Multi-stream Rate Adaptation Using Scalable Video Coding with Medium Grain Scalability

Sergio Cicalò, Abdul Haseeb, and Velio Tralli

Engineering Department of University of Ferrara (ENDIF), Italy
{sergio.cicalo,abdul.haseeb,velio.tralli}@unife.it

Abstract. Multiple video streaming in a shared channel with constant bandwidth requires rate adaptation in order to optimize the overall quality. In this paper we propose a multi-stream rate adaptation framework with reference to the scalable video coding (SVC) extension of the H.264/AVC standard with medium grain scalability (MGS) and quality layer (QL). We first provide a general discrete multi-objective problem formulation with the aim to maximize the sum of assigned rates while minimizing the differences among distortions under a total bit-rate constraint. A single-objective problem formulation is then derived by applying a continuous relaxation to the problem. We also propose a simplified continuous semi-analytical model that accurately estimates the rate-distortion relationship and allows us to derive an optimal and low-complexity procedure to solve the relaxed problem. The numerical results show the goodness of our framework in terms of error gap between the relaxed and its related discrete solutions, the significant performance improvement with respect to an equal-rate adaptation scheme, and the lower complexity with respect to a sub-optimal algorithm proposed in the literature.

Keywords: SVC, MGS, rate-distortion modeling, rate adaptation, quality fairness.

1 Introduction

H.264 Advanced Video coding (AVC) standard with scalable extension, also called Scalable Video Coding (SVC) [1], provides flexibility in rate adaptation by coding an original video sequence into a scalable stream. Three scalability methods are possible in SVC, named temporal, spatial and SNR scalability, that allow to extract a sub-stream in order to meet a particular frame rate, resolution and quality, respectively.

Due to the different complexities of the scenes composing a video sequence, the relationships between the rate and the quality of a set of videos can be really different among them. If individual video streams are transmitted to different users in a broadcast dedicated channel, as for instance in the case of on-demand IPTV services [2], an equal rate allocation can lead to unacceptable distortion

L. Atzori, J. Delgado, and D. Giusto (Eds.): MOBIMEDIA 2011, LNICST 79, pp. 152–167, 2012.

of high-complexity videos with respect to low-complexity ones. Adaptive transmission strategies must be investigated to dynamically optimize the quality of experience (QoE) of each end-user.

In this paper, we focus on rate adaptation, also called in literature statistical multiplexing, of SNR-scalable video streams, with a fixed temporal and spatial resolution. Many contributions exist in the literature that provide rate adaptation exploiting the Fine Granularity Scalability (FGS) tool, e.g. [3],[4] and [5]. FGS coding allows to extract an arbitrary rate-distortion (R-D) point while maintaining the monotonic non-decreasing behavior of the R-D curves. Nevertheless FGS mode has been removed from SVC, due to its complexity.

Two different possibilities for the SNR scalability tool are now available in SVC standard and implemented in the reference software [6], namely Coarse Grain Scalability (CGS) and Medium Grain Scalability (MGS). CGS can be achieved by coding quality refinements of a layer using a spatial ratio equal to 1 and inter-layer prediction. However, CGS scalability can only provide a small discrete set of extractable points equal to the number of coded layers. MGS provides a finer granularity of quality scalability by dividing a CGS layer into up to 16 MGS layers. The granularity can be also improved if a post-processing quality layer (QL) insertion and a consequent quality-based extraction is performed with the aim to optimize the R-D performance [7]. With this tool MGS can be seen as alternative to the FGS coding.

The first aim of this work is to analyze the performance of the MGS with QL and to provide a general R-D model. Other contributions exist in literature that estimate the R-D model for SNR-based scalable stream, with CGS and MGS, e.g. [8],[9], either analytical and semi-analytical. The analytical models are dependent on the probability distribution of discrete cosine transform (DCT) coefficients and often incur in a loss of accuracy. To achieve higher accuracy, semi-analytical R-D models are preferable. The semi-analytical models are based on parametrized functions that follow the shape of analytically derived functions, but are evaluated through curve-fitting from a subset of the rate-distortion empirical data points. In [9], the authors proposed an accurate semi-analytical square-root model for MGS coding and compared it with linear and semi-linear model. They concluded that the best performance is obtained by changing the model according to a parameter that estimates the temporal complexity, evaluated before encoding the entire sequence. However, a general model, that is able to estimate the R-D relationship of a large range of video sequences, is necessary to perform analytical optimization of the rate-adaptation problem. Besides, they did not consider the post-processing QL insertion that produces a variation of the R-D performance.

In [10] the authors proposed a general semi-analytical rate-distortion model for video compression, also verified in [11] for SVC FGS layer, where the rate and the distortion have an inverse relationship. Three sequence-dependent parameters must be estimated through the knowledge of six empirical R-D points. We have also verified this model with reference to SNR scalability with MGS and QL. The high accuracy of the results led us to investigate a simplified model with lower complexity, where the number of R-D points can be reduced by eliminating one of the parameters to estimate. Thus, we propose and compare a simplified two-parameters semi-analytical rate-distortion model. This simplification has two main advantages: (i) only four empirical points are needed by the curve fitting algorithm to achieve good performance, (ii) it allows the derivation of a low-complexity optimal procedure to solve the multi-stream rate-adaptation problem, with a maximum number of iterations equal to the number of streams involved in the optimization.

In summary, this paper collects the following main contributions: in section 2, a general optimization problem is formulated with the aim to provide the maximum quality to each video while minimizing their distortion difference, and by fulfilling the available bandwidth. In section 3 we analyze and verify two similar semi-analytical models for MGS with QL by comparing them with respect to complexity and the normally used goodness parameters: the root mean square error (RMSE) and the coefficient of determination R^2 [12]. An optimum and computationally efficient procedure to solve the relaxed general problem is derived in section 4, with a discussion about complexity and optimality. Finally the numerical results, discussed in section 5, show (i) the goodness of our framework by looking at the error between the relaxed and discrete solutions, (ii) the performance improvement with respect to a blind adaptation, and (iii) the complexity of the proposed algorithm with respect to a sub-optimal golden search algorithm proposed in literature.

