Minimum Variance Optimal Rate Allocation for Multiplexed H.264/AVC Bitstreams

Marco Tagliasacchi, *Member, IEEE*, Giuseppe Valenzise, *Student Member, IEEE*, and Stefano Tubaro, *Member, IEEE*

***Abstract*—Consider the problem of transmitting multiple video streams to fulfill a constant bandwidth constraint. The available bit budget needs to be distributed across the sequences in order to meet some optimality criteria. For example, one might want to minimize the average distortion or, alternatively, minimize the distortion variance, in order to keep almost constant quality among the encoded sequences. By working in the ρ-domain, we propose a low-delay rate allocation scheme that, at each time instant, provides a closed form solution for either the aforementioned problems. We show that minimizing the distortion variance instead of the average distortion leads, for each of the multiplexed sequences, to a coding penalty less than 0.5 dB, in terms of average PSNR. In addition, our analysis provides an explicit relationship between model parameters and this loss. In order to smooth the distortion also along time, we accommodate a shared encoder buffer to compensate for rate fluctuations. Although the proposed scheme is general, and it can be adopted for any video and image coding standard, we provide experimental evidence by transcoding bitstreams encoded using the state-of-the-art H.264/AVC standard. The results of our simulations reveal that is it possible to achieve distortion smoothing both in time and across the sequences, without sacrificing coding efficiency.**

***Index Terms*—Rate control, statistical multiplexing, video coding.**

I. Introduction

VIDEO coding standards, including the state-of-the-art H.264/AVC [1] codec, explicitly define only the bitstream syntax that a compliant decoder can interpret. The encoder is not subject to the standard, leaving the door open to the investigation of nonnormative tools. Among the others, rate control is an important module of the encoder that has received a great deal of attention in the past literature (a recent survey on this topic is given in [2]).

Rate control algorithms are responsible for optimally allocating the available bit budget to the coding units (i.e., group of pictures—GOP, pictures, macroblocks), in order to fulfill some optimality criterion. The general objective of optimal rate control is to minimize the distortion under a rate constraint; this goal can be further specified as minimum average distortion (MINAVE). In some cases, this criterion is not optimal from a subjective point of view, because it introduces annoying quality fluctuations in the encoded sequences. Therefore, an alternative goal might consist of minimizing the distortion variance (MINVAR), aiming at (almost) constant quality [3], [4]. For the case of a single sequence, the latter criterion aims at reducing the distortion variations of the encoded sequence from frame to frame, i.e., to smooth the video quality along time.

The rate control concepts developed for the case of a single sequence can be extended to the multiple video object (MVO) scenario, adopted by the MPEG-4 [5] framework. In this context, the problem is typically formulated by using simple quadratic models that relate the quantization step size q to the rate $R(q)$, and to the distortion $D(q)$. Vetro *et al.* [6] propose a technique that achieves constant bit rate when encoding multiple video objects. A heuristic rate allocation among the video objects is obtained based on the relative size, motion activity and prediction residual variance of the video objects. Then, a quadratic rate-distortion model is used to find the quantization parameters needed to fulfill the target bit rates. Buffer control policies are used to drive frame skipping and avoid buffer overflows. The design of the rate allocation algorithm for MVO's is cast into the optimization of a cost criterion based on signal quality parameters in [7]. Three cost criteria are proposed: weighted distortion (a weighted version of the MINAVE problem), priority based (i.e., an ordered list of distortion targets for the VO's needs to be specified), and constant distortion ratios (a generalization of the constant distortion problem, i.e., similar to MINVAR). More recently, [8] extends the work in [6] by proposing a prediction and feedback control to achieve accurate bit rate while handling buffer fullness. In [9], the authors propose to perform the target bit allocation by considering coding complexity both along the time and VOs. A feedback based approach is described, which adapts the model parameters on a frame-by-frame basis before encoding.

A specific instance of the MVO's case is represented by the joint encoding of frame-based sequences, i.e., the multiple video sequences (MVS) scenario. In fact, 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. In this scenario, at each time instant, the available bandwidth has to be optimally distributed across the sequences. This problem has been addressed in the past literature, and sometimes referred to as *statistical multiplexing* [10]–[12]. The term statistical multiplexing was originally coined to indicate the joint transmission of previously encoded variable bit rate video streams. Due to the diversity among the different video sequences, an approximately constant bit rate channel results from multiplexing several sequences. Nevertheless, it has been shown

Manuscript received August 14, 2007; revised March 10, 2008. This work was supported in part by the EU under the Visnet II Network of Excellence. The material in this paper was presented in part at the ACM Workshop on Mobile Video, Augsburg, Germany, September 2007, and in part at the Picture Coding Symposium, Lisbon, November 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Antonio Ortega.

The authors are with the Dipartimento di Elettronica e Informazione, Politecnico di Milano, 32 20133 Milano, Italy (e-mail: marco.tagliasacchi@polimi.it; valenzise@elet.polimi.it; stefano.tubaro@polimi.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2008.924278

[13] that equally partitioning the bit budget among the different video sequences is suboptimal, and a minimum-distortion approach typically results in a bandwidth saving. Therefore, here we adopt the meaning of statistical multiplexing given in [10], i.e., rate allocation is explicitly performed at encoding time, and the available bit budget is optimally distributed to the video sequences. As in the single sequence scenario, both MINAVE and MINVAR distortion problems can be formulated. Most of the literature on statistical multiplexing refers to encoding of MPEG-2 video, due to the relevance of this standard in the broadcasting application scenario. In [11] 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 [10]. The problem is formulated in the q-domain and, 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 [14], 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 [15] an optimal rate-allocation for MPEG-4 video objects is proposed, in the MINAVE sense. Since this represents the starting point for our work, we briefly summarize the main results of [15] in Section II. More recently, the MVS scenario has been applied to the multiplexing of H.264/AVC encoded video sequences. In [16], the authors propose a joint rate control for multiple H.264/AVC video encoders that uses a look-ahead processing window to allocate the bandwidth resources in order to reduce quality variations along time. This is shown to result also in a quality smoothing among the different sequences. Similar results are obtained in [17] by using a hierarchical approach to allocate rate at a "super-frame" level, composed by the frames of all sequences at a given time instant. Both these works show that using a joint rate-control module to reduce quality fluctuations along time results also in a reduction of distortion variance among programs. However, this phenomenon is not analyzed in detail, and the MINVAR problem between sequences is not clearly addressed. A method analogous to the previous ones that does not make use of a look-ahead window has been recently presented in [18] for the case of multiple H.264/AVC video sequences: the frames of the different sequences at a given time instant are grouped in a "multiframe" (MFRM), and a group of multiframes constitutes a "multi-GOP" (MGOP). The authors estimate the coding complexity of each frame using the ρ-domain model; then they allocate the available bits in such a way that the quality fluctuations between adjacent MGOPs and among the frames inside the MGOPs are minimized. To overcome the mismatch between the ideal ρ-domain model and the actual distortions, the authors propose to use a feedback mechanism for adjusting adaptively the frame level bit allocation. We compare the results of the proposed method with the work in [18] in Section V.

In this paper, we consider an application scenario that consists of multiplexing and transcoding H.264/AVC encoded video streams. Fig. 1 depicts the block diagram of the proposed system architecture. During decoding, a limited number of model parameters is extracted from each 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 sequences, according to the MINVAR criterion.

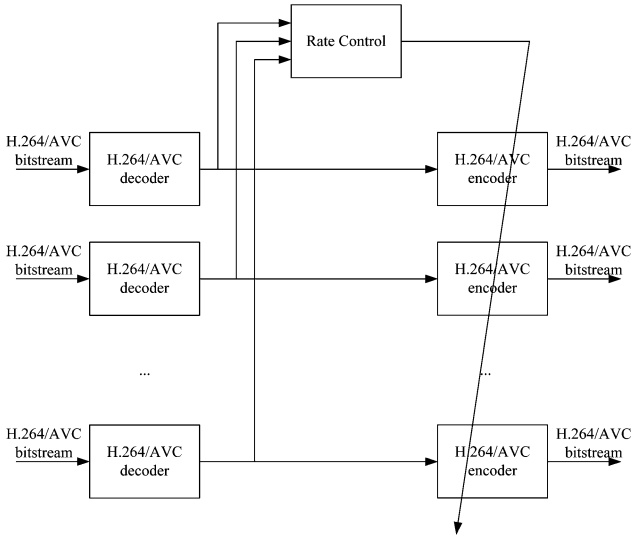


Fig. 1. Block diagram of the proposed transcoding/multiplexing architecture.

Our contribution is novel in the following aspects. First, we formulate the MINVAR problem in the ρ-domain. Since it is not mathematically tractable, we convert the MINVAR problem into an equivalent formulation, which admits a closed form solution. Second, we thoroughly analyze the coding efficiency loss suffered by the MINVAR solution with respect to the MINAVE solution. We prove that the loss factor is bounded and we deterministically relate this factor to a subset of the model parameters. By describing the distribution of model parameters in statistical terms, we show that, for conventional video sequences, the loss factor is typically less than 0.5 dB. In addition, to guarantee almost constant quality also along time, we introduce a global video buffer that can compensate the variability of the allocated bandwidth. The bitstreams produced in output are compliant with the H.264/AVC hypothetical reference decoder [19].

We highlight the fact that the results presented in this paper about the comparison between MINAVE and MINVAR problems are general, and their validity extends beyond the simple transcoding scenario. In fact, the latter differs from a conventional encoding scenario mainly because of the way model parameters can be estimated. Here, transcoding serves as a proof of concept, which allows to validate the results of the proposed analysis. Therefore, we adopt a simple explicit homogeneous (i.e., from H.264/AVC to H.264/AVC) transcoding algorithm that decodes the input sequence up to the pixel domain and re-encodes it, re-using mode decisions and motion vectors, thus avoiding drift. More sophisticated transcoding techniques, such as those reviewed in [20], could be used to further reduce the computational burden. Also, rate adaptation can be achieved by means of scalable video coding, instead of transcoding. In [21], MPEG-4 FGS is employed and quality smoothing is imposed

by controlling the number of residual bit-planes not encoded in the enhancement layer. An optimal bit allocation algorithm for multiplexed scalable streams is proposed in [22]. The method assumes the knowledge of the slope of the rate-distortion curve at the admissible truncation points.

The rest of this paper is organized as follows. In Section II, we briefly review the ρ-domain model proposed in [14], and we explain how to extract the parameters used in the subsequent optimization procedure from the video sequences. In Section III, both the MINAVE and the MINVAR rate allocation criteria are described. Exploiting a ρ-domain rate-distortion model, we reformulate the classical MINVAR problem into a simpler one; we then compare the performance of MINAVE and MINVAR criteria from a statistical point of view in Section III-C. In Section IV, we consider the problem of temporal quality smoothing by adding a global encoder buffer to the system. The theoretical results presented in Sections III and IV are experimentally validated in Section V. Finally, Section VI draws some concluding remarks.

