RESEARCH   **Open Access**

Optimal SP frame selection and bit budget allocation for mobile H.264 video streaming

Stefania Colonnese, Stefano Rinauro, Lorenzo Rossi and Gaetano Scarano[*]

Abstract

Mobile video streaming services are challenging, as they obey several system constraints, such as random access facilities, efficient server storage, and flexible rate adaptation. Rate adaptation can be performed by means of seamless switching among different encoded bitstreams. The H.264 video coding standard explicitly supports bitstream switching using specific frame coding modes, namely switching pictures (SP). Locations of SP frames affect the overall bit rate and quality of streamed video. In this study, we address the issue of optimal joint selection of the SP frames locations and bit budget allocation at frame layer. The optimization is carried out via a game theoretic approach under assigned system constraints on the overall streaming rate and the maximum random access delay. Numerical simulations show that our frame layer optimal encoding procedure brings advantages in terms of several characteristics of the streamed video, encompassing enhanced rate-distortion, reduced transmission buffer occupancy, equalization of the transmission delays, and more efficient switching.

Introduction

Mobile video streaming services are experiencing a boost in cellular networks [1] or in multimedia wireless sensor networks [2] as well as in vehicular applications [3]. In mobile streaming services, the available bandwidth randomly varies, possibly due to changes in network conditions, terminal mobility, and/or handover in heterogeneous networks. The streaming server can react to these variations by streaming content extracted from differently pre-encoded versions of a given video sequence, i.e. by performing bitstream switching. Bitstream switching can be enabled by encoding INTRA frames coded without reference to any other coded frame. This would result in a significant coding cost, a less efficient bandwidth occupancy, an augmented transmission buffer storage, and worsened frame transmission delay and jitter. One of the new features of H.264 is a new coding mode named switching picture (SP), which allows drift-free bitstream switching [4,5].

Since their introduction, SP frames gathered the attention of the research community due to their unique characteristics. Applications space from streaming services [6-8], to error control [9]. The optimal selection of SP frame location has recently been addressed in [10]. Within

the framework of multiview video coding, the use of SP frames is currently under investigation to allow drift-free switching among different views [11,12].

Insertion of the so-called primary SP frames into a mobile streamed video offers a set of candidate switching locations. The switch among the differently encoded versions of the video sequence is realized at need via the transmission of a complementary encoded representation, named secondary SP frame. Since both primary and secondary SP frames encompass a motion compensation stage [13], bitstream switching is provided without resorting to the transmission of a dedicated INTRA frame. During the encoding phase, the locations for switching frames are selected, and both primary and secondary SP frames are pre-encoded and stored at the server side. During the streaming phase, primary or secondary SP frames are transmitted at user convenience, depending on whether a switching is performed or not. As a side effect, SP frames also provide error resilience, which is an important issue in mobile communications. In [4], for instance, SP frames are integrated in a framework where switching is performed within a single compressed stream to achieve both error resilience and rate scalability.

Theoretical and empirical rate-distortion curves of SP frames have been provided in [5]. The rate distortion curve of SP frames unfavorably compares with those of PREDICTED (P) frames, thus limiting the adoption of SP

*Correspondence: gaetano.scarano@uniroma1.it
DIET, Universitá "La Sapienza" di Roma, via Eudossiana 18, 00184 Roma, Italy

coding mode in mobile video streaming. In this respect, it is clear how the choice of the proper frame coding mode itself significantly affects the overall rate (and quality) of the streamed video. On the other hand, the maximum distance between two consecutive SP frames is usually assigned as a system constraint depending on the desired degree of accessibility. Still, there is a degree of freedom on where to locate the SP frames along the sequence. Large margins of quality improvement—or of bit saving—can be observed by allocating SP frames along the video sequence in accordance with a suitable optimization criterion as well as by optimally allocating the available bit budget among the different frames.

On account of these considerations, in this article, we consider a video streaming framework where different versions of the same sequence are encoded at different qualities and switching among these is realized only by means of SP frames, and we jointly address the problems of SP frame location and bit budget allocation, via a game theoretic approach. The former application of game theory in video coding has been presented in the pioneering work [14], where the authors optimize the perceptual quality of the decoded sequence while guaranteeing *fairness* in bit allocation among macroblocks via a game theoretic approach. Here, we select the optimal SP frame locations and the optimal bit allocation that maximize the overall quality of the encoded sequence. Specifically, we extend the preliminary results in [15], and we formulate a game to optimize: (i) the bit allocation between different frames of the video sequence and (ii) the frame coding mode selection. Optimization is carried out under high-level system constraints, such as the temporal distance between successive SP frames and the overall bit budget available to encode the sequence.