2 General Problem Formulation for Multi-stream Rate Adaptation

In general, the aim of multi-stream rate adaptation is to optimize a certain number of utility functions U_i with respect to a quality metric and according to rate constraints [13]. Before or after the encoding process the original high quality video must be adapted, to meet a particular QoE metric depending on spatial, temporal and SNR resolutions.

In this section we provide a general problem formulation for multi-stream rate adaptation. Let K be the number of streams involved in the optimization. Given a set of lossy compression techniques $\{1, \ldots, N_k\}$, we can define in general $\mathcal{D}_k = \{d_{1,k}, ..., d_{N_k,k}\}$, $k = 1, \ldots, K$ as the set of distortion values for the k-th stream. Let us note that its cardinality $|\mathcal{D}_k| = N_k$ is generally not equal for each video source, as in the case of high-flexibility SNR-based compression techniques.

The rate-distortion theory evaluates the minimum bit-rate R_k required to transmit the k-th stream with a given distortion $d_{n,k}$, by defining a function $\mathcal{F}_k$ that maps the distortion to the rate, i.e.

$$\begin{aligned} \mathcal{F}_k : \quad & \mathcal{D}_k \to \mathbb{R}^+ \\ & d_{n,k} \to R_k = \mathcal{F}_k(d_{n,k}) \end{aligned} \tag{1}$$

One of the desirable properties of $\mathcal{F}_k$ is the strictly decreasing monotony, i.e.

$$\mathcal{F}_k(d_{i,k}) > \mathcal{F}_k(d_{j,k}), \quad \forall d_{i,k}, d_{j,k} : d_{i,k} < d_{j,k}. \tag{2}$$

When multiple streams have to be transmitted in a shared channel the rate adaptation algorithm must choose at each time slot and according to one optimization strategy, the best vector $\boldsymbol{D}^* = [D_1^*, ..., D_K^*] \in \boldsymbol{\mathcal{D}} = \mathcal{D}_1 \times \ldots \times \mathcal{D}_K$. $\boldsymbol{\mathcal{D}}$ contains all the possible combinations of the elements of $\mathcal{D}_k$, $k = 1, \ldots, K$ and has cardinality $N = \prod_{k=1}^{K} N_k$.

The main purpose of multi-stream rate adaptation is to provide the minimum distortion, or equivalently the maximum rate according to assumption (2), to each video under a total bit-rate constraints R_c. However, the solution of such problem can generally lead to large distortion variations among different streams, due to the different complexity of video sources. Quality fairness is an important issue that must be addressed when multiple videos from different sources are transmitted in a shared channel. In [4] the authors have shown that, given a continuous decreasing exponential R-D relationship with a constant exponent equal for each source, the solution to the problem of minimizing the distortion variations is also the solution to the problem of minimizing the total average distortion. However, an exponential R-D relationship is not an accurate model for all the different video compression techniques, particularly for the SVC SNR scalable stream [4]. Thus, a general multi-objective problem has to be formulated and a continuous relaxation of the problem leads to some particular simplification under certain assumptions. The general objective of our proposed framework is to minimize the differences among the distortions provided to each video stream while maximizing the sum of the rates until a maximum bit-rate is met. As mentioned above, these two objectives alone can generally lead to different solutions.

Thus, we formulate the general problem as a multi-objective problem:

$$\min_{\boldsymbol{D} \in \boldsymbol{\mathcal{D}}} \sum_{i} \sum_{j<i} \Delta(D_i, D_j) \tag{3}$$

$$\max_{\boldsymbol{D} \in \boldsymbol{\mathcal{D}}} \sum_{k=1}^{K} \mathcal{F}_k(D_k) \tag{4}$$

$$s.t. \quad \sum_{k=1}^{K} \mathcal{F}_k(D_k) \leq R_c \tag{5}$$

where

$$\Delta(D_i, D_j) = \begin{cases} 0 & \text{if } (i,j) \in \mathbb{X}_D \vee (j,i) \in \mathbb{X}_{\mathcal{D}} \\ |D_i - D_j| & \text{otherwise} \end{cases} \tag{6}$$

with

$$\mathbb{X}_{\mathcal{D}} = \{(i,j) \in \mathbb{Z}^2 : (D_i = D_{max,i} \wedge D_j > D_i) \vee (D_i = D_{min,i} \wedge D_j < D_i)\} \tag{7}$$

and $D_{min,i} = \min_n d_{n,i}$, $D_{max,i} = \max_n d_{n,i}$. The operators $\wedge$ and $\vee$ are the logic "AND" and "OR", respectively.

Ideal fairness among the distortion values assigned to the multiple video streams, i.e. $D_i = D_j, \quad \forall i \neq j$, is hard to be achieved. This fact is due to

(i) the discretization of the R-D relationship and (ii) the presence of the minimum and the maximum distortion values for each source that are related to the complexity of each video and which can be very different. The definition of the fairness metric takes this fact into account. In fact, the difference among video distortions $\Delta(D_i, D_j)$ is slightly modified to take into account the minimum and the maximum constraints. It is worth noting that, under the assumption (2), this problem admits a feasible solution only if at least the sum of the minimum rates of the video sequences is supported by the transmission bandwidth R_c, i.e

$$\sum_{k=1}^{K} \mathcal{F}_k(D_{max,k}) \leq R_c \quad (8)$$

otherwise a certain number of videos are not admitted in the transmission until this constraint is not satisfied. The solution of the problem in (3)-(5) requires in general an exhaustive search in the space $\boldsymbol{\mathcal{D}}$ of all possible vectors. If N becomes large the required complexity can be not suitable for real-time adaptation. On the other hand if N is small, i.e there are few video sources as well as few related R-D points, the problem solution can lead to a waste of the available bandwidth and a large distortion differences among multiple videos.

