MINIMUM VARIANCE MULTIPLEXING OF MULTIMEDIA OBJECTS

G. Valenzise, M. Tagliasacchi, S. Tubaro

Dipartimento di Elettronica e Informazione
Politecnico di Milano
P.zza Leonardo da Vinci, 32 20133 - Milano, Italy
Email: {valenzise, tagliasa, tubaro}@elet.polimi.it

ABSTRACT

This paper addresses the problem of simultaneous transmission of multiple multimedia objects (such as images or video sequences) over a bandwidth-limited channel. The trivial strategy of partitioning in equal parts the available rate among the bitstreams is suboptimal, when the multimedia objects have different coding complexities. Exploiting object diversity allows us to allocate the bandwidth according to some optimality criteria, e.g. minimizing the average total distortion or minimizing the variance between the distortions of each object. By describing the rate-distortion characteristics of each multimedia object in terms of a simple exponential model, we provide a closed form solution for both the minimum average and the minimum variance problems. In addition, if we consider the statistical distribution of the rate-distortion model parameters, we can show that the minimum variance solution can effectively reduce the quality fluctuations among the objects, with an overall coding efficiency loss, w.r.t. the minimum average solution, of only 0.5dB on average. Some experiments, carried out on different H.264/AVC video sequences, validate our theoretical results.

***Index Terms*—** Multimedia, coding, statistical multiplexing, rate control

1. INTRODUCTION

The simultaneous transmission of multiple multimedia data streams on the same bandwidth-limited channel is a very common problem in many practical applications, ranging from broadcast television to video and signal-based surveillance systems. A common task in these applications is how to allocate the available bandwidth to each multimedia object in a somehow optimal way. A very simple solution to the problem is to divide the bandwidth among the multimedia objects so that each bitstream receives an equal portion of the bit rate. It is straightforward to see that this method is optimal only in the case that all the multiplexed objects share the same rate-distortion characteristics. In practice this is rarely the case and more sophisticated bit allocation techniques have to be considered.

The problem of distributing the available bandwidth across the objects has been addressed in the literature under the name of *statistical multiplexing* [1]. The term was first used to indicate the joint transmission of multiple objects on the same channel: due to the different coding complexities of each multimedia object, an approximately constant bit rate channel results from multiplexing several data sequences. The diversity of the multiplexed multimedia objects can be further used to allocate the bit rate according to some optimality criteria, e.g. to minimize the average object distortion (MINAVE) or the variance of the object distortions (MINVAR) under some rate constraint. Previous works in the literature have concentrated in particular on the first task [2][3], while for the second there are some specific works [4][5] which deal with the minimization of quality fluctuations of the objects along time, for the case of video sequences. In [6], a joint rate control for multiple H.264/AVC video sequences is proposed, which uses a look-ahead processing window to allocate the bandwidth resources in order to reduce quality variations along time. A different technique which yields similar results is presented in [7]. In both these two works, it is shown that by reducing the quality variations along time, also the differences in distortion between multimedia objects are kept small.

In a previous work [8], the authors have shown that it is possible to achieve the same distortion for the multiplexed objects in an efficient way, for the case of video sequences, using the ρ-domain model proposed in [9]. In this paper, we extend the solution to the MINVAR problem for multimedia objects whose rate-distortion characteristics can be described by a simple exponential model. The goal is to achieve constant quality among the different objects meeting a total rate constraint. Our contribution is novel in the following aspects: first, we find a closed form solution to the MINVAR problem, which holds for those multimedia data whose rate-distortion characteristics can be approximated by the exponential model. Second, we provide a statistical analysis of the performance of the MINVAR solution w.r.t. the MINAVE optimization: we show that, even if the average distortion attained by MINAVE solution is better, the MINVAR distortion in terms of PSNR is only 0.5 dB worse on average. Although our method applies to a broad range of multimedia objects, as a proof of concept we illustrate the results multiplexing some H.264/AVC video sequences in Section 5. The rest of the paper is organized as follows: Section 2 describes and justifies the exponential rate-distortion model adopted in the rest of the paper; Section 3 presents the MINAVE and MINVAR closed form solutions; Section 4 compares the distortion obtained solving the MINVAR problem with the minimum average distortion; Section 5 illustrates an example where objects are video sequences; finally, Section 6 draws the conclusions.

