REFERENCES

[1] L. Boroczky, AY Ngai, and EF Westermann, "Statistical multiplexing using MPEG-2 video encoders," *IBM Journal of Research Development*, vol. 43, no. 4, pp. 511–520, 1999.

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

[3] M. Balakrishnan and R. Cohen, "Global optimization of multiplexed video encoders," in *Proceedings of the 1997 International Conference on Image Processing (ICIP'97)*, 1997, vol. 1.

[4] Z. He and SK Mitra, "A linear source model and a unified rate control algorithm for DCT video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 11, pp. 970–982, 2002.

[5] Z. He and SK Mitra, "Optimum bit allocation and accurate rate control for video coding via ρ-domain source modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 10, pp. 840–849, 2002.

[6] EY Lam and JW Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Transactions on Image Processing*, vol. 9, no. 10, pp. 1661–1666, 2000.

[7] Z. He, W. Zeng, and C.W. Chen, "Low-pass filtering of rate-distortion functions for quality smoothing in real-time video communication," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 8, pp. 973–981, 2005.