In the next section we will propose a semi-analytical R-D model with reference to the SNR scalability tool of SVC with MGS and QL layers [7]. This continuous model will allow us to apply a continuous relaxation to the optimization problem leading to a simplification in a single-objective problem formulation.

3 Rate Distortion Model for MGS with Quality Layer

We consider here SNR scalability obtained through the MGS coding and QL post-processing insertion, with a fixed temporal and spatial resolution. In this case the components of $\mathcal{D}_k$ are the distortion values of the extractable substreams from the high quality original encoded stream.

MGS coding allows to distribute the transform coefficients obtained from a macro-block by dividing them into multiple sets. The number of sets identifies the number of weights, often named MGS layers, in the MGS vector. Thus, the elements of the MGS vector correspond to the cardinality of each set.

The R-D relationship and its granularity depend on the number of MGS layers and the coefficient distribution [14], [15]. In [15] the authors analyzed the impact on performance of different numbers of MGS layers with different configurations used to distribute the transform coefficients. We also verified their results, by noting that more than five MGS layers reduce the R-D performance without giving a substantial increase in granularity. This is mainly due to the fragmentation overhead that increases with the number of MGS layers.

While extracting an MGS stream two possibilities are available in the reference software: a flat-quality extraction scheme, and a QL-based extraction scheme. The second scheme requires a post-encoding process that computes a priority index for each NAL unit, but achieves higher granularity, as well as better R-D-performance [7]. However, differently to flat-quality extraction scheme, the quality-based extraction process does not give substantial variations in granularity and R-D performance when varying the distribution of the coefficients, as also shown in [15]. In our extensive simulation campaign the best results in terms of granularity and R-D performance are obtained with a MGS vector equal to [3 2 4 2 5].

When the SVC video has to be adaptively transmitted it is common practice to analyze the R-D model with respect to a fixed set of frames identified by one group of pictures (GOP). In this way, the adaptation module can follow the complexity variations of the different scenes. Therefore, throughout this paper we assume that the reference time interval used to analyze the R-D relationship as well as to optimize the distortion of multiple streams is the GOP interval.

In [10] the authors propose a general continuous semi-analytical R-D model for video compression, also verified in [11] for SVC FGS layers, with the following relationship :

$$\mathcal{R}_k(D) = \frac{\eta_k}{D + \theta_k} + \phi_k. \tag{9}$$

The distortion D is evaluated as the average mean square error (MSE) of the decoded video. The drawback of this approach is the need to estimate the three sequence/encoder dependents parameters, η_k, θ_k and ϕ_k, by using curve-fitting from a subset of the rate-distortion data points. The curve-fitting algorithm requires a relevant number of iterations and function evaluations and six empirical R-D points. To reduce the complexity, we can simplify this parametrized model by eliminating one parameter, i.e.

$$\mathcal{R}_k(D) = \frac{\alpha_k}{D} + \beta_k \tag{10}$$

In this case, only four R-D points need to be evaluated to estimate the two sequence-dependent parameters α_k and β_k, and as a result the number of iterations and function evaluations decreases. Beside the complexity reduction, this model allows a simple derivation of the solution of the problem (3)-(5), as we will show later.

Table 1 compares the goodness of the two models with respect the coefficient of determination R^2, the RMSE, the number of iterations and function evaluations required by a non-linear Least Square Trust-Region (LSTR) algorithm to converge. It can be noted how the number of function evaluations as well as the number of iterations decrease while a minimum loss occurs in the goodness parameter. In Figure 1, we plot the empirical R-D relationship for the five sequences, used to obtain numerical results, as well as their related R-D curves based on model (10). All of them are referred to the GOP with the worst RMSE value (the minimum in Table 1). We can also appreciate in this figure the achievable granularity of the quality-based extraction.

In the next section we will apply a continuous relaxation to the problem (3)-(5) by exploiting the model (10) and we will provide a low-complexity optimal procedure to solve it.

Table 1. Comparison between the two semi-analytical model in (9) and (10) with respect to the minimum and maximum $RMSE$ and the coefficient of determination R^2 evaluated for each GOP (GOP size equal to 16) of five video sequence with CIF resolution and frame rate of 30 fps. The video are encoded with one base layer (QP equal to 38) and two enhancement layers (QP equal to 32 and 26), both with 5 MGS layers and a weights vector equal to [3 2 4 2 5].

Video	Model	R^2 [min,max]	$RMSE$ [min,max]	Av. No. iteration	Av. No. Function Evaluation
Coastguard	Model (10)	[0.9842 , 0.9934]	[37.895 , 79.992]	30.2	89.6
	Model (9)	[0.9956 , 0.9982]	[22.261 , 36.724]	34.7	155.9
Crew	Model (10)	[0.9752 , 0.9944]	[23.038 , 89.130]	30.9	94.2
	Model (9)	[0.9914 , 0.9972]	[20.019 , 52.489]	35.6	159.9
Football	Model (10)	[0.9662 , 0.9891]	[53.403 , 205.572]	29.0	89.5
	Model (9)	[0.9809 , 0.9993]	[12.940 , 99.810]	38.0	169.3
Foreman	Model (10)	[0.9669 , 0.9955]	[19.710 , 53.371]	25.7	73.2
	Model (9)	[0.9906 , 0.9980]	[13.516 , 33.745]	34.1	154.3
Harbour	Model (10)	[0.9854 , 0.9907]	[51.860 , 73.344]	37.5	129.8
	Model (9)	[0.9952 , 0.9991]	[18.883 , 44.822]	45.3	164.3