II. Overview of the ρ-Domain Model

In [14], it is shown that in any typical transform domain system, there is always a linear relationship between the coding bit rate R and the percentage of zeros among the quantized transform coefficients, denoted by ρ, i.e.,

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

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 [23]. In [15], the exact expression of the distortion D as a function 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 [15]

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

where σ^2 and α are the model parameters that can be computed, as explained in [14], by fitting a Laplacian model to the histogram of quantized DCT coefficients. The ρ-domain rate distortion model has been originally proposed and verified for DCT-based image and video coding standards, including JPEG, H.263 [24] and MPEG-4 [14]. More recently, the same model has been successfully applied to H.264/AVC [25].

Here, we provide a brief summary of the estimation method for the model parameters θ, α and σ^2. The estimation of θ is straightforward. R, i.e., the number of bits needed to encode the prediction residuals, excluding mode decisions and motion vectors bits, is obtained during decoding. At the same time, the fraction ρ of transform coefficients quantized to the zero bin is computed, and $\theta = R/(1-\rho)$.

As for σ^2, let $\mathcal{D}(x)$ be the histogram of the reconstructed and scaled transform coefficients at the decoder after re-scaling, and $p_l(x;\lambda)$ the Laplacian probability density function obtained by fitting $\mathcal{D}(x)$, i.e.,

$$p_l(x) = \frac{\lambda}{2} e^{-\lambda|x|}. \tag{3}$$

As in [15], we set $\sigma^2 = 2/\lambda^2$.

By inverting (2), α can be written as

$$\alpha = \frac{1}{1-\rho} \ln \frac{\sigma^2}{D}. \tag{4}$$

If the original frame were available, the computation of the only missing term, D, would be straightforward. In a transcoding scenario, originals are not available, and an estimation of D needs to be obtained based on the quantization step size q. The H.264/AVC standard adopts a uniform quantizer with dead zone. Let $\Delta = (0.5+b)q$ denote the dead zone and b is a non-negative constant that adjusts the width of the dead zone. In H.264/AVC, $b = 1/6$ for I slices and $b = 1/3$ for P slices. In [15], a closed form expression for the distortion is explicitly derived

$$D = \frac{1 + e^{-\lambda a \Delta} - 2e^{-\lambda\Delta} + \lambda\Delta e^{-\lambda\Delta}(a-1)(1+e^{-\lambda a\Delta})}{\lambda(1+e^{-\lambda a\Delta})} \tag{5}$$

where $a = 1/(1+2b)$. The value obtained using (5) can be inserted in (4) to obtain the estimate of α.

III. Rate Allocation

In this section, we formulate the rate allocation problem among the multiplexed sequences. In order to enable a low-delay implementation of the proposed scheme, rate allocation is performed independently at each time instant t. Therefore, the proposed model does not explicitly address the interframe dependencies introduced by the prediction loop. More sophisticated techniques such as those indicated in [26] could be adopted, at the cost of a significant increase in computational complexity, which grows with the length of the group of pictures. In fact, the complexity of the optimization algorithm in [26] is dominated by the data collection phase, which requires the computation of the Lagrangian costs associated with any pair of quantizers for two dependent frames. Conversely, the computational complexity of the proposed approach is much lower, yet the goal of minimizing intersequence variance is achieved, as supported by experimental evidence in Section V. This can be explained by observing that, in the MINVAR scenario, the reference frames of the multiplexed sequences have comparable distortion levels, thus justifying the adopted simplification.

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 the following, we consider two distinct optimization problems: MINAVE (minimum average distortion) and MINVAR (minimum variance distortion).

A. MINAVE—Minimum Average Distortion

In order to find the optimal rate allocation that minimizes the average distortion of the output, we need to solve the following nonlinear constrained optimization problem:

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

Using (2), the minimization problem (6) becomes

$$\min_{\mathbf{R}} \frac{1}{S}\sum_{i=1}^{S}\sigma_i^2 e^{-\alpha_i(1-\rho_i)}, \quad \text{s.t.} \quad \sum_{i=1}^{S}\theta_i(1-\rho_i) \le R. \quad (7)$$

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

$$R_i^* = \xi_i \log\frac{\sigma_i^2}{\xi_i} + \frac{\xi_i}{\sum_{j=1}^{S}\xi_j}\left(R - \sum_{j=1}^{S}\xi_j \log\frac{\sigma_j^2}{\xi_j}\right) \quad (8)$$

where $\xi_i = \theta_i/\alpha_i$. This formulation has already been addressed in the past in [15], for the problem of optimally allocating the rate among multiple video objects in MPEG-4.

B. MINVAR—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

$$\min_{\mathbf{R}} \frac{1}{S}\sum_{i=1}^{S}(D_i(\rho_i) - \bar{D})^2, \quad \text{s.t.} \quad \sum_{i=1}^{S}\theta_i(1-\rho_i) \le R \quad (9)$$

where $\bar{D} = (1/S)\sum_{i=1}^{S} D_i(\rho_i)$. This problem is difficult to solve in closed form, since $\bar{D}$ depends on the whole set of distortion values D_i of each sequence. To overcome this limitation, in [27], we have shown that it is possible to reformulate problem (9) into an equivalent one, in order to achieve equal distortion for all the sequences. We evaluate in Section III-C the "goodness" of the obtained distortion value against the MINAVE solution.

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 minimization problems $\mathcal{P}_1, \mathcal{P}_2, \ldots \mathcal{P}_n, \ldots$, respectively, where the problem $\mathcal{P}_n$ is described as

$$\mathcal{P}_n: \quad \min_{\mathbf{R}} \sum_{i=1}^{S} D_i^n(\rho_i) = \min_{\mathbf{R}} \sum_{i=1}^{S}\sigma_i^{2n} e^{-n\alpha_i(1-\rho_i)}$$
$$\text{s.t.} \quad \sum_{i=1}^{S}\theta_i(1-\rho_i) \le R. \quad (10)$$

In this problem, the distortion terms are decoupled from the distortions of the other sequences; therefore, we can easily solve (10) using Lagrange multipliers method and find

$$\tilde{R}_i^{(n)} = \xi_i \log\sigma_i^2 + \frac{\xi_i R}{\sum_{j=1}^{S}\xi_j} - \frac{\xi_i \sum_{j=1}^{S}\xi_j \log\sigma_j^2}{\sum_{j=1}^{S}\xi_j} + \cdots + \frac{1}{n}\left[\frac{\xi_i \sum_{j=1}^{S}\xi_j \log\xi_j}{\sum_{j=1}^{S}\xi_j} - \xi_i \log\xi_i\right]. \quad (11)$$

The following property holds (refer to Appendix I for the proof).

Property 1: The sequence of distortion variances, $\{\mathrm{var}\tilde{\mathbf{D}}^{(n)}\}$, converges to 0 as $n \to \infty$.

Property 1 states that, by solving $\mathcal{P}_n$ for $n \to \infty$, the solution is such that $\tilde{D}_1^{(n)} = \tilde{D}_2^{(n)} = \cdots = \tilde{D}_S^{(n)} = \tilde{D}$, where

$$\tilde{D} = \exp\left[\frac{\sum_{i=1}^{S}\xi_i \log\sigma_i^2 - R}{\sum_{i=1}^{S}\xi_i}\right] \quad (12)$$

as derived in Appendix I, thus achieving minimum variance, which is the goal of (9). The set of rates that produce the distortion of (12) can be easily found by computing the limit, for $n \to \infty$, of the rates (11)

$$\begin{aligned}\tilde{R}_i &= \lim_{n\to\infty}\tilde{R}_i^{(n)} \\ &= \xi_i \log\sigma_i^2 + \frac{\xi_i}{\sum_{j=1}^{S}\xi_j} \cdot \left(R - \sum_{j=1}^{S}\xi_j \log\frac{\sigma_j^2}{\xi_j}\right) \\ &= R_i^* + \xi_i \log\xi_i - \frac{\xi_i \sum_{j=1}^{S}\xi_j \log\xi_j}{\sum_{j=1}^{S}\xi_j}. \end{aligned} \quad (13)$$

This result allows us to find a closed form solution of the MINVAR problem for the allocation of target bit rates R_i so that all the video programs have the same distortion level $\tilde{D}$.

The equivalence between solving problems (10) as $n \to \infty$ and the solution of (9) needs further considerations. Since the variance is always greater than or equal to zero, and $\mathrm{var}\{\mathbf{D}_i^{(n)}\}$ converges to zero, clearly the rates of (13) correspond to a point of global minimum for the variance minimization problem, as written in (9). The question now is whether or not there exist other solutions besides (13) that minimize (9).

We start by considering that, if a solution exists, in order to be a global minimum it should satisfy

$$D_1 = D_2 = \cdots = D_S = \bar{D} \quad (14)$$

where $\bar{D}$ is the average distortion that may be different from $\tilde{D}$. The distortion-rate function $D(R)$, parameterized in the ρ-domain, may be rewritten explicitly as

$$D_i = \sigma_i^2 e^{-R_i/\xi_i} \quad (15)$$

from which it is easy to write the inverse relationship $R(D)$

$$R_i = \xi_i \log\sigma_i^2 - \xi_i \log D_i. \quad (16)$$

Using (16), (14) becomes an undetermined system of S equations in the unknown $\bar{D}$

$$\begin{aligned} R_1 &= \xi_1 \log\sigma_1^2 - \xi_1 \log\bar{D} \\ R_2 &= \xi_2 \log\sigma_2^2 - \xi_2 \log\bar{D} \\ &\vdots \\ R_S &= \xi_S \log\sigma_S^2 - \xi_S \log\bar{D} \end{aligned}$$

which can be conveniently written in vector form as

$$\mathbf{r} = \mathbf{a} - \log\bar{D} \cdot \boldsymbol{\xi} \quad (17)$$

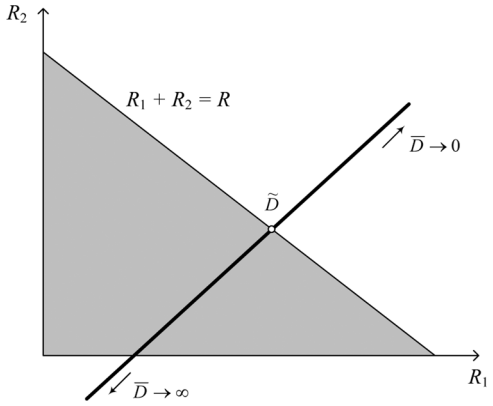


Fig. 2. Geometrical interpretation of (9) and (10) for the case of two sequences ($S = 2$). The shaded area encloses the feasible solutions and is limited by the constraint (19). The bold line is the locus of equal distortion solutions, given by (17). Solving our MINVAR problem (10) is equivalent to solve the minimum variance problem (9) with the equality constraint.

