

A Smoothed, Minimum Distortion-Variance Rate Control Algorithm for Multiplexed Transcoded Video Sequences *

Giuseppe Valenzise
valenzise@elet.polimi.it

Marco Tagliasacchi
tagliasa@elet.polimi.it

Stefano Tubaro
tubaro@elet.polimi.it

Dipartimento di Elettronica e Informazione, Politecnico di Milano
P.za Leonardo da Vinci 32, 20133 Milano, Italy

ABSTRACT

This paper considers the problem of transcoding multiple video sequences in order to enable transmission across a channel characterized by a limited bandwidth. At each time instant, the available bit budget needs to be optimally allocated to the sequences in order to minimize the output average distortion. Nevertheless, this criterion does not guarantee that the quality of the reconstructed sequences is the same. Therefore, we seek for a solution that minimizes the output distortion variance, and we show that we achieve constant quality while keeping the penalty in average distortion small. We formulate this problem in the ρ -domain, and we show that the model parameters needed to solve the optimization problem can be easily extracted from the incoming compressed bitstreams. In addition, if a global video buffer can be accommodated, we show that it is possible to smooth the overall video quality along time. Experimental results on H.264/AVC compressed video data validate the performance of the proposed algorithm.

Categories and Subject Descriptors: I.4.2 [Computing Methodologies]: Image Processing and Computer Vision; E.4 [Data]: Coding and Information Theory

General Terms: Algorithms, Performance.

1. INTRODUCTION

In several applications, from digital TV broadcast to video surveillance, there is the need of transmitting simultaneously several video sequences over a bandwidth-limited channel. Therefore, the available bandwidth has to be distributed across the sequences according to some optimality criterion. This problem has been addressed in the literature under the name of statistical multiplexing [2][7][1], where the goal is to maximize the average output quality, or to achieve constant quality. It can be easily demonstrated that simple equal partitioning of the available bandwidth among the sequences

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

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MV'07, September 28, 2007, Augsburg, Bavaria, Germany.
Copyright 2007 ACM 978-1-59593-779-7/07/0009 ...\$5.00.

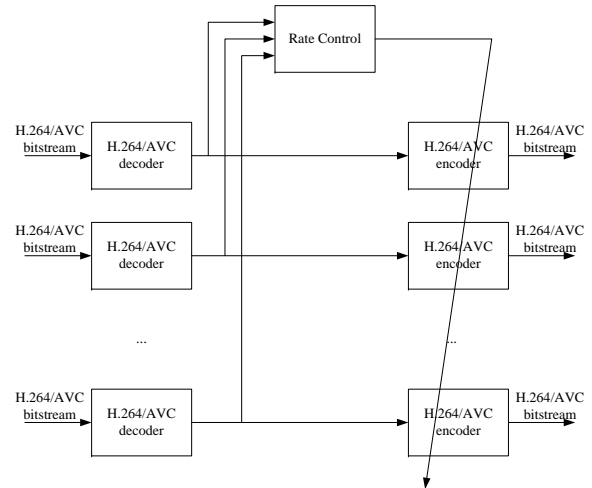


Figure 1: Block diagram of the proposed transcoding/multiplexing architecture.

(also called programs in this context) is sub-optimal, in the sense that it does not take advantage of the diversity in video complexity.

Most of the literature on statistical multiplexing refers to encoding of MPEG-2 video. In [7] 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 [2]. These works formulate the problem in the q -domain, by using simple models to relate the quantization step size q to the rate, i.e. $R(q)$. In addition, they make the implicit assumption that by keeping the quantization parameter equal, constant quality is guaranteed. This is in general not true, as it can be demonstrated by adopting more accurate rate-distortion models. For example, in [3] 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 [4] an optimal rate-allocation for MPEG-4 Visual video objects is proposed, with the goal of minimizing the average distortion.