2. RATE-DISTORTION MODEL

We consider here multimedia objects whose rate R and distortion D are related by the following exponential model:

$$D(R) = \sigma^2 e^{-\frac{R}{\beta}}, \tag{1}$$

where σ^2 is the variance of the object data and β is a parameter related to the coding complexity of the multimedia data. Given two objects with the same variance σ^2, we need to spend more bits to encode the object with the higher value of β in order to attain the same distortion between the bitstreams. Model (1) is linear in the $R - \log D$ plane, and the parameters β and $\log \sigma^2$ are, respectively, the inverse of the slope and the intercept at zero. In [10], it is considered the case of average rate-distortion functions with the same form of (1); however, all objects are supposed to have the same coding complexity β and different variances σ_i^2, i.e. all the data sequences are represented by parallel lines in the $R - \log D$ plane. In that case it is shown that the coding gain due to a minimum average optimization w.r.t. a constant bit allocation is the ratio between the arithmetic and geometric mean of the σ_i^2. Examples of exponential $D(R)$ curves can be found when a high-resolution uniform quantizer is used and the distortion metrics is the mean square error. This means that, at high rates, the rate-distortion characteristics can be well approximated by (1). As an example, Figure 1(a) shows the rate-distortion function of a QCIF frame coded using JPEG, JPEG 2000 and H.264/AVC intra: each curve exhibits a linear behavior in the $R - \log D$ plane at medium to high rates. If we consider objects coded using the same algorithm, as shown in Figure 1(b) for the case of three CIF video sequences encoded with H.264/AVC, it is apparent that each $D(R)$ curve is very well approximated by an exponential function at high rates [11]; therefore, each video sequence characteristics can be summarized by its parameters β_i and σ_i^2.

3. OPTIMAL BIT ALLOCATION

In this section, we formulate and solve the MINAVE and MINVAR rate allocation problems among the multiplexed objects. In the next section, the average distortions obtained in the two cases are compared.

3.1. Minimum Average distortion – MINAVE

In order to find the optimal rate allocation that minimizes the average distortion of the N multiplexed objects, we need to solve the following non-linear constrained optimization problem:

$$\min_{\mathbf{R}} \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 e^{-\frac{R_i}{\beta_i}}, \quad \text{s.t.} \quad \sum_{i=1}^{N} R_i \leq R_T, \tag{2}$$

where R_T is the total available rate. Since each term of the summation is decoupled from the others, we can easily solve this problem by means of the Lagrange multiplier method. We thus obtain that (2) is minimized by the rates:

$$R_i^* = \beta_i \log(\sigma_i^2/\beta_i) + \frac{\beta_i}{\beta_0} \left(R_T - \sum_{j=1}^{N} \beta_j \log\left(\sigma_j^2/\beta_j\right) \right), \tag{3}$$

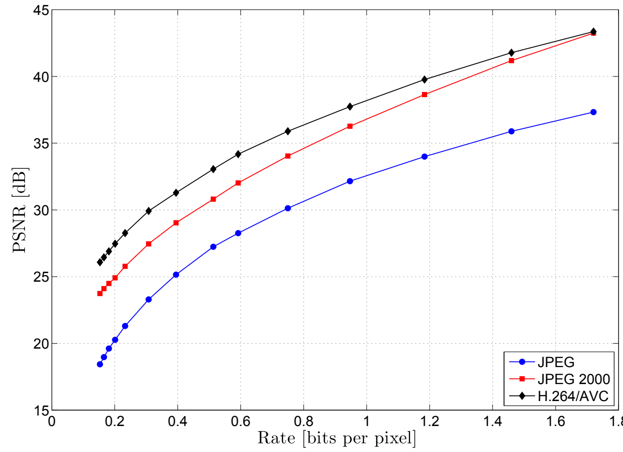
where $\beta_0 = \sum_{i=1}^{N} \beta_i$, and the value of the objective function evaluated at R_i^* is

$$D_{\text{MINAVE}} = \frac{1}{N} \sum_{i=1}^{N} \beta_i \exp\left(\frac{\sum_{j=1}^{N} \beta_j \log(\sigma_j^2/\beta_j) - R_T}{\beta_0} \right). \tag{4}$$