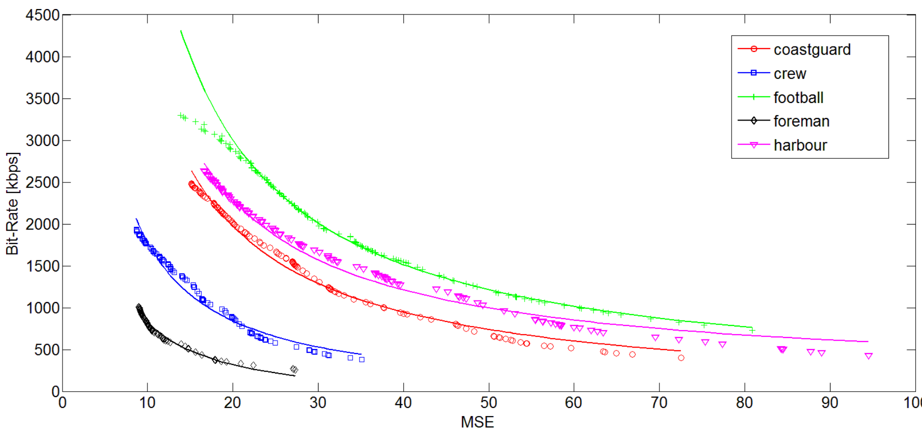

Fig. 1. R-D Model (straight line), according to eq. (10) fitting the empirical R-D relationship for the GOP with the worst $RMSE$ with reference to Table 1

4 GOP-Based Multi-stream Rate Adaptation Framework

Without loosing generality we assume that each video is coded with the same GOP size and the rate allocation is performed at the GOP boundaries. Thus, from now on we focus on one GOP interval. Considering all the discussions in the previous sections, we apply a continuous relaxation to the optimization problem based on the model (10). Therefore we assume that the discrete variable D_k becomes continuous (denoted by $\tilde{D}_k$), but limited by the minimum and maximum distortion, i.e.

$$\tilde{D}_k \in [D_{min,k}, D_{max,k}]. \tag{11}$$

With reference to the SNR scalability, the points $\{D_{max,k}, \mathcal{F}_k(D_{k,max})\}$ and $\{D_{min,k},\ \mathcal{F}_k(D_{k,min})\}$ are the base layer and the highest enhancement layer points, respectively. Those values are two of the four R-D points required by the curve-fitting algorithm.

It is worth noting that a trivial solution can be derived if the sum of the full quality encoded stream rates is less then or equal to the available bandwidth, that corresponds to transmit the entire encoded streams without adaptation. Thus, we analyze the non-trivial case where the following constraint holds :

$$\sum_{k=1}^{K} \mathcal{F}_k(D_{k,min}) > R_c \tag{12}$$

According to the continuous relaxation (11) and the assumptions (8) and (12), a feasible solution is obtained when the constraint on the overall channel bandwidth is active with equality. A single-objective problem, where the second objective, i.e (4) in the problem formulation, is eliminated and replaced by an equality constraints can be then formulated. Nevertheless, as a result of the relaxation of the problem, the two constraints referred to the maximum and minimum available rates of each stream must be added. They imply that each video sequence has to obtain at least the base layer and not more than the maximum available bit-rate must be allocated to each video source to save bandwidth.

Thus, the relaxed problem can be formulated as

$$\min_{\tilde{\boldsymbol{D}} \in \mathbb{R}^{\boldsymbol{K}}} \sum_{i} \sum_{j<i} \varDelta(\tilde{D}_i, \tilde{D}_j) \tag{13}$$

$$s.t. \quad \sum_{k=1}^{K} \mathcal{R}_k(\tilde{D}_k) = R_c \tag{14}$$

$$\mathcal{R}_k(\tilde{D}_k) \geq \mathcal{F}_k(D_{k,max}) \quad \forall k \tag{15}$$

$$\mathcal{R}_k(\tilde{D}_k) \leq \mathcal{F}_k(D_{k,min}) \quad \forall k \tag{16}$$

Note that the model $\mathcal{R}_k(\tilde{D}_k)$ replaces the actual R-D relationship $\mathcal{F}_k(D_k)$. In the next subsection we will derive an optimal procedure to solve this relaxed problem using methods that are computationally efficient and without the use of heuristics or brute-force search.

4.1 Problem Solution

A solution to the relaxed problem (13)-(16) can be derived by using sub-optimal procedures as the golden search algorithm proposed in [3] for a piecewise linear model. Nevertheless, the continuous formulation of model (10) allows us to derive a low-complexity optimal procedure, by noting that the solutions to the problem without the constraints (15) and (16) can be easily derived as follows:

$$\tilde{D}^* = \tilde{D}_k^* = \frac{\sum_{k=1}^{K} \alpha_k}{R_c - \sum_{k=1}^{K} \beta_k}, \quad \forall k. \tag{17}$$

Since those constraints imply that a minimum (maximum) or a maximum (minimum) rate (distortion) has to be allocated to each video stream, these solutions can be improved successively through a simple iterative procedure.