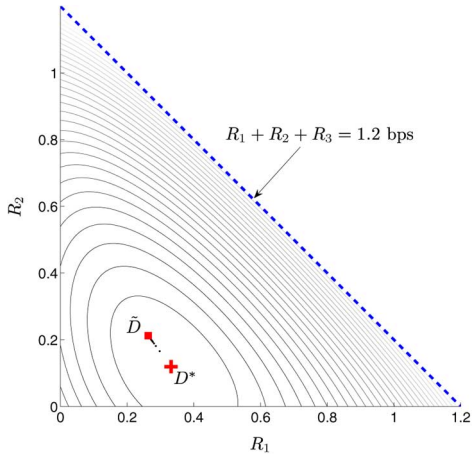


Fig. 3. MSE distortion surface versus rate allocation for three video sequences, when the total available bit budget is 1.2 bpp. The smaller points are the trajectory of the sequence $1/3 \cdot \sum_{i=1}^{3} \bar{D}_i^{(n)}$, for $n = 2 \ldots 30$.

where

$$\mathbf{r} = \begin{pmatrix} R_1 \\ \vdots \\ R_S \end{pmatrix}, \quad \mathbf{a} = \begin{pmatrix} \xi_1 \log \sigma_1^2 \\ \vdots \\ \xi_S \log \sigma_S^2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_S \end{pmatrix}. \tag{18}$$

Equation (17) says that there are infinite solutions to the minimum variance problem (9). In fact, when $\bar{D}$ is varied in the interval $(0, +\infty)$, (17) spans a 1-D subspace (i.e., a line) in $\mathbb{R}^S$ (since $\log \bar{D}$ is a monotonic mapping of $\bar{D}$), as illustrated in Fig. 2, for $S = 2$. We show now that the solution of (10) when $n \to \infty$, which lies on this line, is the set of rates $\tilde{R}_i$ that produces the least average distortion among all the possible solutions of (9). In fact, since the ρ-domain parameters ξ_i are all positive, as the average distortion $\bar{D}$ decreases towards zero, the rates R_i grow indefinitely in the direction of $\boldsymbol{\xi}$, which is a vector pointing outwards the half-space of $\mathbb{R}^S$ defined by the constraint

$$\sum_{i=1}^{S} R_i \leq R. \tag{19}$$

Clearly, (19) with the equality sign is the boundary of that region: we can move from a point inside the region towards the boundary along the iso-distortion line (17) by reducing the value of $\bar{D}$. It turns out that the *only* point at which the line intersects the boundary corresponds to the minimum feasible distortion $\bar{D}^*$, which is given by the set of rates that fulfill (19) with equality. We show in the proof of Property 1 that $\tilde{D}$ is the average distortion obtained by solving the convex problem (10) using the Lagrange multipliers method, i.e., with the equality in (19). Thus, we can conclude that $\bar{D}^* = \tilde{D}$, i.e., the MINVAR solution is, among all possible solutions of (9), the one with the *least* distortion or, equivalently, the one which solves (9) with the equality constraint. Fig. 2 provides a geometrical interpretation of this fact, when $S = 2$.

Using this geometric argument, one could obtain easily the value of $\tilde{D}$, by imposing the equality constrain in (19)

$$\sum_{i=1}^{S} \xi_i \log \sigma_i^2 - \log \tilde{D} \cdot \sum_{i=1}^{S} \xi_i = R \tag{20}$$

which gives as solution

$$\log \tilde{D} = \frac{\sum_{i=1}^{S} \xi_i \log \sigma_i^2 - R}{\sum_{i=1}^{S} \xi_i} \tag{21}$$

that can be recognized to be exactly the same expression in (45). However, our formulation with $\mathcal{P}_n$ gives more insight into the problem, since it allows to find solutions that are in-between the MINAVE distortion D^* and the MINVAR distortion $\tilde{D}$. To exemplify this fact, in Fig. 3, we show the MINAVE distortion, which is the special case of $\mathcal{P}_n$ when $n = 1$, and the MINVAR distortion, obtained for $n \to \infty$, when the total rate budget is $R = 1.2$ bpp, for three multiplexed video sequences. The small dots represent the solutions for $n = 2, 3, \ldots 30$. Sweeping all possible values of $n \in [1, \infty)$ one can achieve all possible tradeoffs between D^* and $\tilde{D}$.

C. MINAVE Versus MINVAR Comparison

In this section, we investigate the coding efficiency loss incurred by solving the MINVAR problem instead of the MINAVE problem. We demonstrate that while seeking minimum variance, the average distortion level is increased.

Let D^* denote the average distortion level attained by solving the MINAVE problem (6), while $\tilde{D}$ is the (equal) distortion level attained by solving the MINVAR problem, as defined in (12). In addition, let $\zeta_i = (\xi_i)/(\sum_{i=1}^{S} \xi_i)$, $0 \leq \zeta_i \leq 1$, be the normalized values of ξ_i and $H(\zeta) = -\sum_{i=1}^{S} \zeta_i \log \zeta_i$ the entropy function of a discrete memoryless source having the set $\zeta_i, i = 1 \ldots S$ as the probability mass function of its S symbols. Based on these definitions, in Appendix II, we prove the following property.

Property 2: The coding efficiency loss $D^*/\tilde{D}$ between the MINVAR rate allocation and the MINAVE solution is

$$\frac{D^*}{\tilde{D}} = \frac{e^{H(\zeta)}}{S}. \tag{22}$$

Property 2 states that the increase in distortion attained when solving the MINVAR problem can be quantified exactly as a function of the model parameters α_i, θ_i, $i = 1, \ldots, S$, from which the values of ζ_i are immediately computed.

From information theory, we know that

$$0 \leq H(\zeta) \leq \log(S). \tag{23}$$

Therefore, we obtain the following bounds as a corollary of Property 2:

$$\frac{1}{S} \leq \frac{D^*}{\tilde{D}} \leq 1. \tag{24}$$

We can take advantage of the result in (22) in two different ways.

- Given the model parameters α_i and θ_i, the value of (22) can be deterministically computed. Therefore, at each time instant, before solving any rate allocation problem, the coding efficiency loss due to the MINVAR problem can be determined. If such a loss is large, one can temporarily switch to the MINAVE problem.
- Characterizing the distribution of model parameters, it is possible to derive the average loss incurred by the MINVAR problem in statistical terms. This is further investigated in the following, where we show that a lower bound tighter than the one presented in (24) typically holds, for the problem of multiplexing real video sequences.

We are ultimately interested in estimating a lower bound for the following quantity:

$$E\left[\frac{D^*}{\tilde{D}}\right] = \frac{1}{S} E\left[e^{H(\zeta)}\right] \tag{25}$$

where $E[\,\cdot\,]$ denotes the expectation with respect to the statistical distribution of the model parameters. By the Jensen's inequality, we can write

$$E\left[\frac{D^*}{\tilde{D}}\right] = \frac{1}{S} E\left[e^{H(\zeta)}\right] \geq \frac{1}{S} e^{E[H(\zeta)]}. \tag{26}$$

In order to find $E[H(\zeta)]$, we need to define an appropriate statistical model describing the distribution of ξ; the distribution of ζ can then be found accordingly to the model of ξ. Given the actual parameter values estimated by decoding several H.264/AVC bitstreams, we found that $\theta_i/\alpha_i = \xi_i \sim \mathrm{Gamma}(\xi; a_i, b_i)$ (see Fig. 4)

$$\mathrm{Gamma}(\xi; a, b) = \xi^{a-1} \frac{b^a e^{-b\xi}}{\Gamma(a)} \tag{27}$$

where a is the shape parameter and b is the inverse of the scale parameter, whereas $\Gamma(\,\cdot\,)$ denotes the gamma function.

It can be shown [29] that, if $\xi_i \sim \mathrm{Gamma}(\xi; a_i, b)$, then

$$\begin{aligned}(\zeta_1, \zeta_2, \ldots, \zeta_S) &= \left(\frac{\xi_1}{\sum_{i=1}^S \xi_i}, \frac{\xi_2}{\sum_{i=1}^S \xi_i}, \ldots, \frac{\xi_S}{\sum_{i=1}^S \xi_i}\right) \\ &\sim \mathrm{Dir}(\zeta_1, \zeta_2, \ldots, \zeta_S; a_1, a_2, \ldots, a_S)\end{aligned} \tag{28}$$

where $\mathrm{Dir}(\zeta_1, \zeta_2, \ldots, \zeta_S; a_1, a_2, \ldots, a_S)$ denotes the multivariate Dirichlet distribution, which is defined as

$$\mathrm{Dir}(\zeta_1, \zeta_2, \ldots, \zeta_S; a_1, a_2, \ldots, a_S) = \frac{1}{B(\mathbf{a})} \prod_{i=1}^S \zeta_i^{a_i - 1} \tag{29}$$

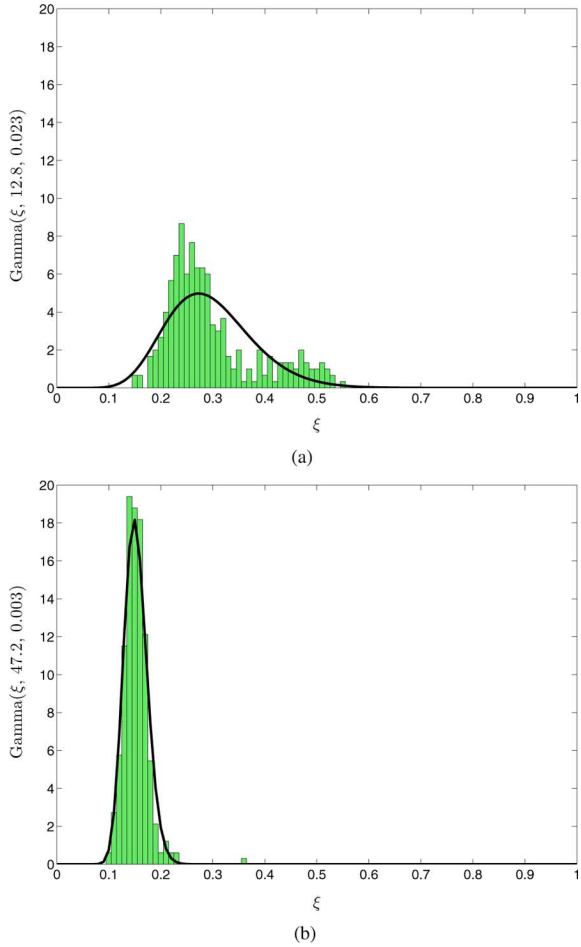


Fig. 4. Gamma fitting of the histograms of ξ. (a) Foreman. (b) Hall Monitor.

where $\zeta_i \geq 0$, $\sum_{i=1}^S \zeta_i = 1$, and $a_i > 0$. The normalizing constant is the multinomial beta function, which can be expressed in terms of the gamma function

$$B(\mathbf{a}) = B(a_1, a_2, \ldots, a_S) = \frac{\prod_{i=1}^S \Gamma(a_i)}{\Gamma(\sum_{i=1}^S a_i)}. \tag{30}$$

We notice that the Dirichlet distribution is defined over the S-dimensional simplex given by the constraints $\zeta_i \geq 0$, $\sum_{i=1}^S \zeta_i = 1$. Fig. 5 illustrates the Dirichlet probability density function in $\mathbb{R}^3$ for different parameter vectors.