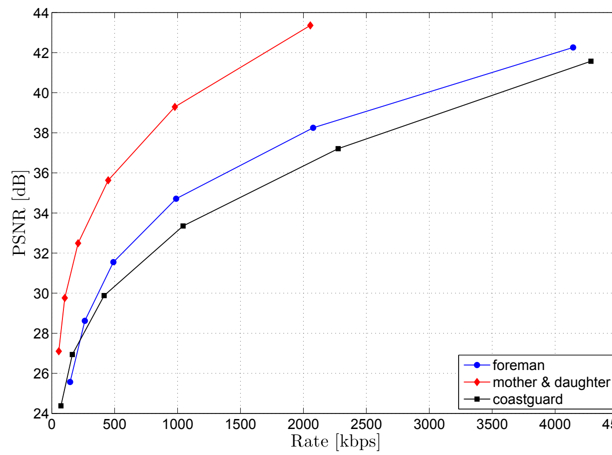
3.2. Minimum Distortion Variance – MINVAR

We now want to minimize the variance of the output distortions, i.e.

$$\min_{\mathbf{R}} \frac{1}{N} \sum_{i=1}^{N} \left(D_i(R_i) - \overline{D} \right)^2, \quad \text{s.t.} \quad \sum_{i=1}^{N} R_i \leq R_T, \tag{5}$$



(a) $D(R)$ characteristics for different types of multimedia contents



(b) $D(R)$ characteristics of different H.264/AVC inter-coded CIF video sequences

Fig. 1. Exponential behavior of $D(R)$ at high rates

where $\overline{D} = \frac{1}{N} \sum_{j=1}^{N} D_j(R_j)$. This problem is no longer decoupled, since each term of the summation, D_i, depends on all the others distortion terms D_j, $j \neq i$ through the average distortion $\overline{D}$. Therefore, we are not able to find a closed form solution using the Lagrange multiplier method. However, we have shown in a previous work [8], for the case of video sequences, that problem (5) is equivalent to the following, for $n \to \infty$:

$$\min_{\mathbf{R}} \frac{1}{N} \sum_{i=1}^{N} D_i^n(R_i), \quad \text{s.t.} \quad \sum_{i=1}^{N} R_i \leq R_T. \tag{6}$$

This latter problem can be solved in closed form to obtain the optimal (in the minimum-variance sense) rates

$$\tilde{R}_i = \beta_i \log \sigma_i^2 + \frac{\beta_i}{\beta_0} \left(R_T - \sum_{j=1}^{N} \beta_j \log \sigma_j^2 \right), \tag{7}$$

which yields the same distortion D_{MINVAR} for each multimedia object:

$$D_{\text{MINVAR}} = \exp\left(\frac{\sum_{i=1}^{N}\beta_i \cdot \log\sigma_i^2 - R_T}{\beta_0}\right). \quad (8)$$

4. PERFORMANCE ANALYSIS

In this section we compare the average distortions obtained by the MINAVE and the MINVAR bit allocation strategies. Since the average distortion is a convex function (it is a sum of exponentials), there exists only a global minimum, which is given by (4), i.e. by the MINAVE solution. Therefore, by allocating the rate to each multiplexed multimedia object according to the MINVAR optimization, we expect that the average distortion D_{MINVAR} is higher than D_{MINAVE}. To quantify this quality deterioration, we consider the *coding efficiency loss* ratio $D_{\text{MINVAR}}/D_{\text{MINAVE}}$. We have already found in our previous work [8] that this ratio can be rewritten as

$$\frac{D_{\text{MINVAR}}}{D_{\text{MINAVE}}} = \frac{e^{H(\boldsymbol{\zeta})}}{N}, \quad (9)$$

where we define $\boldsymbol{\zeta} = [\zeta_1, \zeta_2, \ldots \zeta_N]^T$, $\zeta_i = \frac{\beta_i}{\beta_0}$ and $H(\boldsymbol{\zeta}) = -\sum_{i=1}^{N}\zeta_i \log\zeta_i$, i.e. $H(\boldsymbol{\zeta})$ is the entropy function of a discrete memoryless source having the set ζ_i, $i = 1\ldots N$ as the probability mass function of its N symbols. From basic information theory, it follows that:

$$\frac{1}{N} \leq \frac{D_{\text{MINVAR}}}{D_{\text{MINAVE}}} \leq 1. \quad (10)$$