Let $x_k, y_k \in \{0, 1\}$, $k = 1, \ldots, K$, be binary variables that indicate whether or not the two constraints are active for the video stream k and will be updated during the procedure. We can then define:

$$A_{\boldsymbol{x},\boldsymbol{y}} = \sum_{k=1}^{K} x_k y_k \alpha_k \tag{18}$$

$$B_{\boldsymbol{x},\boldsymbol{y}} = \sum_{k=1}^{K} x_k y_k \beta_k \tag{19}$$

$$R^{av}_{\boldsymbol{x},\boldsymbol{y}} = R_c - \sum_{k=1}^{K} (1 - x_k)\mathcal{F}_k(D_{k,max}) - \sum_{k=1}^{K} (1 - y_k)\mathcal{F}_k(D_{k,min}) \tag{20}$$

where $R^{av}_{\boldsymbol{x},\boldsymbol{y}}$ is the available rate for the videos which have not active constraints. The iterative procedure works as follows:

1. Initialize: $x_k = 1$ and $y_k = 1 \quad \forall k = 1, \ldots, K$
2. For each $k : x_k \cdot y_k = 1$ Compute:
   $\tilde{D}_k^* = \frac{A_{\boldsymbol{x},\boldsymbol{y}}}{R^{av}_{\boldsymbol{x},\boldsymbol{y}} - B_{\boldsymbol{x},\boldsymbol{y}}}$
   $\tilde{R}_k^* = \mathcal{R}_k(\tilde{D}_k^*)$, based on model (10)
   $condition = 0$
   2a. If $\tilde{R}_k^* > \mathcal{F}_k(D_{k,min})$ then
      $\tilde{R}_k^* = \mathcal{F}_k(D_{k,min})$
      $\tilde{D}_k^* = D_{k,min}$
      $y_k = 0$
      $condition = 1$
   2b. elseif $\tilde{R}_k^* < \mathcal{F}_k(D_{k,max})$ then
      $\tilde{R}_k^* = \mathcal{R}_k(D_{k,max})$
      $\tilde{D}_k^* = D_{k,max}$
      $x_k = 0$
      $condition = 1$
3. If $condition = 1$
   Go to step 2.
4. else break

The final relaxed solutions, given x_k and y_k, $k = 1, \ldots, K$, are then given by:

$$\tilde{R}_k^* = \begin{cases} \frac{\alpha_k}{\tilde{D}_k^*} + \beta_k & \text{if } x_k \cdot y_k = 1 \\ \mathcal{F}_k(D_{k,max}), & \text{if } x_k = 0 \\ \mathcal{F}_k(D_{k,min}), & \text{if } y_k = 0 \end{cases} \tag{21}$$

with

$$\tilde{D}_k^* = \begin{cases} \frac{A_{\boldsymbol{x},\boldsymbol{y}}}{R^{av}_{\boldsymbol{x},\boldsymbol{y}} - B_{\boldsymbol{x},\boldsymbol{y}}} & \text{if } x_k \cdot y_k = 1 \\ D_{k,max}, & \text{if } x_k = 0 \\ D_{k,min}, & \text{if } y_k = 0 \end{cases} \tag{22}$$

The algorithm requires in the worst case, a maximum of K iterations with $(K-1)/2$ rate and distortion evaluations. At the first iteration, due to the initialization, $\tilde{D}_k^*$ is computed as in (17). At each iteration the algorithm checks if the related rate solutions violate one of the constraints (15), (16). If it happens for one video, the algorithm assigns the relative minimum or maximum rate to this particular video and re-evaluates the distortion for the other video streams.

The optimality of the solutions (21) and (22) can be easily proved, by noting that the sum of the difference functions in (13) is always kept to zero, i.e. $\sum_i \sum_{j<i} \Delta(\tilde{D}_i^*, \tilde{D}_j^*) = 0$ and the sum of the rates is always equal to the available bandwidth. In fact, if at the n-th iteration a maximum rate constraint (condition of step 2a) is violated for the i-th video, the distortion of the other videos at the next iteration, $\tilde{D}_k^*[n+1]$, will decrease, i.e.

$$\tilde{D}_k^*[n+1] < \tilde{D}_k^*[n] < D_{i,min}, \quad \forall k \neq i : x_k[n+1] \cdot y_k[n+1] = 1, y_i[n] = 0. \tag{23}$$

Vice versa, when the second constraint (condition of step 2b) is violated for the j-th video the distortion $\tilde{D}_k^*[n+1]$ of the other video will increase, i.e.

$$\tilde{D}_k^*[n+1] > \tilde{D}_k^*[n] > D_{j,max}, \quad \forall k \neq j : x_k[n+1] \cdot y_k[n+1] = 1, x_j[n] = 0. \tag{24}$$

For all other videos with $x_k \cdot y_k = 1$ the solutions are left untouched, as shown in (22). The inequalities (23) and (24) follow from the monotony property of the R-D function.

Let us finally note that the conditions of steps 2a and 2b are auto-exclusive for each video source if

$$D_{s,max} > D_{p,min}, \quad \forall s \neq p, \quad s, p = 1, \ldots, K \tag{25}$$

When two or more video streams have a very different scene complexity in the same GOP, the inequality (25) may not be verified and the evaluated distortion $\tilde{D}_k^*$ may fall inside the interval $[D_{s,max}, D_{p,min}]$. In this particular case, to assure the best fairness, the algorithm would require some temporary additional steps to evaluate which constraints has to be applied first, which leads to a small increase in the complexity. In order to keep the complexity low we propose for this case to prioritize the distortion minimization. Thus, we first apply the constraints on the maximum rate (step 2a) by assigning the minimum distortion

$D_{p,min}$ to the p-th video. At the next iteration, the distortion will decrease, due to the convexity of the R-D functions. If the distortion decreases in such way that the evaluated rate of the s-th video do not violate its maximum distortion constraint, the algorithm will be able to assign a lower distortion to it. Let us note that this choice does not compromise the optimality of the solution of the problem according to eq. (6).

From a mathematical perspective the optimal discrete solution $\boldsymbol{D}^*$, starting from the relaxed one $\tilde{\boldsymbol{D}}^*$, should be derived by applying optimization techniques, e.g. branch & bound search. Nevertheless, such techniques require the knowledge of all the empirical discrete R-D points or a subset of R-D points close to the relaxed optimum solutions, with an increase in complexity. To keep the complexity low, it is common practice to extract the higher discrete bit-rate under the optimal relaxed solution, by paying a minimum waste of bandwidth due to the granularity of the empirical R-D relationship.