We can exploit the knowledge about the probability distribution of the parameters ξ_i to find the bound (26). After some laborious calculations (see Appendix III), one finds that the expected entropy $E[H(\zeta)]$ is

$$E[H(\zeta)] = \psi(a_0 + 1) - \frac{1}{a_0} \sum_{i=1}^S a_i \psi(a_i + 1) \tag{31}$$

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

If we observe a long temporal interval and we do not make prior assumptions about the distribution of the individual multiplexed sequences, we can assume that ξ_is, $i = 1, \ldots, S$ are i.i.d., i.e., $\xi \sim \mathrm{Gamma}(\xi; a, b)$. Fig. 6 shows the fitting of the

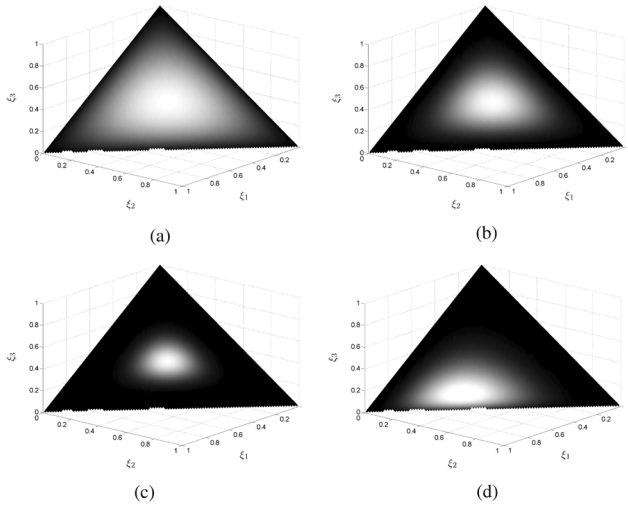


Fig. 5. Probability density of the Dirichlet distribution when $S = 3$. (a) $a_1 = a_2 = a_3 = 2$. (b) $a_1 = a_2 = a_3 = 4$. (c) $a_1 = a_2 = a_3 = 10$. (d) $a_1 = 6, a_2 = 4, a_3 = 2$‘.

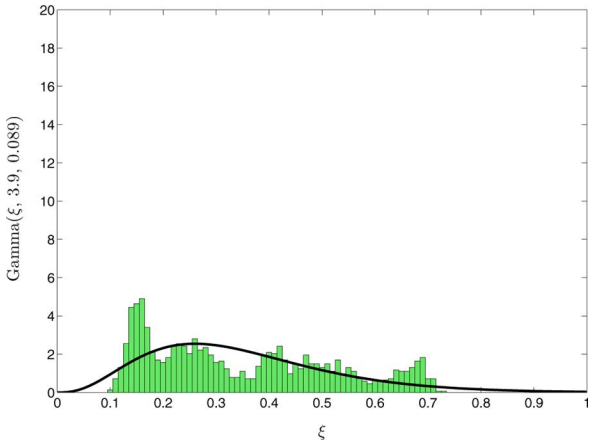


Fig. 6. Gamma fitting of the histograms of ξ, obtained by concatenating *Foreman*, *Hall Monitor*, *Soccer, Coastguard*, and *Mobile* sequences.

gamma probability density function on the histogram of ξ's collected by concatenating the sequences *Foreman*, *Hall Monitor, Soccer, Coastguard,* and *Mobile*. We notice that, for real video sequences, the shape parameter a is typically larger than 3–4. When all the a_i are the same $(a_i = a, \forall i)$, (31) becomes

$$E[H(\zeta); a_i = a] = \psi(Sa + 1) - \psi(a + 1). \qquad (32)$$

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 decrease as a grows, as illustrated in Fig. 5.

The result in (26) allows us to impose a much tighter bound than the one shown in (24), once the value of a is known. Fig. 7 illustrates the bounds obtained for different values of a. We also show the values of $E[(D^*)/(\tilde{D})] = (1/S)E[e^{H(\zeta)}]$ computed by means of Montecarlo simulations, by sampling the Dirichlet distribution. We notice that, for a fixed value of a, the coding efficiency gap between D^* and $\tilde{D}$ increases with the number of

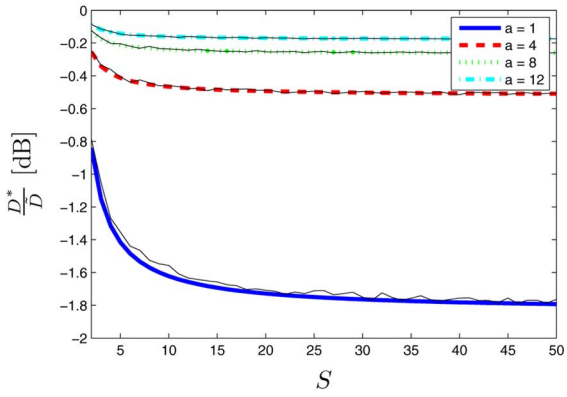


Fig. 7. Coding efficiency loss as a function of the number of multiplexed sequences. Thick lines represent the lower bound in (26) for different values of a. Thin lines are the result of Montecarlo simulations for $E[D^*/\tilde{D}]$ obtained by sampling the Dirichlet distrubution.

multiplexed sequences S. In addition, for value of a typically encountered in practice, i.e., $a \approx 4$, the coding efficiency loss is smaller than 0.5 dB, even for $S \to \infty$. This result suggests that, for the problem at hand, solving the MINVAR problem achieves the goal of minimum variance, without compromising the overall coding efficiency.

In order to validate the aforementioned results, we estimated the parameters of the ρ-domain model as explained in Section II for 300 frames of the following test sequences: *Hall Monitor*, *Foreman*, and *Soccer*, all at CIF spatial resolution (352×288 pixels). All the sequences are first encoded at a fixed $\mathrm{QP} = 20$ using H.264/AVC. The first frame is intracoded (I slices) and the remaining frames are all interframe coded (P slices). Fig. 8(a) shows the frame-by-frame rate allocation computed solving the MINAVE problem when the bit budget for each sequence is set to $1/3$ bpp (i.e., $R = S \cdot 1/3 = 1$ [bpp×number of sequences]). We notice that, at each time instant, the distortion level of the individual sequences, as predicted by the exponential R-D model (2), can be significantly different from each other, with a gap as large as 6 dB. Fig. 8(b) shows the rate allocation computed solving the MINVAR problem. Using these rates as input for the R-D model, we notice that all sequences attain exactly the same distortion level. It is instructive to compare the average distortion curves of Fig. 8(a) and (b). Fig. 9 shows $D^*/\tilde{D}$ expressed in decibels. We notice that the maximum loss is equal to -0.70 dB while the average loss is equal to -0.15 dB.

In Section V, we will repeat this experiment, using the rates shown in Fig. 8(a) and (b) in a real transcoding scenario. Instead of the distortions given as the solution of problem (6) and (10), we will show the actual PSNR tracks for each sequence when the set of rates R_i^* and $\tilde{R}_i$ are imposed at transcoding: we anticipate that the quality variations among sequences in the MINVAR case, although not null, are significantly reduced w.r.t. the MINAVE optimization.

IV. Temporal Smoothing

In the previous section, we considered a strictly constant bit rate (CBR) optimization, i.e., $R(t) = R \quad \forall t$, either in the MINAVE or MINVAR sense. However, this can produce distortion profiles with large fluctuations along time [see Fig. 8(a) and

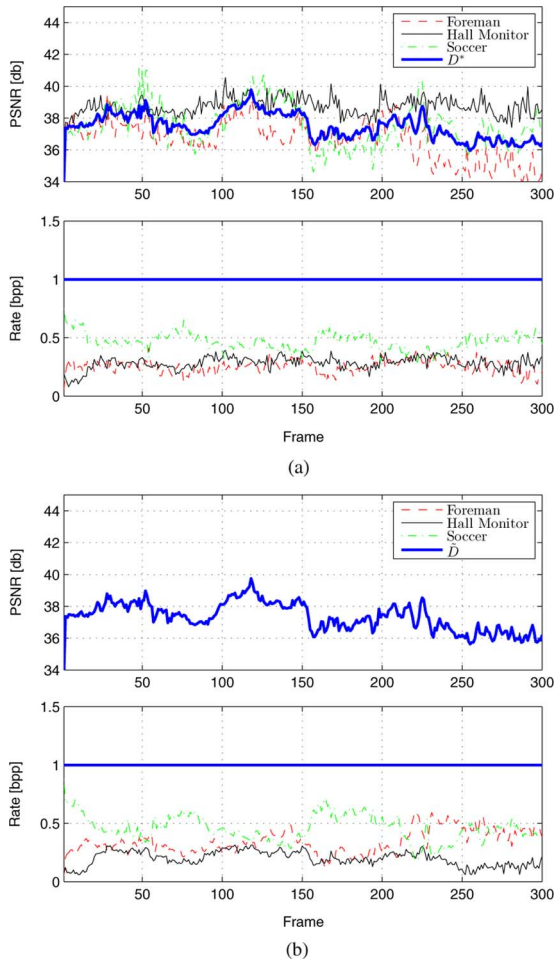


Fig. 8. Distortion and rate profiles using MINAVE and MINVAR rate allocation for three video sequences. The rates for each sequence are obtained from (13), using the ρ-domain model parameters extracted from each video frame. The bold constant line in the rate plots shows that the CBR constraint is satisfied. The expected distortion profile for each sequence is then computed according to the exponential model (2). (a) MINAVE. (b) MINVAR.

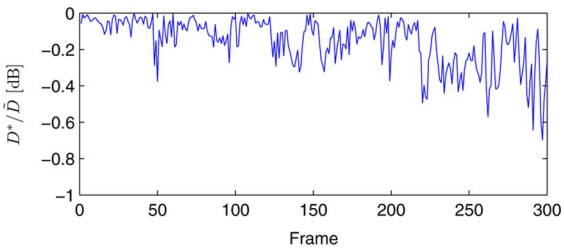


Fig. 9. Quality loss, expressed in decibels, between the optimal MINVAR and the MINAVE bit allocation strategies.