In words, equation (11) states that in the best case scenario the two distortions (MINAVE and MINVAR) are the same; in the worst case, the average distortion incurred by solving the MINVAR problem is N times larger than the global optimum, where N is the number of multiplexed objects.

The lower bound in (11) is not very optimistic, since it says that the coding efficiency loss due to MINVAR allocation maybe be very large w.r.t. the MINAVE solution. Fortunately, it was empirically shown in [8] that this quality loss is much smaller in practice. We extend here that analysis using a more formal, statistical framework. We are ultimately interested in finding a bound on the expectation of the coding efficiency loss. By the Jensen's inequality, we can write:

$$E\left[\frac{D_{\text{MINVAR}}}{D_{\text{MINAVE}}}\right] = \frac{1}{N}E\left[e^{H(\boldsymbol{\zeta})}\right] \geq \frac{1}{N}e^{E[H(\boldsymbol{\zeta})]}. \quad (11)$$

In order to find $E[H(\boldsymbol{\zeta})]$, we need to define an appropriate statistical model describing the distribution of β_i. Since the parameters β_i are all positive, their distribution can be well approximated by a gamma distribution:

$$\text{Gamma}\,(\beta_i; a, b) = \beta_i^{a-1}\frac{b^a e^{-b\beta_i}}{\Gamma(a)}, \quad (12)$$

where a is the shape parameter and b is the inverse of the scale parameter, whereas $\Gamma(\cdot)$ denotes the gamma function. Some experiments with video sequences have confirmed that the fitting with the gamma distribution approximates very well the histograms of β_i. It can be shown [12] that, if β_i is gamma distributed, then $\zeta_i = \frac{\beta_i}{\beta_0}$ follows a multivariate Dirichlet distribution:

$$p(\boldsymbol{\zeta}) = \text{Dir}(\boldsymbol{\zeta}; a_1, a_2, \ldots, a_N) = \frac{1}{B(\mathbf{a})}\prod_{i=1}^{N}\zeta_i^{a_i-1} \quad (13)$$

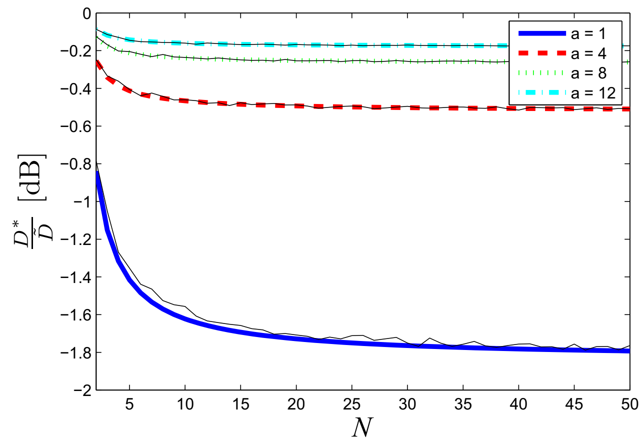


Fig. 2. Coding efficiency loss for different values of a. Thick lines are the lower bounds of (11); thin lines are the result of a Montecarlo simulation, obtained by sampling the Dirichlet distribution.

where $a_i > 0$, and the normalizing constant is the multinomial beta function:

$$B(\mathbf{a}) = B(a_1, a_2, \ldots, a_N) = \frac{\prod_{i=1}^{N}\Gamma(a_i)}{\Gamma(\sum_{i=1}^{N}a_i)} \quad (14)$$