5 Numerical Results

In this section we evaluate the performance of the proposed rate adaptation framework by using the JSVM reference software [6]. We encode five video sequences with different scene complexity, i.e. *coastguard*, *crew*, *football*, *foreman*, *harbour* in CIF resolution with a frame rate of 30 fps. The SNR-scalability is obtained through 2 enhancement layers, each one split in 5 MGS layers with vector distribution [3 2 4 2 5]. The quantization parameter (QP) of the base and enhancement layers are equally spaced and set to 38, 32 and 26, respectively. Each sequence is coded GOP-by-GOP with a GOP size equal to 16, and the post-processing quality-based process is then applied, as mentioned throughout the paper. We first provide the performance metrics for a particular case of bandwidth, i.e. $R_c = 3000$ kbps, then we study the impact of different R_c values. The fairness is evaluated through two metrics: the average MSE difference $\delta_{av} = (1/S)\sum_i \sum_{j<i} |D_i^* - D_j^*|$, where the average is computed with respect to the number $S = K(K-1)/2$ of terms in the sum, and the most used MSE variance for each GOP. We first compare the solution of our algorithm (OPT) with an equal-rate (ER) scheme where no adaptation is performed, i.e. the same proportion of the available bandwidth is assigned to each video. To have a fair comparison we apply to ER scheme the constraints (15) and (16) in order to guarantee the base-layer to each video and to fulfill the available bandwidth. Therefore, after sorting the streams in two vectors into decreasing order according to base-layer bit-rate and into increasing order according to highest layer bit-rate, respectively, we iteratively check if the bit-rate $R_k = R_c/K$ required by each ordered stream violates one of those constraints. If it happens, we assign the corresponding bit-rate and equally re-distribute the remaining bandwidth to the other streams.

Table 2 shows the average MSE resulting from the rate assigned to each video sequences for the first 15 GOPs. As expected, the ER scheme is able to provide less distortion to the low-complexity video, i.e. *crew*, *foreman*, by compromising

Table 2. Average MSE of each video sequence with equal-rate (ER) assignment and rate adaptation with the proposed algorithm (OPT). Total bandwidth is equal to 3000 kbps.

Gop index	Guard		Crew		Football		Foreman		Harbour	
	ER	**OPT**	**ER**	**OPT**	**ER**	**OPT**	**ER**	**OPT**	**ER**	**OPT**
1	53.71	53.71	18.59	34.64	80.86	55.87	18.40	31.66	74.28	55.52
2	57.35	54.57	19.79	37.85	74.65	59.56	18.24	29.96	81.23	56.98
3	69.45	54.63	23.52	38.67	64.02	54.06	24.63	29.99	94.54	58.27
4	81.35	59.02	39.87	39.87	63.69	56.29	17.75	33.34	75.92	57.75
5	53.71	47.36	24.89	41.67	49.53	43.55	17.73	31.58	71.93	50.97
6	55.16	41.70	28.22	38.26	16.85	24.55	19.51	34.00	73.82	46.48
7	49.11	42.22	39.87	44.31	20.23	31.36	12.40	27.35	82.14	49.31
8	49.38	42.64	33.87	38.57	31.49	39.12	14.21	28.35	73.47	48.10
9	45.79	44.11	37.47	41.71	43.89	44.20	19.20	36.12	73.51	50.37
10	42.02	46.06	42.85	43.02	47.94	45.19	19.51	32.24	69.64	52.51
11	44.49	49.17	34.40	45.68	59.81	48.88	17.77	31.33	67.82	53.78
12	42.07	40.36	25.56	39.42	41.44	41.17	18.73	30.32	71.87	46.47
13	40.17	43.18	27.09	41.48	50.24	43.84	16.55	27.87	72.23	50.91
14	42.11	56.76	23.86	35.08	82.50	56.45	25.39	45.48	68.08	57.95
15	38.29	60.28	24.81	38.76	84.63	56.84	25.92	57.12	69.48	55.82
Av.	50.95	49.05	29.64	39.93	54.12	46.73	19.06	33.78	74.66	52.74

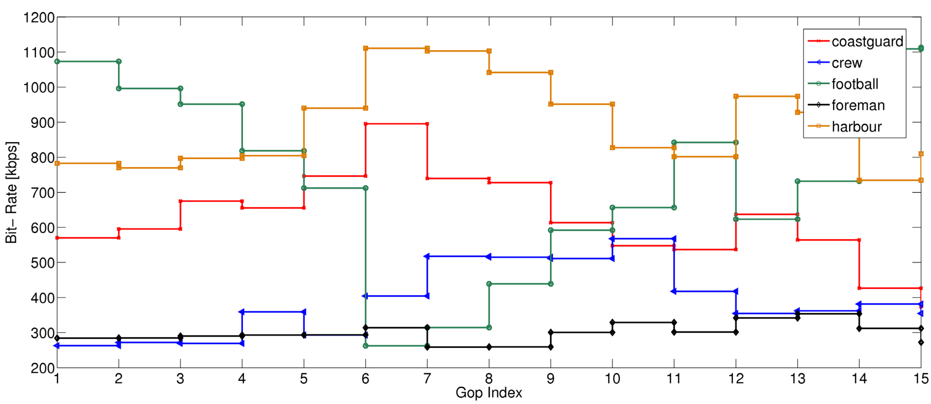

Fig. 2. Rate assigned by our adaptation algorithm in each GOP, with bandwidth equal to 3000 kbps

the distortion of the video sequences with more complexity. Our algorithm, while providing fairness, is able to improve the performance of the complex videos, by allocating more bits to video with more complex scenes. This is more clear in figure 2 where we plot the rate assigned to each video sequence GOP-by-GOP.