(b)]. 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, we need, therefore, to add an *encoder buffer* into the system. In this paper we describe the case of a global encoder buffer, which should be placed on the right of the H.264/AVC encoders in Fig. 1. In the following, we describe the temporal smoothing algorithm adopted in this paper, and we will refer to the resulting rate allocation as S-MINVAR (smoothed MINVAR).

In [30], it is proved that using a geometric averaging filter, it is possible to smooth the optimal minimum distortion profile while achieving, on average, the target bit rate. Let $\tilde{D}(t)$ be the distortion at frame t computed as the solution of the MINVAR problem under a constant bit rate (CBR) $R(t) = R, \forall t$, as explained in Section III. We define the smoothed target distortion at time t as

$$D_{\text{smooth}}(t) = \prod_{k=0}^{M-1} [\tilde{D}(t-k)]^{\frac{1}{M}} \tag{33}$$

where M is the length of the averaging window (e.g., $M = 30$ frames). Therefore, to maintain a temporally-smooth average distortion, we need:

1) to compute the CBR distortion profile;
2) to smooth it using (33);
3) to set $D_{\text{smooth}}(t)$ as target distortion and find the rates $R_i(t)$ which meet $D_{\text{smooth}}(t)$ in a MINVAR 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 expressed in bits; B_0 is the desired buffer level; $b(t)$ denotes the buffer fullness at time t; finally, let C be the channel rate, i.e., the rate at which the buffer is drained. The buffer state evolves according to the difference equation

$$b(t) = b(t-1) + P \cdot \sum_{i=1}^{S} R_i(t) - C \tag{34}$$

where P is the number of pixels in a frame. 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.

In the following, in order to compare the temporally smoothed solution with the solution of the MINAVE and MINVAR problems illustrated in Section III, we set $C = R$.

A. Unconstrained Smoothed Distortion Tracking

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 unconstrained minimization problem:

$$\min_{\mathbf{R}} \sum_{i=1}^{S} (D_i(t) - D_{\text{smooth}}(t))^2. \tag{35}$$

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

$$R_i(t) = \xi_i(t)(\log \sigma_i^2(t) - \log D_{\text{smooth}}(t)). \tag{36}$$

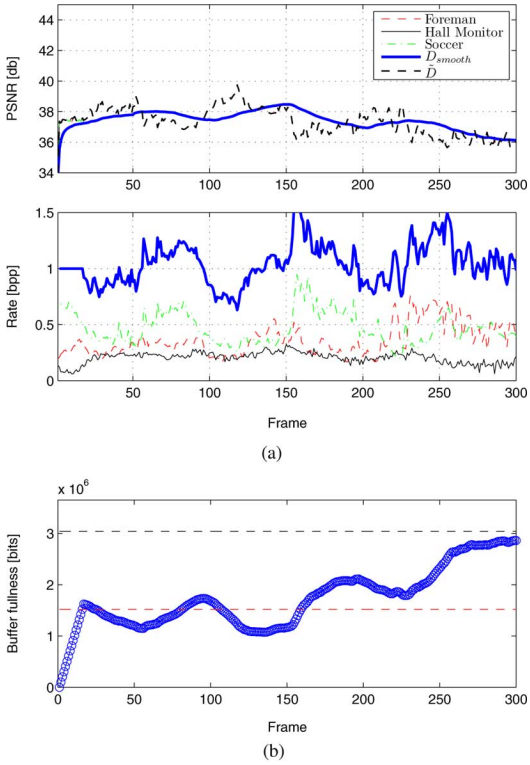


Fig. 10. Smoothed MINVAR rate allocation. (a) S-MINVAR. (b) Buffer fullness.

B. Constrained Smoothed Distortion Tracking

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 [30]). Therefore, the new CBR target becomes

$$R' = R - \frac{B_{\text{res}}}{K}. \tag{37}$$

The value of distortion $\tilde{D}_{\text{CBR}}(t)$ is smoothed with (33), and the target distortion $D_{\text{smooth}}(t)$ is used to find the rates for each sequence with (36). The result of quality smoothing on the three CIF video sequences *Foreman*, *Hall monitor* and *Soccer* is shown in Fig. 10(a). 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, $1/3$ bpp). Fig. 10(b) shows the buffer fullness level at each time instant. $B_{\max}$ is set to be 1s of delay, corresponding to 30 frames, and e.g., $B_0 = 0.5 \cdot B_{\max}$. Initially, the buffer level $b(0)$ is set to 0.

V. Experimental Results

In the following, we apply the MINAVE, MINVAR and S-MINVAR rate allocation algorithms illustrated in Sections III and IV to the problem of transcoding and multiplexing H.264/AVC encoded sequences.

The test sequences are at CIF spatial resolution (352×288 pixels) and 30 frames/s. The input bitstreams are produced by encoding each sequence at a fixed $\text{QP} = 20$ using H.264/AVC, baseline profile. In the transcoding module, model parameters are estimated as explained in Section II, and rate allocation is performed solving either the MINAVE, MINVAR or S-MINVAR problem. At each time instant, a modified total rate constraint $R'(t) = R(t) - \sum_{i=1}^{S} R_i^H(t)$ is imposed, where $R_i^H(t)$ denotes the number of header bits. In fact the transcoder does not perform rate-distortion optimization, but it inherits mode decisions and motion vectors from the incoming bit stream. The transcoder simply adjusts the quantization parameter, at the macroblock level, in order to attain the desired bit budget $R_i(t) = R_i'(t) + R_i^H(t)$. The rate control algorithm proposed in [25] is used for this purpose. Although this is not optimal in a rate-distortion sense, it enables fast transcoding of multiple video streams, making this approach suitable for real-time low-delay implementations.

Before detailing the results of the experiments, we introduce a few metrics that will be used to assess the goodness of the rate allocation schemes.

- MINVAR coding efficiency loss w.r.t. MINAVE rate allocation

$$\Delta\text{PSNR}_{\text{MINVAR}} = 10\log_{10} \times \frac{\sum_{i=1}^{S}\sum_{t=1}^{T} D_i^*(t)}{\sum_{i=1}^{S}\sum_{t=1}^{T} \tilde{D}_i(t)}. \tag{38}$$

- S-MINVAR coding efficiency loss w.r.t. MINAVE rate allocation

$$\Delta\text{PSNR}_{\text{S-MINVAR}} = 10\log_{10} \times \frac{\sum_{i=1}^{S}\sum_{t=1}^{T} D_i^*(t)}{\sum_{i=1}^{S}\sum_{t=1}^{T} D_{\text{smooth},i}(t)}. \tag{39}$$

- MINAVE variance

$$\sigma^2_{MINAVE} = \frac{1}{T}\sum_{t=1}^{T} \text{var}\{\mathbf{D}^*(t)\}. \tag{40}$$

- MINVAR variance

$$\sigma^2_{\text{MINVAR}} = \frac{1}{T}\sum_{t=1}^{T} \text{var}\{\tilde{\mathbf{D}}(t)\}. \tag{41}$$

- S-MINVAR variance

$$\sigma^2_{\text{S-MINVAR}} = \frac{1}{T}\sum_{t=1}^{T} \text{var}\{\mathbf{D}_{\text{smooth}}(t)\}. \tag{42}$$

$\text{var}\{\cdot\}$ denotes the sample variance across the S sequences.

Fig. 11 shows the PSNR tracks obtained using the same settings used for the simulations in Section III. Three sequences (*Foreman*, *Hall Monitor*, and *Soccer*) characterized by different motion characteristics are multiplexed and transcoded. In order to match the same test conditions of Section III-C, the bit budget for each sequence is set to 1/3 bpp, i.e., the bit budget is $R = S \cdot 1/3 = 1$ [bpp×number of sequences]. The overall channel bit rate can be computed as $352 \cdot 288 \cdot R \cdot 30$ bps (bits/s), using CIF resolution sequences at a frame rate of 30 frames/s. The first frame is intracoded, while the remaining frames are intercoded. By applying the rate allocation determined as the solution of the MINAVE problem [see Fig. 8(a)], we notice large quality variability both across the sequences ($\sigma^2_{\text{MINAVE}} = 35.01$) and along time. The intersequence variance is significantly reduced

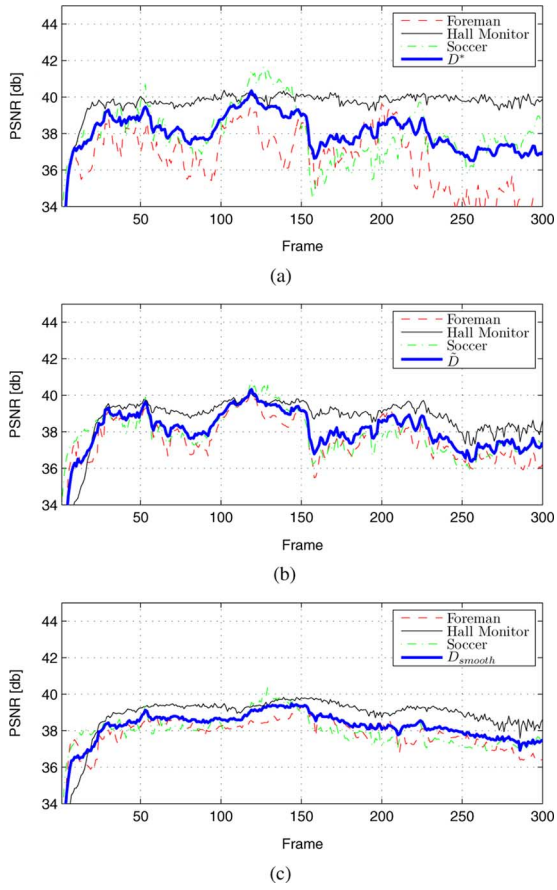


Fig. 11. MINAVE versus MINVAR versus S-MINVAR optimization applied to the transcoding of 3 H.264/AVC coded video sequences. (a) MINAVE. (b) MINVAR. (c) S-MINVAR.

by enforcing the solution of the MINVAR problem, as shown in Fig. 11(b) ($\sigma^2_{\mathrm{MINVAR}} = 4.08$), at the price of a small coding efficiency loss ($\Delta\mathrm{PSNR}_{\mathrm{MINVAR}} = -0.29$ dB on average). The S-MINVAR rate allocation is illustrated in Fig. 11(c), using a buffer size of 1 s. Besides preserving the smoothness across the multiplexed sequences ($\sigma^2_{\mathrm{S-MINVAR}} = 3.34$), temporal fluctuations are significantly reduced, producing a more visually pleasing result. The coding efficiency loss $\mathrm{PSNR}_{\mathrm{S-MINVAR}}$ with respect to the MINAVE scenario is -0.48 dB.