We notice that the Dirichlet distribution is defined over the N-dimensional simplex $\mathcal{S}_N$ given by the constraints $\zeta_i \geq 0$, $\sum_{i=1}^{N}\zeta_i = 1$. Given the distribution of $\boldsymbol{\zeta}$, we can now compute the expectation of the entropy $H(\boldsymbol{\zeta})$ by solving the following integral over the simplex surface:

$$E[H(\boldsymbol{\zeta})] = \int_{\mathcal{S}_N} H(\boldsymbol{\zeta})p(\boldsymbol{\zeta})ds_N. \quad (15)$$

By changing the coordinates to $\eta_i = \sqrt{\zeta_i}$, after some laborious calculations one obtains:

$$E[H(\boldsymbol{\zeta})] = \psi(a_0 + 1) - \frac{1}{a_0}\sum_{i=1}^{S}a_i\psi(a_i + 1), \quad (16)$$

where $a_0 = \sum_{i=1}^{S}a_i$, and ψ is the *digamma* function $\psi(t) = \frac{d}{dt}\log\Gamma(t)$.

By substituting (16) into (11), we can obtain the expected coding efficiency loss we incur when we carry out a MINVAR minimization in place of a minimum average optimization. To better illustrate these concepts, we quantify the coding efficiency loss in a simple case. If we observe a long sequence of multiplexed objects (e.g. frames of a long video sequence), and we do not make prior assumptions about the distribution of the individual objects, we can assume that β_i are i.i.d., i.e. $\beta_i \sim \text{Gamma}(\beta_i; a, b)$. When all the a_i are the same ($a_i = a$, $\forall i$), equation (16) becomes:

$$E[H(\boldsymbol{\zeta}); a_i = a] = \psi(Na + 1) - \psi(a + 1). \quad (17)$$

This corresponds to the expected entropy of a Dirichlet distribution having its mean value at the barycenter of the simplex. The parameter a controls the peakedness of the distribution about the mean: higher values of a result in sharper peaks about the mean, i.e. the variance of the distributions decreases as a grows. Figure 2 shows the coding efficiency loss as the number of multiplexed objects N increases, for different values of the parameter a. It can be seen that

Sequences	MINAVE variance	MINVAR variance	ΔPSNR
F-H	5.80	0.77	-0.12
F-S	10.48	3.58	-0.29
F-C	27.18	3.09	-0.25
H-S	4.71	1.09	-0.13
H-C	51.49	7.45	-0.08
S-C	68.81	2.59	-0.04
F-H-S	6.44	0.89	-0.15
F-H-C	23.60	3.20	-0.12
H-S-C	27.61	2.19	-0.09
F-S-C	36.22	7.33	-0.05
F-S-C-S	57.50	9.71	-0.02

Table 1. MINAVE vs. MINVAR optimization. The average rate for each sequence is $1/2$ bpp. F = *Foreman*, H = *Hall Monitor*, S = *Soccer*, C = *Coastguard*

as the number of multiplexed objects gets larger, the gap between MINVAR and MINAVE average distortions becomes larger, even if it can be shown that the curves pictured in the figure are asymptotically bounded. Some more interesting considerations can be drawn by looking at the quality loss for different parameters of the Dirichlet distribution. When the variance of the distribution goes to zero, the pdf of $\boldsymbol{\zeta}$ becomes a Dirac delta centered at $\boldsymbol{\zeta} = [1/N, \ldots 1/N]$. In this case, $H(\boldsymbol{\zeta}) = \log N$ and, by (9), we have that the coding efficiency loss is null; this corresponds to the case in which all the objects have the same coding complexity β. This latter result is in agreement with information theory, which states that, for bitstreams that have a $D(R)$ characteristics with the same slope (at high rates), the optimal rate allocation in a MINAVE sense is the one that achieves equal distortion for all the multiplexed objects [13].

For common multimedia objects, such as images or video sequences, the shape parameter a is typically larger than 3-4. Therefore, on average, the expected quality loss incurred when achieving the same distortion between objects is equal to or less than 0.5 dB.