In this paper we consider an application scenario that consists of jointly transcoding/multiplexing H.264/AVC encoded video streams. Figure 1 depicts the block diagram of the

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

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

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

2. OVERVIEW OF THE ρ -DOMAIN MODEL

In [4], it is shown that in any typical transform domain system, there is always a linear relationship between the coding bitrate R and the percentage of zeros among the quantized transform coefficients, denoted by ρ , i.e.:

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

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

If the transform coefficients have a positive distribution, there is a one-to-one mapping between the quantization parameter q and ρ . Therefore any function in the q domain can be mapped into the ρ domain and vice-versa. Let us consider an encoder that adopts a uniform quantizer with dead zone. Let q denote step size and $\Delta = (1 - b)q$, $b \in [0, 1/2]$ the dead zone threshold. The parameter b adjusts the width of the dead zone. Let $\mathcal{D}(x)$ be the distribution of the (un-quantized) DCT coefficients. For any q , the corresponding percentage of zeros can be obtained as follows:

$$\rho(q) = \frac{1}{N} \sum_{|x| < (1-b)q} \mathcal{D}(x) \quad (2)$$

where N is the number of coefficients in the current video frame.

It is widely recognized in the literature that the distribution of DCT coefficients of prediction residuals can be recognized as Laplacian [6]. In [4] 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 [4]:

$$D(\rho) = \sigma_x^2 e^{-\alpha(1-\rho)}, \quad (3)$$

where σ_x^2 denotes the variance of the DCT coefficients and α is a parameter that can be simply computed at the decoder, as explained in [4]. Equation (2) constitutes the basis of the rate-distortion model used in the rest of this paper.

3. RATE ALLOCATION

3.1 Minimum average distortion

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

$$\min_{\mathbf{R}} \frac{1}{S} \sum_{i=1}^S D_i(R_i), \quad (4)$$

subject to the constraint:

$$\sum_{i=1}^S R_i \leq R. \quad (5)$$

Using equation (3), the minimization problem (4) becomes:

$$\min_{\mathbf{R}} \frac{1}{S} \sum_{i=1}^S \sigma_i^2 e^{-\alpha_i(1-\rho_i)}, \quad (6)$$

subject to the constraint:

$$\sum_{i=1}^S \theta_i(1 - \rho_i) \leq R. \quad (7)$$

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

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

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

3.2 Minimum variance distortion

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

$$\min_{\mathbf{R}} \frac{1}{S} \sum_{i=1}^S (D_i(R_i) - \bar{D})^2, \quad (9)$$

subject to the constraint

$$\sum_{i=1}^S R_i \leq R, \quad (10)$$

or, adopting the ρ -domain model:

$$\min_{\mathbf{R}} \frac{1}{S} \sum_{i=1}^S (D_i(\rho_i) - \bar{D})^2, \quad (11)$$

subject to the constraint

$$\sum_{i=1}^S \theta_i (1 - \rho_i) \leq R, \quad (12)$$

where $\bar{D} = \frac{1}{S} \sum_{i=1}^S D_i(R_i)$ or $\bar{D} = \frac{1}{S} \sum_{i=1}^S D_i(\rho_i)$, depending on the choice of the domain. This problem is difficult to solve in closed form, since \bar{D} depends on the whole set of distortions D_i of each sequence. To overcome this limitation, we first reformulate problem (9)-(12) into a simpler one in order to achieve equal distortion for all the sequences. Then, we evaluate the ‘‘goodness’’ of the obtained distortion value against the minimum average distortion solution.

3.2.1 Equal distortion bit allocation

The first problem we deal with is how to allocate the available rate to each video sequence so that the variance of the output distortions is minimized, subject to rate constraint. Hereafter, we assume that the rate-distortion profile is well approximated by the exponential distortion model (3).

Let $\{\tilde{D}_i^{(n)}\}$, $n = 1, 2, 3, \dots$ be the sequence of distortions of each video program i , found by solving the sequence of minimization problems $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_n, \dots$, where the problem \mathcal{P}_n is described as:

$$\mathcal{P}_n : \min_{\mathbf{R}} \sum_{i=1}^S D_i^n(\rho_i) = \min_{\mathbf{R}} \sum_{i=1}^S \sigma_i^{2n} e^{-n\alpha_i(1-\rho_i)}, \quad (13)$$

subject to the constraint

$$\sum_{i=1}^S \theta_i (1 - \rho_i) \leq R. \quad (14)$$

Then, we can prove the following property:

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

PROOF. The constrained minimization problem \mathcal{P}_n can be easily converted into an unconstrained one using Lagrange multipliers:

$$\min_{\mathbf{R}} \left[\sum_{i=1}^S \sigma_i^{2n} e^{-n\alpha_i(1-\rho_i)} + \lambda \cdot \left(\sum_{i=1}^S \theta_i (1 - \rho_i) - R \right) \right] \quad (15)$$

which is minimized by the following mix of distortions:

$$\tilde{D}_i^{(n)} = \exp \left[\frac{1}{n} \left(\log \xi_i - \frac{\sum_{i=1}^S \xi_i \log \xi_i}{\sum_{i=1}^S \xi_i} \right) + \frac{\sum_{i=1}^S \xi_i \log \sigma_i^2 - R}{\sum_{i=1}^S \xi_i} \right]. \quad (16)$$

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:

$$\lim_{n \rightarrow \infty} \tilde{D}_i^{(n)} = \exp \left[\frac{\sum_{i=1}^S \xi_i \log \sigma_i^2 - R}{\sum_{i=1}^S \xi_i} \right] = \tilde{D}, \quad (17)$$

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

$$\bar{D}^{(n)} = \frac{1}{S} \sum_{i=1}^S \tilde{D}_i^{(n)} \xrightarrow{n \rightarrow \infty} \tilde{D} \quad (18)$$

and the sequence of variances

$$\text{var } \tilde{D}_i^{(n)} = \frac{1}{S} \sum_{i=1}^S \left(\tilde{D}_i^{(n)} - \bar{D}^{(n)} \right)^2 \quad (19)$$

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

Property 1 provides an explicit way for allocating the bit budget R among the video sequences. In fact, it is possible to compute the limit of the sequence of rates $\tilde{R}_i^{(1)}, \tilde{R}_i^{(2)}, \dots$ which minimize the problems $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_n, \dots$ as follows:

$$\begin{aligned} \tilde{R}_i &= \lim_{n \rightarrow \infty} \tilde{R}_i^{(n)} \\ &= \lim_{n \rightarrow \infty} \xi_i \log \sigma_i^2 + \frac{\xi_i R}{\sum_{i=1}^S \xi_i} - \frac{\xi_i \sum_{i=1}^S \xi_i \log \sigma_i^2}{\sum_{i=1}^S \xi_i} + \\ &\quad + \frac{1}{n} \left[\frac{\xi_i \sum_{i=1}^S \xi_i \log \xi_i}{\sum_{i=1}^S \xi_i} - \xi_i \log \xi_i \right] \\ &= \xi_i \log \sigma_i^2 + \frac{\xi_i R}{\sum_{i=1}^S \xi_i} - \frac{\xi_i \sum_{i=1}^S \xi_i \log \sigma_i^2}{\sum_{i=1}^S \xi_i}. \end{aligned} \quad (20)$$

This result allows us to allocate very efficiently the bit budgets R_i so that all the video programs have the same distortion level \tilde{D} .

3.2.2 Evaluation against minimum average distortion

Once the minimum variance solution \tilde{D} has been found, we investigate how much this value is close to the solution of (4), i.e. the solution of the minimum average distortion problem (4) or, equivalently, the solution of (13) when $n = 1$. Since the average distortion function with the exponential model (3) 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 (16):

$$\begin{aligned} D^* &= \frac{1}{S} \sum_{i=1}^S \tilde{D}_i^{(1)} \\ &= \frac{1}{S} \sum_{i=1}^S \exp \left[\log \xi_i + \frac{\sum_{i=1}^S \xi_i \log \left(\frac{\sigma_i^2}{\xi_i} \right) - R}{\sum_{i=1}^S \xi_i} \right] \end{aligned} \quad (21)$$

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

$$\begin{aligned} D^* - \tilde{D} &= \frac{1}{S} \sum_{i=1}^S \xi_i \exp \left[\frac{\sum_{i=1}^S \xi_i \log \left(\frac{\sigma_i^2}{\xi_i} \right) - R}{\sum_{i=1}^S \xi_i} \right] \\ &= \tilde{D} \cdot [\mathcal{E} - 1], \end{aligned} \quad (22)$$

where

$$\mathcal{E} = \frac{1}{S} \exp \left(\frac{-\sum_{i=1}^S \xi_i \log \xi_i}{\sum_{i=1}^S \xi_i} \right) \sum_{i=1}^S \xi_i \quad (23)$$