Table 3. Average modified MSE difference Δ_{av}, average MSE difference δ_{av} and MSE variance in each GOP interval. Comparison between the proposed algorithm (OPT) and equal-rate (ER) assignment with bandwidth equal to 3000 kbps.

GOP index	Δ_{av}		δ_{av}		**Variance**	
	ER	**OPT**	**ER**	**OPT**	**ER**	**OPT**
1	36.12	0.43	36.12	13.86	884.40	145.41
2	35.51	1.00	36.17	15.67	889.50	171.43
3	33.78	0.84	37.37	14.50	941.76	148.35
4	19.53	0.55	32.65	13.85	705.43	139.62
5	24.79	1.48	27.44	8.89	489.84	53.75
6	29.92	1.64	29.92	10.31	614.97	69.38
7	33.67	1.42	33.67	11.37	752.18	84.72
8	27.28	2.21	27.28	8.72	495.50	52.27
9	23.39	2.01	23.39	6.20	382.93	26.39
10	21.24	1.93	21.24	8.72	319.33	54.28
11	24.30	1.46	25.10	9.68	398.50	73.46
12	24.56	1.22	24.56	6.81	420.64	34.09
13	26.90	1.54	26.90	9.69	463.11	70.69
14	32.00	0.30	32.00	11.40	680.44	98.23
15	32.64	1.05	32.64	8.87	730.21	73.16
Av.	28.37	1.21	29.76	10.57	611.25	86.35

More bit-rate is assigned to *coastguard*, *football* and *harbour* video sequences, allowing them to achieve more quality.

In Table 3, we show the improvements of our proposed schemes with respect to ER. The average MSE difference is significantly reduced and equivalently the variance is decreased up to ten times. However, in this particular case of bandwidth, the MSE difference (variance) is still quite high, due to the minimum rate constraints. The average modified MSE difference $\Delta_{av} = (1/S)\sum_i \sum_{j<i} \Delta(D_i^*, D_j^*)$ according to definition in (6), is also evaluated in Table 3. Let us note that this metric also give us the information of the error generated when the discrete solution replaces the continuous solution of the relaxed problem, whose Δ_{av} is zero. This error includes two contributions: the estimation error of the model and the integrality gap. As expected the average error is not small due to mainly the low granularity of the low-rate points.

In figure 3, the MSE variance averaged over 15 GOPs is evaluated for different bandwidths. In the bandwidth interval considered, the assumptions (8) and (12) hold for each GOP. When the bandwidth is very low the two schemes provide approximately the same MSE because the optimization range is limited by the minimum rate constraints. When the bandwidth increases, our procedure improves the fairness leading the variance close to 0. A slight variance increase occurs at large bandwidths when the maximum rate constraints limit the achievable distortion. On the other hand the ER scheme generally increases the MSE variance until the base-layer constraints are active for most of the streams.

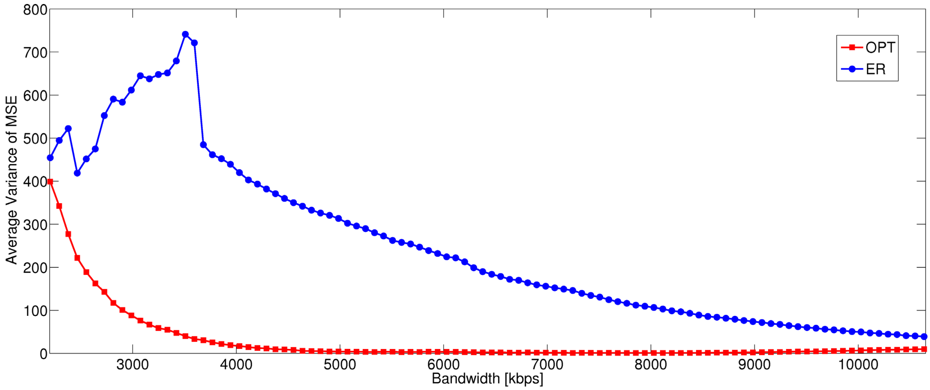

Fig. 3. Variance of the MSE averaged over 15 GOPs, with different bandwidth values. Comparison between the proposed algorithm (OPT) and equal-rate (ER) assignment.

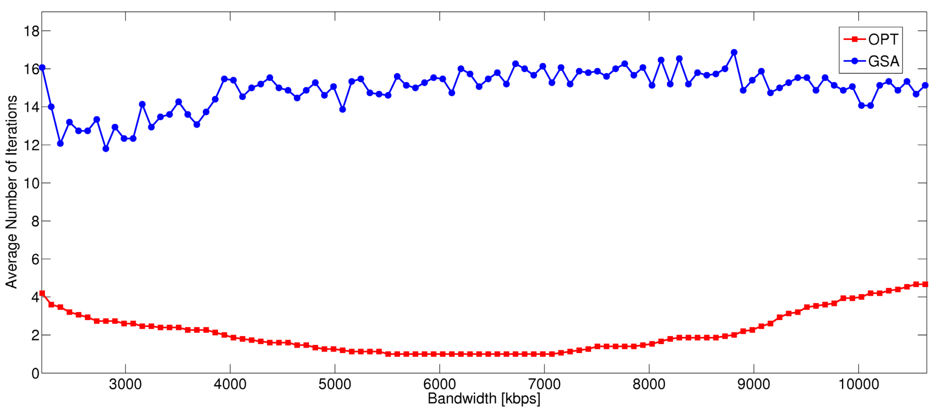

Fig. 4. Average number of iterations required by our adaptation algorithm (OPT) and golden search algorithm (GSA) to converge

This behavior can be partially reduced by controlling the base-layer bit-rate [16] to each video according to their complexity as performed for instance in [3].