We have compared these results with the joint rate allocation algorithm proposed in [18], denoted hereafter as the MVS (multivideo sequence) algorithm. This technique uses a hierarchical approach to allocate bits to "multiframes" (MFRMs), i.e., the set of frames of different video sequences at a given time instant, and to "multi-GOPs," i.e., groups of multiframes. The bit budget is then partitioned inside MFRMs according to the coding complexity of each sequence, which is given by the estimated ρ-domain parameters. To reduce the mismatch between the ideal model and the actual H.264/AVC video sequences distortions, the MVS algorithm introduces a feedback mechanism that allows to yield *both* intersequence and temporal quality smoothing. Evaluation metrics $\Delta\mathrm{PSNR}_{\mathrm{MVS}}$ and σ^2_{MVS} can be defined in a similar manner as in (38)–(42). Table I shows the results of the MINAVE, MINVAR, S-MINVAR and MVS techniques, obtained by mixing different sequences together, at a rate equal to $1/3$ bpp respectively. We notice that the largest coding efficiency loss incurred by the MINVAR solution is -0.35 dB. Also the S-MINVAR solution is characterized by a coding efficiency loss of the same order of magnitude. The MVS algorithm, instead, can achieve in some cases a better distortion than the MINAVE solution: this is due to the fact that the MVS method estimates the model parameters re-encoding each frame, while in the cases of MINAVE and MINVAR, the ρ-domain parameters are given by the previously encoded sequences that are passed as input to the transcoder. Therefore, while this last approach is computationally simpler, it gives a suboptimal estimate w.r.t. MVS, especially when the gap between the input rate and the target output rate R is significant. In terms of intersequence distortion variability, both MINVAR and S-MINVAR consistently outperform MINAVE. For all the combinations of sequences, the distortion variance produced by MINVAR and S-MINVAR is less than the variability obtained using MVS. Table II shows similar results, but at a larger rate, equal to $1/2$ bpp. We notice that increasing the bit budget the intersequence distortion variability is significantly reduced. Also, the gain of the MVS approach in terms of average distortion is quite negligible, while the distortion variance of MINVAR and S-MINVAR is always less than the one obtained with MVS. This is due to the fact that the model parameters are extracted from sequences encoded at $\mathrm{QP} = 20$, resulting in a bit rate typically larger than $1/2$ bpp. For this reason, the $R(\rho)$ and $D(\rho)$ models do not fit equally well over a large range of bit rates, but they tend to be more accurate in the neighborhood of the input sequence bit rate.

Fig. 12 shows the PSNR tracks resulting from the proposed rate allocation algorithms, when one frame every 30 frames is intracoded. In the MINAVE and MINVAR scenarios, the CBR constraints imposes large quality fluctuations, resulting in large dips when intracoded frames occur. Conversely, in the S-MINVAR scenario temporal smoothing is achieved, thanks to encoder buffer that can tolerate rate variability.

VI. Conclusion

In this paper, we have considered the problem of allocating the available bit rate among different video sequences, in such a way that the quality variance among sequences at each time instant is minimized. Our contribution is novel in two aspects. First, we have provided a closed-form (and, hence, efficient) solution to the MINVAR problem, using an exponential rate-distortion model in the ρ-domain. Second, we have carried out an extensive comparison between the average quality attained by the MINVAR rate allocation and the MINAVE average distortion. We have found that the coding efficiency loss depends only on a small number of parameters which can be easily extracted from each video sequence. By characterizing these parameters in terms of the Dirichlet distribution, we have shown that, on average, the coding efficiency loss incurred by MINVAR allocation w.r.t. MINAVE is on the order of a fraction of dB, using a PSNR distortion metrics. Since in most systems one wants to reduce not only intersequence variance, but also frame-to-frame quality fluctuations, we have introduced a global encoder buffer to obtain temporal quality smoothing. The two goals are achieved by using classical buffer control techniques together

TABLE I
MINAVE VERSUS MINVAR VERSUS S-MINVAR. THE AVERAGE RATE FOR EACH SEQUENCE IS 1/3 BPP. F = Foreman, H = Hall Monitor, S = Soccer, C = Coastguard

Sequences	$\text{PSNR}_{\text{MINAVE}}$	$\Delta\text{PSNR}_{\text{MINVAR}}$	$\Delta\text{PSNR}_{\text{S-MINVAR}}$	$\Delta\text{PSNR}_{\text{MVS}}$	σ^2_{MINAVE}	σ^2_{MINVAR}	$\sigma^2_{\text{S-MINVAR}}$	σ^2_{MVS}
F-H	38.54	-0.35	-0.48	0.31	26.8	3.54	2.48	6.55
F-S	36.61	-0.14	-0.43	0.27	45.47	6.11	4.56	8.95
F-C	35.19	-0.34	-0.46	-0.03	70.54	8.95	9.13	16.64
H-S	38.24	-0.04	-0.21	-1.83	20.28	10.47	4.61	27.73
H-C	37.61	-0.15	-0.06	-0.13	177.23	16.47	9.69	49.24
S-C	34.20	-0.35	-0.09	0.45	332.42	8.68	17.81	21.00
F-H-S	37.73	-0.29	-0.48	-1.31	35.01	4.90	3.34	8.08
F-H-C	36.47	-0.16	-0.03	0.32	79.57	14.62	19.84	26.24
H-S-C	35.91	-0.13	-0.20	0.57	164.43	11.47	7.01	20.72
F-S-C	35.26	-0.33	-0.35	0.25	111.04	55.47	60.87	75.01
F-S-C-S	36.28	-0.02	-0.14	0.42	90.57	14.90	12.04	14.91

TABLE II
MINAVE VERSUS MINVAR VERSUS S-MINVAR. THE AVERAGE RATE FOR EACH SEQUENCE IS 1/2 BPP. F = Foreman, H = Hall Monitor, S = Soccer, C = Coastguard

Sequences	$\text{PSNR}_{\text{MINAVE}}$	$\Delta\text{PSNR}_{\text{MINVAR}}$	$\Delta\text{PSNR}_{\text{S-MINVAR}}$	$\Delta\text{PSNR}_{\text{MVS}}$	σ^2_{MINAVE}	σ^2_{MINVAR}	$\sigma^2_{\text{S-MINVAR}}$	σ^2_{MVS}
F-H	40.01	-0.12	-0.25	0.01	5.80	0.77	0.52	2.84
F-S	38.73	-0.29	-0.23	0.06	10.48	3.58	4.31	7.86
F-C	36.91	-0.25	-0.21	-0.05	27.18	3.09	2.28	6.41
H-S	39.87	-0.13	-0.20	-0.02	4.71	1.09	0.57	2.98
H-C	37.75	-0.08	-0.15	0.30	51.49	7.45	6.49	15.36
S-C	36.38	-0.04	-0.06	0.08	68.81	2.59	2.43	7.96
F-H-S	39.49	-0.15	-0.32	0.02	6.44	0.89	0.56	3.16
F-H-C	38.18	-0.12	-0.19	0.16	23.60	3.20	2.83	7.25
H-S-C	37.28	-0.09	-0.08	0.01	27.61	2.19	2.16	5.69
F-S-C	37.91	-0.05	-0.07	0.18	36.22	7.33	4.30	8.94
F-S-C-S	38.21	-0.02	-0.11	0.11	57.50	9.71	6.75	15.95

with our MINVAR rate allocation algorithm. As a proof of concept, we have applied the MINVAR and the S-MINVAR algorithms to the case of multiplexed transcoded video sequences. Experimental results with different H.264/AVC video sequences validate the theoretical results.

APPENDIX I
PROOF OF PROPERTY 1

We have to prove that the sequence $\{\text{var}\tilde{\mathbf{D}}^{(n)}\}$ converges to 0 as $n \to \infty$. The constrained minimization problem $\mathcal{P}_n$ (10) can be turned into an unconstrained one using the method of Lagrange multipliers

$$\min_{\mathbf{R}} \left[\sum_{i=1}^{S} \sigma_i^{2n} e^{-n\alpha_i(1-\rho_i)} + \lambda \left(\sum_{i=1}^{S} \theta_i(1-\rho_i) - R \right) \right] \quad (43)$$

which is minimized by the optimal rates shown in (13). Inserting (13) into the R-D model (2) gives the following mix of distortions

$$\tilde{D}_i^{(n)} = \exp\left[\frac{1}{n} \left(\log \xi_i - \frac{\sum_{j=1}^{S} \xi_j \log \xi_j}{\sum_{j=1}^{S} \xi_j} \right) + \frac{\sum_{j=1}^{S} \xi_j \log \sigma_j^2 - R}{\sum_{j=1}^{S} \xi_j} \right]. \quad (44)$$

The sequence of the $\tilde{D}_i^{(n)}$ is a function of n, so we can take the limit as n goes to infinity and get

$$\lim_{n\to\infty} \tilde{D}_i^{(n)} = \exp\left[\frac{\sum_{j=1}^{S} \xi_j \log \sigma_j^2 - R}{\sum_{j=1}^{S} \xi_j} \right] = \tilde{D} \quad (45)$$

which is independent from the index i. Therefore, the sequence of average distortions $\bar{D}^{(n)}$ is

$$\bar{D}^{(n)} = \frac{1}{S} \sum_{i=1}^{S} \tilde{D}_i^{(n)} \overset{n\to\infty}{\longrightarrow} \tilde{D} \quad (46)$$

and the sequence of variances

$$\text{var}\tilde{\mathbf{D}}^{(n)} = \frac{1}{S} \sum_{i=1}^{S} \left(\tilde{D}_i^{(n)} - \bar{D}^{(n)} \right)^2 \quad (47)$$

converges to 0 as $n \to \infty$.