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

PROPERTY 2.

$$D^* \leq \tilde{D} \leq S \cdot D^*, \quad (24)$$

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

$$\begin{aligned} \mathcal{E} &= \frac{1}{S} \sum_{i=1}^S \xi_i \cdot \exp \left[-\sum_{i=1}^S \zeta_i \log \left(\zeta_i \sum_{i=1}^S \xi_i \right) \right] \\ &= \frac{1}{S} \sum_{i=1}^S \xi_i \cdot \exp \left[H(\zeta) - \sum_{i=1}^S \zeta_i \log \left(\sum_{i=1}^S \xi_i \right) \right] \\ &= \frac{1}{S} e^{H(\zeta)} \sum_{i=1}^S \xi_i \cdot \exp \left[-\sum_{i=1}^S \frac{\xi_i}{\sum_{i=1}^S \xi_i} \log \left(\sum_{i=1}^S \xi_i \right) \right] \\ &= \frac{1}{S} e^{H(\zeta)}, \end{aligned} \quad (25)$$

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

$$0 \leq H(\zeta) \leq \log(S). \quad (26)$$

Therefore bounds on \mathcal{E} are:

$$\frac{1}{S} \leq \mathcal{E} \leq 1. \quad (27)$$

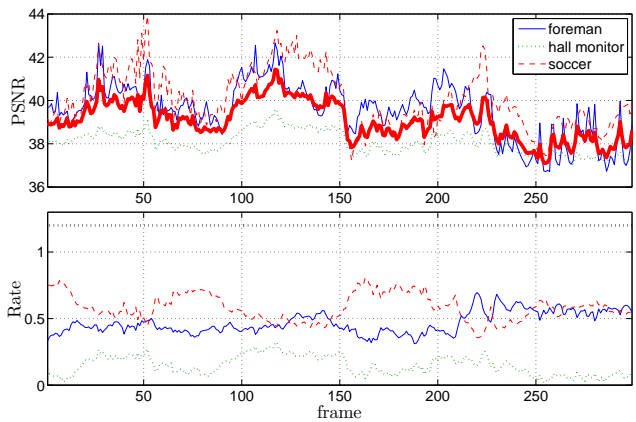
Putting (27) into (22), we get the bounds on the minimum variance distortion $\tilde{D} = D^*/\mathcal{E}$:

$$D^* \leq \tilde{D} \leq S D^*. \quad (28)$$

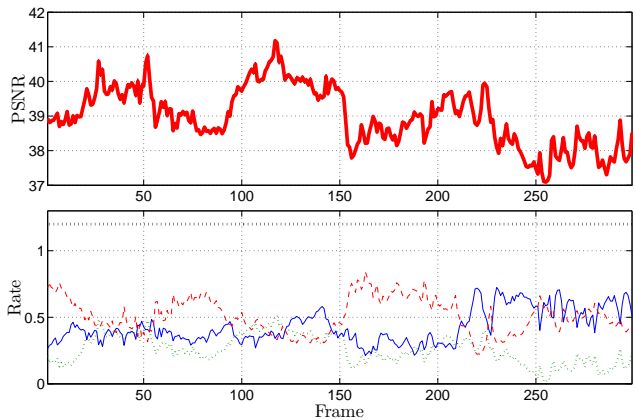
□

In proving Property 2, we have found that the performance loss in terms of average MSE distortion, due to the constraint that all video sequences must have the same visual quality, depends on how the video programs differ to each other. More precisely, the connection between the characteristics of the video sequences and the quality loss is given by the entropy of the normalized parameters ζ_i : when the ζ_i are the same for all sequences, then $\mathcal{E} = 1$ and \tilde{D} reaches the lower bound D^* . This was expected, since allocating bit budget to S identical sequences, so that the average distortion is minimized, leads to assigning the same number of bits to each program, which results in equal distortion for each sequence. Similar arguments justify the loosely tight upper bound in (24). In fact, if we consider the problem of mixing two sequences which considerably differ to each other, it's apparent that achieving the same distortion for both the sequences is sub-optimal in an average distortion sense, when a bit-rate constraint is set.