5. EXPERIMENTAL RESULTS WITH H.264/AVC VIDEO SEQUENCES

As a proof of concept, we have applied the MINAVE and MINVAR allocation strategies to the problem of multiplexing some H.264/AVC video sequences on a bandwidth limited channel. We have used input sequences encoded at a fixed QP $= 20$; from the test sequences, we have extracted the parameters of the rate-distortion model for each frame; then, we have used the σ_i^2 and β_i to optimally allocate, for each frame, the available rate to the different objects, according to the MINAVE or MINVAR criteria.

Table 1 shows the results obtained multiplexing the three CIF video sequences *Foreman*, *Hall monitor*, and *Coastguard*; the available rate R_T depends on the number of multiplexed sequences: for this example, $R_T = N \cdot 0.5$ bpp. The second and third columns of the table show the average variance, computed for 300 frames, across the distortions of the video sequences. Differently from what stated in Section 3.2, the variance of the MINVAR distortions is greater than zero, due to the non-perfect linearity of the $D(R)$ characteristics in the $R - \log D$ plane, even at high rates. It can be clearly seen, however, that using the MINVAR optimization the inter-object variance is noticeably reduced compared to the one obtained through MINAVE allocation. The last column of the table shows the coding efficiency loss in terms of the average PSNR for the multiplexed objects. The largest coding efficiency loss is 0.29 dB.

6. CONCLUSIONS

In this paper we have considered the problem of multiplexing different multimedia objects over a bandwidth-limited channel. We have considered two kinds of optimal bit allocation strategies: the minimization of the average object distortion and the minimization of the variance of inter-object distortions. Adopting an exponential rate-distortion characteristics, which has been classically used in literature to model data sequences quantized with a fine granularity, we have provided a closed solution for both the MINAVE and the MINVAR problems. We have then investigated how much the MINVAR allocation deteriorates the average object distortion w.r.t. the optimal MINAVE solution, by providing a statistical analysis of the *expected* coding efficiency loss. Using an example with some multiplexed video sequences, we have proved that our method effectively reduces the inter-object distortion variance, with just a negligible loss in term of the overall object quality.

7. REFERENCES

[1] L. Boroczky, A. Y. Ngai, and E. F. Westermann, "Statistical multiplexing using MPEG-2 video encoders," *IBM Journal of Research Development*, vol. 43, pp. 511–520, July 1999.

[2] Z. He and S. K. Mitra, "Optimum bit allocation and accurate rate control for video coding via ρ-domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 840–849, November 2002.

[3] A. Vincent, P. Corriveau, P. Blanchfield, and R. Renaud, "Modeling of the coding gain of joint coding for multi-program video transmission," in *Proc. IEEE Int. Conf. on Multimedia and Expo (ICME)*, 2000, pp. 1309–1312.

[4] Z. Chen and K. Ngan, "Distortion variation minimization in real-time video coding," *Signal Processing: Image Communication*, vol. 21, pp. 273279, January 2006.

[5] L. J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 446–459, 1998.

[6] Z. He and D. Wu, "Look-ahead processing and joint rate control for multiple JVT video encoders," in *Proc. IEEE Int. Packet Video Workshop*, December 2004.

[7] J. Yang, X. Fang, and H. Xiong, "A joint rate control scheme for H. 264 encoding of multiple video sequences," *IEEE Trans. Consum. Electron.*, vol. 51, no. 2, pp. 617–623, 2005.

[8] G. Valenzise, M. Tagliasacchi, and S. Tubaro, "A smoothed, minimum distortion-variance rate control algorithm for multiplexed transcoded video sequences," in *ACM Int. Workshop on Mobile Video,* MV'07, Augsburg, Germany, September 2007.

[9] Z. He and S. K. Mitra, "A liner source model and a unified rate control algorithm for DCT video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 970–982, November 2002.

[10] J. Choi and D. Park, "A stable feedback control of the buffer state using the controlledLagrange multiplier method," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 546–558, 1994.

[11] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G.J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 688–703, 2003.

[12] L. Devroye, *Non-Uniform Random Variate Generation*, Springer-Verlag, New York, 1986.

[13] Thomas M. Cover and Joy A. Thomas, *Information Theory*, Wiley Interscience, John Wiley & Sons, Inc., 1991.