APPENDIX II
PROOF OF PROPERTY 2

We have to show that

$$\frac{D^*}{\tilde{D}} = \frac{e^{H(\zeta)}}{S}. \quad (48)$$

The value of $\tilde{D}$ has already been determined in (45). In order to find D^*, we seek for the solution of the MINAVE problem (6) or, equivalently, the solution of $\mathcal{P}_1$ in (10). Since the average

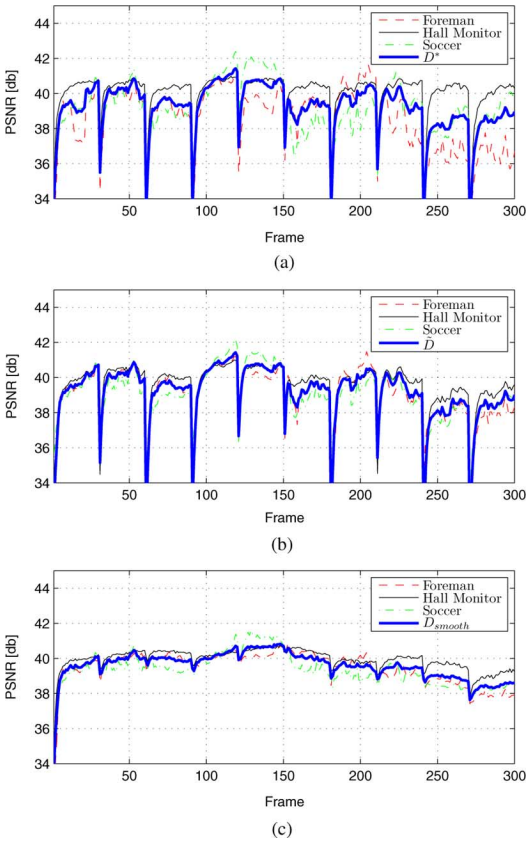


Fig. 12. MINAVE versus MINVAR versus S-MINVAR optimization applied to the transcoding of 3 H.264/AVC coded video sequences. One every 30 frames is intracoded. In the S-MINVAR case, a 1-s buffer has been used. (a) MINAVE. (b) MINVAR. (c) S-MINVAR.

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 setting $n = 1$ in (44)

$$D^* = \frac{1}{S}\sum_{i=1}^{S}\tilde{D}_i^{(1)} = \frac{1}{S}\sum_{i=1}^{S}\exp\left[\log\xi_i + \frac{\sum_{j=1}^{S}\xi_j\log\left(\frac{\sigma_j^2}{\xi_j}\right) - R}{\sum_{j=1}^{S}\xi_j}\right]. \quad (49)$$

To evaluate the performance of the minimum variance distortion, we analyze the ratio

$$\frac{D^*}{\tilde{D}} = \mathcal{E} \quad (50)$$

where

$$\mathcal{E} = \frac{1}{S}\exp\left(\frac{-\sum_{i=1}^{S}\xi_i\log\xi_i}{\sum_{i=1}^{S}\xi_i}\right)\sum_{i=1}^{S}\xi_i. \quad (51)$$

Let $\zeta_i = \frac{\xi_i}{\sum_{i=1}^{S}\xi_i}$, $0 \le \zeta_i \le 1$, be the normalized values of ξ_i and $H(\zeta) = -\sum_{i=1}^{S}\zeta_i\log\zeta_i$ the entropy function of a discrete memoryless source having the set $\zeta_i, i = 1\ldots S$, as the probability mass function of its S symbols. The expression of $\mathcal{E}$ can be rewritten in a more compact form

$$\begin{aligned}\mathcal{E} &= \frac{1}{S}\sum_{i=1}^{S}\xi_i\cdot\exp\left[-\sum_{i=1}^{S}\zeta_i\log\left(\zeta_i\sum_{j=1}^{S}\xi_j\right)\right]\\ &= \frac{1}{S}\sum_{i=1}^{S}\xi_i\cdot\exp\left[-\sum_{i=1}^{S}\zeta_i\log\zeta_i - \sum_{i=1}^{S}\zeta_i\log\left(\sum_{j=1}^{S}\xi_j\right)\right]\\ &= \frac{1}{S}\sum_{i=1}^{S}\xi_i\cdot\exp\left[H(\zeta) - \left(\log\sum_{i=1}^{S}\xi_i\right)\left(\sum_{j=1}^{S}\zeta_j\right)\right]\\ &= \frac{1}{S}e^{H(\zeta)}\sum_{i=1}^{S}\xi_i\cdot\exp\left[-\log\left(\sum_{j=1}^{S}\xi_j\right)\right]\\ &= \frac{1}{S}e^{H(\zeta)} \end{aligned} \quad (52)$$

from which (48) follows.

Appendix III
Derivation of the Entropy Expectation

To simplify calculations, we first apply a nonlinear coordinate transformation $\eta_i = \sqrt{\zeta_i}$. In such a way, the parameters ζ_i are mapped from the unit simplex to the first hyperoctant of the unit hypersphere $\sum_{i=1}^{S}\eta_i^2 = 1$ (which is commonly referred to as Bhattacharyya's statistical space [31]). We can express the η_is with the polar coordinates $(r, \phi_1, \phi_2, \ldots\phi_{S-1})$, with $r = 1$ and

$$\begin{aligned}\eta_S &= \cos\phi_{S-1} \quad (53)\\ \eta_{S-1} &= \sin\phi_{S-1}\cos\phi_{S-2}\\ \eta_{S-2} &= \sin\phi_{S-1}\sin\phi_{S-2}\cos\phi_{S-3}\\ &\vdots \qquad\qquad \vdots\\ \eta_3 &= \sin\phi_{S-1}\sin\phi_{S-2}\sin\phi_{S-3}\ldots\cos\phi_2\\ \eta_2 &= \sin\phi_{S-1}\sin\phi_{S-2}\sin\phi_{S-3}\ldots\sin\phi_2\cos\phi_1\\ \eta_1 &= \sin\phi_{S-1}\sin\phi_{S-2}\sin\phi_{S-3}\ldots\sin\phi_2\sin\phi_1. \quad (54)\end{aligned}$$

The expected entropy $E[H(\zeta)]$ is obtained by taking the expectation of the entropy function over the S-dimensional Bhattacharyya hyperoctant surface Ω_S

$$E[H(\zeta)] = \int_1\int_2\ldots\int_{S-1}H(\zeta)\mu d\Omega_S \quad (55)$$

where $\int_i$ is a short-hand notation for $\int_{\phi_i=0}^{\phi_i=\pi/2}$ and $\mu = \mu(\phi_1, \phi_2, \ldots\phi_{S-1})$ is a probability density function over the Bhattacharyya space. The infinitesimal surface area $d\Omega_S$ is given by [31], [32]

$$d\Omega_S = \sin\phi_2\sin^2\phi_3\sin^3\phi_4\cdots\sin^{S-2}\phi_{S-1}d\phi_1 d\phi_2\ldots d\phi_{S-1}. \quad (56)$$

The relation between $\mu = \mu(\phi_1,\ldots\phi_{S-1})$ defined over the Bhattacharyya space, and a generic distribution

$\nu = \nu(\zeta_1, \dots \zeta_S)$ defined over the S-dimensional simplex is given in [31]

$$\mu = 2^{S-1}\sqrt{S}\prod_{i=1}^{S}\eta_i\nu. \tag{57}$$

In the case of $\nu = \mathrm{Dir}(\zeta_1, \zeta_2, \dots, \zeta_S; a_1, a_2, \dots, a_S)$, we get

$$\mu = \frac{1}{B(\mathbf{a})}2^{S-1}\prod_{i=1}^{S}\eta_i^{2a_i-1} \tag{58}$$

where we have used the definition of Dirichlet distribution (29) in (57), and the nonlinear change of variables $\eta_i = \sqrt{\zeta_i}$. We add a normalizing constant $\sqrt{S}$ to obtain unit area on the hyperoctant. The expected entropy in (55) then becomes

$$\begin{aligned} E[H(\zeta)] &= -\frac{1}{B(\mathbf{a})}2^{S-1}\int_1\int_2\cdots\int_{S-1}\prod_{j=1}^{S}\eta_j^{2a_j-1} \\ &\quad\cdot\sum_{i=1}^{S}\zeta_i\log(\zeta_i)d\Omega_S \\ &= -\frac{1}{B(\mathbf{a})}2^{S-1}\sum_{i=1}^{S}\int_1\int_2\cdots\int_{S-1}\prod_{j=1}^{S}\eta_j^{2a_j-1} \\ &\quad\times\zeta_i\log(\zeta_i)d\Omega_S \end{aligned} \tag{59}$$

which can be solved in closed form, as illustrated in the following.

Because of the S-fold symmetry of the integration space, we can exchange the positions of the variables η_i in (59) so that only one type of integral has to be solved. Let $\mathcal{T}_{i,S}(\mathbf{a}), i = 1, \dots S$, denote the ith transposition of the parameter vector $\mathbf{a} = [a_1, a_2, \dots a_S]^T$, which is obtained by swapping element a_i with the last element a_S. There are S possible transpositions of $\mathbf{a}$

$$\begin{aligned} \mathcal{T}_{S,S}(\mathbf{a}) &= \{a_1, a_2, a_3, \dots a_{S-2}, a_{S-1}, a_S\} \\ \mathcal{T}_{S-1,S}(\mathbf{a}) &= \{a_1, a_2, a_3, \dots a_{S-2}, a_S, a_{S-1}\} \\ \mathcal{T}_{S-2,S}(\mathbf{a}) &= \{a_1, a_2, a_3, \dots a_S, a_{S-1}, a_{S-2}\} \\ \vdots \quad &= \qquad \vdots \\ \mathcal{T}_{1,S}(\mathbf{a}) &= \{a_S, a_2, a_3, \dots a_{S-2}, a_{S-1}, a_1\}. \end{aligned}$$

Using S-fold symmetry, (59) becomes

$$\begin{aligned} &E[H(\zeta)] \\ &\quad= -\frac{1}{B(\mathbf{a})}2^{S-1}\sum_{\mathcal{T}_{i,S}(\mathbf{a})}\int_1\int_2\cdots\int_{S-1}\prod_{j=1}^{S}\eta_j^{2a_j-1} \\ &\quad\cdot\zeta_S\log(\zeta_S)d\Omega_S \end{aligned} \tag{60}$$

where the notation $\sum_{\mathcal{T}_{i,S}(\mathbf{a})}$ denotes the sum over all the transpositions of the vector of parameters $\mathbf{a}$. Therefore, we can solve S integrals of the form

$$\begin{aligned} &E[H(\zeta_S)] \\ &\quad= -\frac{1}{B(\mathbf{a})}2^{S-1}\int_1\int_2\cdots\int_{S-1}\prod_{j=1}^{S}\eta_j^{2\tilde{a}_j-1} \\ &\quad\zeta_S\log(\zeta_S)d\Omega_S \end{aligned} \tag{61}$$

where we have used $\tilde{a}_j$ to denote the jth element of $\mathcal{T}_{i,S}(\mathbf{a})$. Using the spherical coordinate system (53), we observe that

$$\begin{aligned} &\prod_{j=1}^{S}\eta_j^{2a_j-1}d\Omega_S \\ &\quad= \prod_{j=1}^{S-1}\cos^{(2a_{j+1}-1)}\phi_j \\ &\qquad\cdot\sin^{(-1+2\sum_{k=1}^{j}a_k)}\phi_j d\phi_j \end{aligned} \tag{62}$$

and, thus, (61) becomes

$$\begin{aligned} &E[H(\zeta)] \\ &\quad= -\frac{1}{B(\mathbf{a})}2^{S-1}\int_1\cos^{2a_2-1}\phi_1\sin^{2a_1}\phi_1 d\phi_1 \\ &\quad\cdot\int_2\cos^{2a_3-1}\phi_1\sin^{2(a_1+a_2)-1}\phi_2 d\phi_2\dots \\ &\quad\cdot\int_{S-1}\cos^{2a_S+1}\phi_{S-1} \\ &\quad\cdot\sin^{(-1+2\sum_{k=1}^{S-1}a_k)}\phi_{S-1} \\ &\quad\cdot\log(\cos^2\phi_{S-1})d\phi_{S-1}. \end{aligned} \tag{63}$$