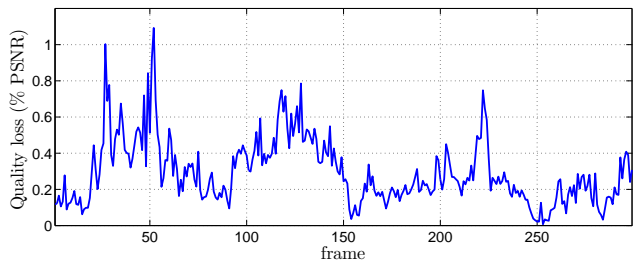
In practice, experimental results have shown that the minimum variance distortion is very close to minimum average distortion for real video sequences, and the loss factor \mathcal{E} is typically greater than 0.99. Figure 3 shows the locations of D^* and \tilde{D} on the average distortion surface for three video



(a) Minimum average distortion. The bold line is the average distortion for the three sequences.



(b) Minimum distortion variance.



(c) Quality loss, expressed in percentage PSNR, between the optimal minimum average distortion and the minimum distortion variance bit allocation strategies.

Figure 2: Distortion and rate profiles using minimum average and minimum variance distortion optimization, for three video programs, using the exponential model. Rates are expressed in bits per pixel.

programs. The average quality loss, in terms of PSNR, due to minimization of variance is of about 0.5dB. This result has been verified for all the frames of the three video sequences in Figure 2(c). The probability of having substantial losses of quality due to variance minimization, as the number of programs is varied, will be a subject of future investigation.

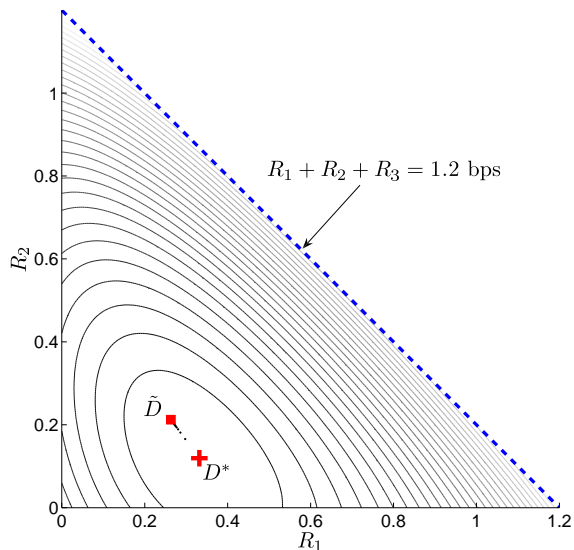


Figure 3: MSE distortion surface vs. rate allocation for three sequences (*Foreman*, *Hall monitor* and *Soccer*), when the total available bit budget is 1.2 bps. Here, D^* is 8.2459 (38.97dB PSNR), while $\tilde{D} = 8.337$ (38.92dB PSNR). The smaller points are the trajectory of the sequence $1/3 \cdot \sum_{i=1}^3 \tilde{D}_i^{(n)}$, for $n = 2 \dots 30$.

4. TEMPORAL SMOOTHING

The strictly constant bit rate (CBR) optimization considered in previous sections allows to achieve optimal distortion (either in the minimum average or minimum variance sense) while satisfying the total rate constraint. However, this can produce distortion profiles with large fluctuations along time (see Figures 2(a)-2(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 in our minimum distortion variance transcoder, we need therefore to add an *encoder buffer* into the system. In this paper we limit to describe the case of a global encoder buffer, which should be placed on the right of the H.264/AVC encoders in Figure 1.