To further assess our proposed scheme, we compared it to the golden search algorithm (GSA) proposed in [3], to solve the problem (13)-(16). This algorithm can be seen as a suboptimal version of our procedure. The initial solution is computed as function of the golden-section value and the difference between the lower and higher bounds, i.e. $a = \min_k D_{k,min}$ and $b = \max_k D_{k,max}$, identified by the minimum and the maximum distortion among the videos. At each iteration the solution is updated by applying the per-video constraints and by compressing the search interval consequently. The GSA terminates when the difference between the sum of the assigned rates and the available bandwidth is less of a chosen value ϵ. Nevertheless, an additional termination condition must be introduced to assure the convergence of the algorithm, that is usually indicated by the tolerance τ, i.e. $|a - b| \leq \tau$. In order to provide a fair comparison we set $\epsilon = 0.0002R_c$, and $\tau = 0.01$, leading to a sub-optimality error under 0.5% over all the investigated cases. In figure 4 we plot the average number of iterations required by the two algorithms for different bandwidths. The number of iterations of our algorithm is limited by the number of video sequences, as mentioned in sub-section 4.1, and decreases away from the minimum and the maximum bandwidths obtained as the sum of minimum and maximum rates of each video. The GSA algorithm requires generally more iterations due to the sub-optimal choice of the starting-point. This result does not change by increasing the number of videos involved in the optimization, as we also verified.

6 Conclusions

In this work we proposed a multi-stream rate adaptation framework with reference to SNR-scalability of SVC with MGS and QL. We formulated a general discrete problem with the aim to minimize the average distortion while providing fairness to different video sources. Two similar semi-analytical model that estimate the R-D relationship of each video source GOP-by-GOP are evaluated and compared with respect to goodness parameters and complexity. The general discrete problem was then relaxed and an optimal procedure was derived based on a low-complexity model. In the numerical results we showed the feasibility of our framework by analyzing the gap between the relaxed and discrete solution according to fairness metrics, the improvements with respect to an equal-rate scheme and the lower complexity of the proposed procedure with respect to an existing algorithm in the literature.

Acknowledgments. The work has been supported in part by the Italian Ministero dello Sviluppo Economico within the project WEBS, and in part by University of Ferrara, Italy. The work has been inspired by research and discussions within the cooperation framework of project e-GAP^2 of the Royal Society, UK.

References

1. Schwarz, H., Marpe, D., Wiegand, T.: Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. IEEE Transactions on Circuits and Systems for Video Technology 17(9) (September 2007)
2. Wiegand, T., Noblet, L., Rovat, F.: Scalable Video Coding for IPTV Services. IEEE Transactions on Broadcasting 55(2) (June 2009)
3. Wang, Y., Chau, L., Yap, K.: Joint Rate Allocation for Multiprogram Video Coding Using FGS. IEEE Transactions on Circuits and Systems for Video Technology 20(6) (June 2010)
4. Zhang, X., Vetro, A., Shi, Y., Sun, H.: Constant Quality Constrained Rate Allocation for FGS Coded Video. IEEE Transactions on Circuits and Systems for Video Technology (February 2003)

5. Jacobs, M., Tondeur, S., Paridaens, T., Barbarien, J., Van de Walle, R., Schelkens, P.: Statistical Multiplexing Using SVC. In: Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast, pp. 1–6 (2008)
6. JSVM 9.19.11, Reference Software (February 2011)
7. Amonou, I., Cammas, N., Kervadec, S., Pateux, S.: Optimized Rate-Distortion Extraction with Quality Layers in the Scalable Extension of H.264/AVC. IEEE Transactions on Circuits and Systems for Video Technology 17(9) (September 2007)
8. Mansour, H., Krishnamurthy, V., Nasiopoulos, P.: Rate and Distortion Modeling of Medium Grain Scalable Video Coding. In: Proc. of 2008 IEEE 15th International Conference on Image Processing, San Diego, October 12-15 (2008)
9. Cesari, M., Favalli, L., Folli, M.: Quality Modeling for the Medium Grain Scalability Option of H.264/SVC. In: Mobimedia 2009, London, UK, September 7-9 (2009)
10. Stuhlmuller, K., Fraber, N., Link, M., Girod, B.: Analysis of Video Transmission over Lossy Channels. IEEE Journal On Selected Areas In Communications 18(6) (June 2000)
11. Mansour, H., Krishnamurthy, V., Nasiopoulos, P.: Channel Adaptive Multi-User Scalable Video Streaming with Unequal Erasure Protection. In: Eight International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2007 (2007)
12. Kwon, D., Shen, M., Kuo, C.: Rate Control for H.264 Video with Enhanced Rate and Distortion Models. IEEE Transactions on Circuits and Systems for Video Technology 17(5), 517–529 (2007)
13. Thang, T., Kim, J., Kang, J., Yoo, J.: SVC Adaptation: Standard Tools and Supporting Methods. Journal, Image Communication Archive 24(3) (March 2009)
14. Gupta, R., Pulipaka, A., Seeling, P., Karam, L.J., Reisslein, M.: H.264 Coarse Grain Scalable (CGS) and Medium Grain Scalable (MGS) Encoded Video: A Trace Based Traffic and Quality Evaluation (2011) (submitted)
15. Gorkemli, B., Sadi, Y., Tekalp, M.: Effects Of MGS Fragmentation, Slice Mode And Extraction Strategies on the Performance of SVC With Medium-Grained Scalability. In: Proc. of 2010 IEEE 17th International Conference on Image Processing, Hong Kong, September 26-29 (2010)
16. Liu, Y., Guo Li, Z., Chai Soh, Y.: Rate Control of H.264/AVC Scalable Extension. IEEE Transactions on Circuits and Systems for Video Technology 18(1), 111–116 (2008)