This expression can be rewritten in a more compact form

$$\begin{aligned} &E[H(\zeta_S)] \\ &\quad= -\frac{1}{B(\mathbf{a})}2^{S-1}\cdot\mathrm{L}\left(2\sum_{k=1}^{S-1}a_k - 1, 2a_S+1\right) \\ &\qquad\cdot\prod_{j=1}^{S-2}\mathrm{M}\left(2\sum_{k=1}^{j}a_k - 1, 2a_{j+1}-1\right) \end{aligned} \tag{64}$$

where M and L are defined as follows:

$$\begin{aligned} \mathrm{M}(n,m) &= \int_0^{\pi/2}\sin^n x\cos^m x dx \\ &= \frac{1}{2}B\left(\frac{m+1}{2}, \frac{n+1}{2}\right) \end{aligned} \tag{65}$$

$$\begin{aligned} \mathrm{L}(n,m) &= \int_0^{\pi/2}\sin^n x\cos^m x\log(\cos^2 x)dx \\ &= \frac{1}{2}B\left(\frac{m+1}{2}, \frac{n+1}{2}\right) \\ &\quad\cdot\left[\psi\left(\frac{m+1}{2}\right) - \psi\left(\frac{m+n}{2}+1\right)\right] \end{aligned} \tag{66}$$

and ψ is the digamma function. Putting the previous results into (64) and recalling the definition of the beta function (30) and the identity $\Gamma(x+1) = x\Gamma(x)$, one obtains

$$\begin{aligned} E[H(\zeta_S)] &= -\frac{\Gamma(a_0)}{\prod_{j=1}^{S}\Gamma(a_j)} \\ &\quad \cdot \prod_{j=1}^{S-2} \frac{\Gamma(a_{j+1})\Gamma\left(\sum_{k=1}^{j} a_k\right)}{\Gamma\left(\sum_{k=1}^{j+1} a_k\right)} \\ &\quad \cdot \frac{\Gamma\left(\sum_{k=1}^{S-1} a_k\right)\Gamma(a_S+1)}{\Gamma(a_0+1)} \\ &\quad \cdot [\psi(a_S+1) - \psi(a_0+1)] \\ &= -\frac{a_S}{a_0}[\psi(a_S+1) - \psi(a_0+1)] \end{aligned} \tag{67}$$

where $a_0 = \sum_{i=1}^{S} a_i$. By virtue of the S-fold symmetry of (60), we can combine the a_i's to obtain the overall expected entropy

$$E[H(\zeta)] = -\sum_{i=1}^{S} \frac{a_i}{a_0}(\psi(a_i+1) - \psi(a_0+1)) \tag{68}$$

from which (31) follows.

REFERENCES

[1] *Information Technology—Coding of Audio-visual Objects—Part 10: Advanced Video Coding*, ISO/IEC International Standard 14496-10:2003, May 2003, ITU-T.

[2] Z. Chen and K. N. Ngan, "Recent advances in rate control for video coding," *Signal Process.: Image Commun.*, vol. 22, pp. 19–38, Jan. 2007.

[3] Z. Chen and K. Ngan, "Distortion variation minimization in real-time video coding," *Signal Process.: Image Commun.*, vol. 21, pp. 273–279, Jan. 2006.

[4] L. J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 8, pp. 446–459, Aug. 1998.

[5] J. R. Ohm, Introduction to MPEG-4 Video (Rectangular) Jul. 2005, ISO/IEC JTC1/SC29/WG11, MPEG2005/N7297.

[6] A. Vetro, H. Sun, and Y. Wang, "MPEG-4 rate control for multiple video objects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 2, pp. 186–199, Feb. 1999.

[7] J. I. Ronda, M. Eckert, F. Jaureguizar, and N. Garcia, "Rate control and bit allocation for MPEG-4," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 12, pp. 1243–1258, Dec. 1999.

[8] Y. Sun and I. Ahmad, "A robust and adaptive rate control algorithm for object-based video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 10, pp. 1167–1182, Oct. 2004.

[9] P. Nunes and F. Pereira, "Scene level rate control algorithm for MPEG-4 video coding," *Proc. SPIE Vis. Commun. Image Process.*, vol. 4310, 2001.

[10] A. Y. Ngai, L. Brczky, and E. F. Westermann, "Statistical multiplexing using MPEG-2 video encoders," *IBM J. Res. Develop.*, vol. 43, pp. 511–520, July 1999.

[11] L. Wang and A. Vincent, "Joint rate control for multi-program video coding," *IEEE Trans. Consum. Electron.*, vol. 42, no. 3, pp. 300–305, Mar. 1996.

[12] M. Balakrishan and R. Cohen, "Global optimization of multiplexed video encoders," in *Proc. Int. Conf. Image Processing*, Santa Barbara, CA, Oct. 1997, pp. 377–380.

[13] 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. Multimedia and Expo*, 2000, pp. 1309–1312.

[14] 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, no. 11, pp. 970–982, Nov. 2002.

[15] 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, no. 11, pp. 840–849, Nov. 2002.

[16] Z. He and D. Wu, "Look-ahead processing and joint rate control for multiple JVT video encoders," presented at the IEEE Int. Packet Video Workshop, Dec. 2004.

[17] 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, May 2005.

[18] P. Nunes, G. Pastuszak, A. Pietrasiewicz, and F. Pereira, "Joint bit-allocation for multi-sequence H.264/AVC video coding rate control," presented at the Proc. Picture Coding Symp., Nov. 2007.

[19] J. Ribas-Corbera, P. A. Chou, and S. L. Regunathan, "A generalized hypothetical reference decoder for H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 674–687, Jul. 2003.

[20] A. Vetro, C. Christopoulos, and H. Sun, "Rate-distortion optimization for video compression," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 18–29, Mar. 2003.

[21] X. K. Yang and N. Ling, "Statistical multiplexing based on mpeg-4 fine granularity scalability coding," *Int. J. VLSI Signal Process. Syst.*, vol. 42, pp. 69–77, Jan. 2006.

[22] S. Dumitrescu and X. Wu, "Optimal variable rate multiplexing of scalable code streams," presented at the IEEE Data Compression Conf., Snowbird, UT, Mar. 2003.

[23] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000.

[24] *Video Coding for Low Bitrate Communication*, ITU-T Recommendation H.263, Version 1, Nov. 1995, ITU-T.

[25] Z. He and T. Chen, "Linear rate control for JVT video coding," presented at the IEEE Int. Conf. Information Technology: Research and Education, Newark, NJ, Aug. 2003.

[26] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multi-resolution and MPEG video coding," *IEEE Trans. Image Process.*, vol. 3, no. 9, pp. 533–545, Sep. 1994.

[27] G. Valenzise, M. Tagliasacchi, and S. Tubaro, "A smoothed, minimum distortion-variance rate control algorithm for multiplexed transcoded video sequences," in *Proc. Int. Workshop on Mobile Video*, Sep. 2007, pp. 55–60.

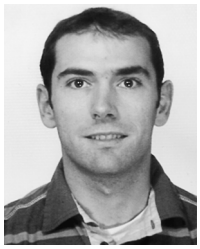
[28] G. Valenzise, M. Tagliasacchi, S. Tubaro, and L. Piccarreta, "A ρ-domain rate controller for multiplexed video sequences," presented at the Picture Coding Symp., Nov. 2007.

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

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

[31] A. De Vos, "The entropy of a mixture of probability distributions," *Entropy*, vol. 7, p. 15, 2005.

[32] M. G. Kendall, *A Course in the Geometry of N Dimensions*. London, U.K.: Charles Griffin, 1961.

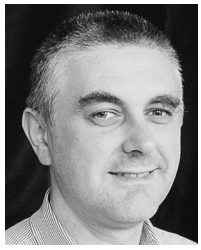


Marco Tagliasacchi (M'06) was born in 1978. He received the Laurea degree (cum Laude) in computer engineering and the Ph.D. degree in electrical engineering and computer science from the Politecnico di Milano, Milano, Italy, in 2002 and 2006, respectively.

He is currently an Assistant Professor at the Dipartimento di Elettronica e Informazione, Politecnico di Milano. He has authored more than 40 scientific papers in international journals and conferences. His research interests include distributed video coding, scalable video coding, non-normative tools in video coding standards, applications of compressive sensing, robust classification of acoustic events, and localization of acoustic sources through microphone arrays.

Giuseppe Valenzise (S'07) was born in Monza, Italy, in 1982. He received a double M.S. degree with honors in computer engineering from the Politecnico di Milano, Milano, Italy, and the Politecnico di Torino, Torino, Italy, in April 2007, and the Diploma from the Alta Scuola Politecnica, Italy, in June 2007. He is currently pursuing the Ph.D. degree in information technology at the Politecnico di Milano.

From May 2007 to December 2007, he was a research assistant in the Image and Sound Processing Group (ISPG), Computational Acoustics and Sound Engineering Lab, Como, Italy. His research interests span different fields of signal processing, such as resource-constrained signal processing, applications of compressive sensing, microphone array processing, and robust classification of acoustic events.



Stefano Tubaro (M'01) was born in Novara, Italy, in 1957. He completed his studies in electronic engineering at the Politecnico di Milano, Milano, Italy, in 1982.

He joined the Dipartimento di Elettronica e Informazione, Politecnico di Milano, first as a researcher of the National Research Council, then as an Associate Professor in November 1991. Since December 2004, he has been a Full Professor. In the early years of his activities, he worked on problems related to speech analysis, motion estimation/compensation for video analysis/coding, and vector quantization applied to hybrid video coding. Over the past few years, his research interests have focused on image and video analysis for the geometric and radiometric modeling of 3-D scenes and advanced algorithms for video coding and sound processing. He has authored more than 150 scientific publications in international journals and congresses. He coauthored two books on the digital processing of video sequences. He is also a coauthor of several patents relative to image processing techniques. He coordinates the research activities of the Image and Sound Processing Group (ISPG), Dipartimento di Elettronica e Informazione, Politecnico di Milano, which is involved in several research programs funded by industrial partners, the Italian Government, and the European Commission.