In [5] it is proved that using a geometric averaging filter, it is possible to smooth the optimal minimum distortion profile while achieving the target bit rate on average. Let $\tilde{D}_{\text{CBR}}(t)$ be the minimum variance CBR distortion at frame t ; we define the smoothed distortion target at time t as

$$D_S(t) = \prod_{k=0}^{M-1} \left[\tilde{D}_{\text{CBR}}(t-k) \right]^{\frac{1}{M}}, \quad (29)$$

where M is the length of the averaging window (e.g. $M = 15$ frames). To maintain a temporally-smooth average distortion, we need: 1) to compute the CBR distortion profile; 2) to smooth it using (29); 3) to set $D_S(t)$ as target distortion and find the rates $R_i(t)$ which meet $D_S(t)$ in a minimum dis-

tortion variance sense, for each frame t . Quality smoothing requires therefore that we relax or tighten the rate constraint according to the current buffer level. Let B_{max} be the size of the buffer (which conditions the latency experienced at the decoder); B_0 is the desired buffer level (e.g. $B_0 = 0.5 \cdot B_{\text{max}}$); $b(t)$ denotes instead the buffer fullness at time t ; finally let C be the channel rate, i.e. the rate at which the buffer is drained. The buffer state evolves according to the difference equation:

$$b(t) = b(t-1) + \sum_{i=1}^S R_i(t) - C. \quad (30)$$

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.

4.1 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 straightforward minimization problem:

$$\min_{\mathbf{R}} \sum_{i=1}^S (D_i(t) - D_S(t))^2. \quad (31)$$

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

$$R_i(t) = \xi_i(t) (\log \sigma_i^2(t) - \log D_S(t)). \quad (32)$$

4.2 Constrained buffer draining

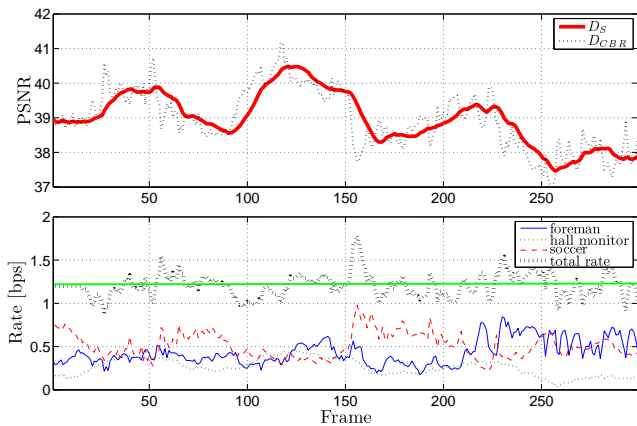
When the number of bits in the buffer exceeds B_0 , the target bit rate R of the CBR distortion profile is reduced to prevent buffer overflow. Let $B_{\text{res}} = b(t) - B_0$; if $B_{\text{res}} > 0$, the encoder needs to reduce the output bits by B_{res} within the next K frames (e.g. K can be set to $0.5M$, as suggested in [5]). Therefore, the new CBR target becomes:

$$R' = R - \frac{B_{\text{res}}}{K}. \quad (33)$$

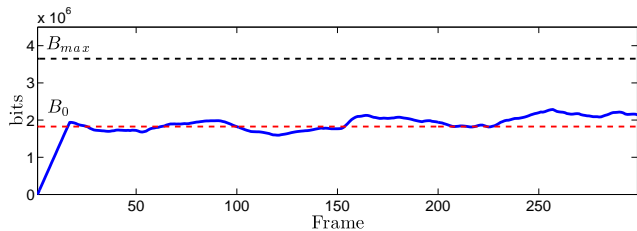
The value of distortion $\tilde{D}_{\text{CBR}}(t)$ is smoothed with (29), and the target distortion $D_S(t)$ is used to find the rates for each sequence with (32). The result of quality smoothing on the three CIF video sequences *Foreman*, *Hall monitor* and *Soccer* is shown in Figure 4. Note that, on average, the total bit rate for the three sequences is equal to the bit rate of the CBR problem (in this example, 1.2 bits per pixel).

5. EXPERIMENTAL RESULTS

We have tested the smoothed minimum distortion variance algorithm against the minimum average distortion bit allocation on *Foreman*, *Hall monitor* and *Soccer* CIF sequences encoded with H.264/AVC baseline profile at 30 fps with a fixed $QP = 15$. The first frame was encoded in INTRA mode, while the rest of the sequence is composed by P slices only. The test procedure consists of extracting the parameters $\{\alpha_i, \theta_i, \sigma_i^2\}$ [4] for each frame of the encoded video sequences, and use them to find the bit rates $R_i(t)$ as described in section 4. The rates R_i computed in the minimum distortion or in the smoothed minimum variance cases are



(a) Smooth-quality bit allocation for three video sequences.



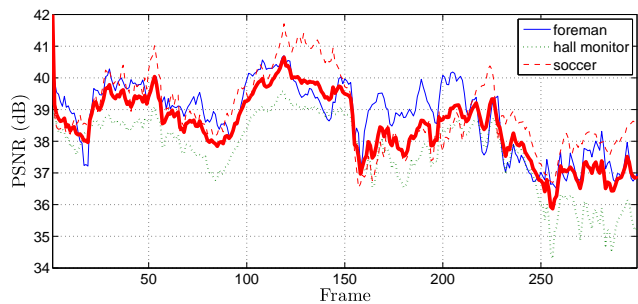
(b) Buffer occupancy. B_{\max} is set to be 1s of delay, corresponding to 30 frames. Initially, the buffer level $b(0)$ is set to 0.

Figure 4: Quality smoothing with 3 CIF video sequences, using the exponential distortion model.

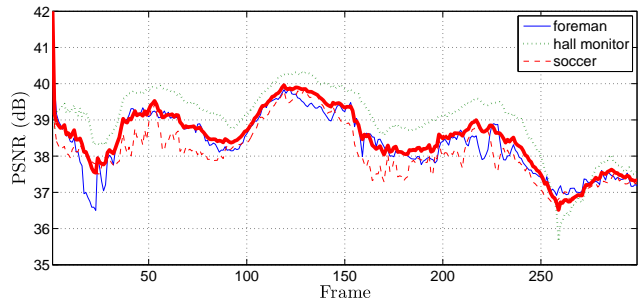
represented in the bottom part of Figures 2(a) and 4(a), respectively. We then use these rates to re-encode the video sequences, using a heuristic macroblock-level rate allocation algorithm, which ensure that each frame is encoded spending the bits indicated by R_i with very high accuracy (the cost of encoding motion vectors has been neglected). The results are presented in Figure 5. The mean standard deviation in PSNR between the three programs in the case of minimum average distortion is 0.85 dB, while it is 0.52 dB in the smooth case: thus, not only the visual quality has been smoothed across time (as it can be clearly noticed from the graph), but also the quality variance between video sequences has been reduced of about 40%.

6. CONCLUSIONS

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



(a) Minimum average distortion.



(b) Minimum distortion variance plus smoothing.

Figure 5: Minimum average distortion vs. smoothed minimum distortion variance optimization applied to the transcoding of 3 H.264/AVC video sequences (in bold the average distortion of the 3 sequences).

7. ACKNOWLEDGEMENTS

We would like to thank Luca Piccarreta for having provided us with essential help and advice in the efficient implementation of a H.264/AVC transcoder.

8. REFERENCES

- [1] M. Balakrishnan and R. Cohen. Global optimization of multiplexed video encoders. In *Proceedings of the 1997 International Conference on Image Processing (ICIP'97)*, volume 1, 1997.
- [2] L. Boroczky, A. Ngai, and E. Westermann. Statistical multiplexing using MPEG-2 video encoders. *IBM Journal of Research Development*, 43(4):511–520, 1999.
- [3] Z. He and S. Mitra. A linear source model and a unified rate control algorithm for DCT video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(11):970–982, 2002.
- [4] Z. He and S. Mitra. Optimum bit allocation and accurate rate control for video coding via ρ -domain source modeling. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(10):840–849, 2002.
- [5] Z. He, W. Zeng, and C. Chen. Low-pass filtering of rate-distortion functions for quality smoothing in real-time video communication. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(8):973–981, 2005.
- [6] E. Lam and J. Goodman. A mathematical analysis of the DCT coefficient distributions for images. *IEEE Transactions on Image Processing*, 9(10):1661–1666, 2000.
- [7] L. Wang and A. Vincent. Joint rate control for multi-program video coding. *IEEE Transactions on Consumer Electronics*, 42(3):300–305, 1996.