Related studies
State-of-the-art works on SP frames mostly focus on the realization of novel coding techniques aimed at reducing the allocated budget for SP frames. Sun et al. [16] describe a technique to improve the coding efficiency of the SP frames by limiting the mismatch between the prediction reference and the frames to be encoded. In [17], it is shown that by appropriately choosing reference pictures, the size of secondary SP frames can be reduced by up to 40% for random-access and up to 2% for rate-switching, without affecting the decoded sequence quality. In recent literature, the problem of coding mode selection has been deeply discussed. In [18], a low-complexity procedure to address INTRA mode selection is proposed, while in [19] a rate-distortion approach is employed to derive coding mode assignment procedure for intra, predicted and bidirectional predicted slices. In [10], a scheme to select the best switching points among the encoded bitstream has been introduced in the framework of a

specific bandwidth reservation scheme, namely the so-called downstairs reservation scheme. The downstairs reservation scheme is based on reserving the maximum bit rate of the encoded sequence until the frame corresponding to such a maximum is transmitted; then, the reserved rate is reduced to the next highest bit rate and so on. Altaf et al. [10] propose to select as switching points in the bitstream those frames where a change in the reserved bit rate is observed. The resulting SP frame allocation scheme allows the SP frame to be transmitted when the receiver buffer is supposed to be empty, with a minimization of the wasted bits after the bitstream switching [10]. The SP frame selection scheme proposed in [10] is not explicitly related to any optimality criterion on the encoded sequence quality. Besides, in spite of its simplicity, the scheme in [10] is suitable only under a particular reservation scheme and, since the SP frames must coincide with the changes in the reserved bit rate, the degree of accessibility may be severely limited. Thereby, it is worth seeking a procedure for coding mode selection and bit allocation independent of the rate reservation scheme possibly implemented in the streaming system; besides, the procedure should allow the user to choose the desired degree of accessibility.

Organization of the paper
In this study, we introduce an optimization procedure for coding mode selection and bit allocation derived under a game theoretic framework. Specifically, we formulate the optimization problem by representing the frames of a sequence as players whose strategy is the choice of the coding mode and the allocated bits and whose goal is the maximization of the overall sequence quality. Such an encoding optimization procedure is beneficial in different respects, ranging from rate/distortion of the encoded sequences, to network-related issues such as equalization of the transmission delay and transmission/playout buffer load.

Optimal SP selection and bit budget allocation
Here, we carry out a frame layer optimization of SP frame selection and bit budget allocation for mobile H.264 video streaming resorting to a game theoretic approach. Since the video encoder controls the resource allocation among different frames in a joint fashion, we recast the problem of coding mode selection and bit allocation in terms of strategy selection in a cooperative game.

Let us then consider a reference streaming framework where the server is equipped with K versions of the same sequence. Each of these flows is encoded at a different quality. The server simultaneously transmits these flows in multicast to the clients. Each client automatically synchronizes to the flow that better matches the experienced channel conditions and the expected video

quality. Seamless bitstream switching among the different flows is enabled only via the employment of primary and secondary SP frames.

Setting the maximum temporal distance $\tau_{\max}$ between two consecutive SP frames results in a system constraint on the random access delay. Thereby, the maximum temporal distance $\tau_{\max}$ between two consecutive SP frames is chosen so as to cope with the achievable degree of flexibility. The choice of the maximum distance between two consecutive SP frames corresponds to a maximum number of frames between switching points, $N_{\max} = f_0 \cdot \tau_{\max}$, f_0 being the video sequence frame rate. To satisfy this constraint on the maximum number of frames between switching points, the video sequence is partitioned in shorter subsequences of $N = \text{floor}(N_{\max}/2)$ frames. In each subsequence, exactly one frame shall be coded as a switching one, so as to comply with the choice of $\tau_{\max}$. Here, we exploit a game theoretic approach to jointly address the problem of coding mode assignment, i.e. the problem of the selection of the frame where the switching is enabled to occur, and the problem of resource allocation, once the coding modes are correctly assigned.

The game is described as follows:

- the players of the game are the N frames within a subsequence;
- the player's strategy is given by its coding mode and by the number of bits allocated for coding;
- the player's utility is its visual quality after decoding.

We wish to encode the N frames in the sub-sequence at the target bit rate of R [bit/s]. The overall bit budget available for the N frames is $B = RN/f_0$ (bit).

To elaborate, let us denote by c_i the coding mode assigned to the frame $i = 0, \ldots, N-1$, and by $\mathbf{c} = [c_0, \ldots, c_{N-1}]$ the coding mode N-tuple corresponding to the entire subsequence. The coding mode c_i takes a value in a finite set $\mathcal{L}$ of cardinality L representing the coding modes provided by the video encoder. The generic N-tuple $\mathbf{c}$ takes a value in a finite set $\mathcal{M}$ of cardinality M, i.e. $\mathcal{M} = \{\mathbf{c}^{(0)}, \ldots, \mathbf{c}^{(M-1)}\}$. Due to coding constraints, generally $\mathcal{M} \subseteq \mathcal{L}^N$, so that $M \leq L^N$. Here, we consider the case where c_i represents a binary choice between P and SP coding mode.[a] Let us remark here that once the number of allowed frames for each coding mode in each subsequence has been set, the N-tuples $\mathbf{c} \in \mathcal{M}$ are different permutations of the same values.[b]

Let r_i be the number of bits allocated to the ith frame and let $u_i = u_i(c_i, r_i)$ denote the utility of the ith player, i.e. the visual quality of the ith frame, $i = 0, \ldots, N-1$. Each player is characterized by the initial utility u_i^0, which measures the minimal visual quality that must be guaranteed, and by the corresponding number of allocated bits r_i^0 required to achieve the quality u_i^0. In assigning the

minimal quality that must be guaranteed to each frame u_i^0, $i = 0, \ldots, N-1$, different priors can be adopted.

In the video coding framework, the Nash-Bargaining solution [20] can be found by maximizing the following objective function [21]:

$$\mathcal{G}(r_0, \ldots, r_{N-1}; \mathbf{c}) = \prod_{i=0}^{N-1} \left(u_i(c_i, r_i) - u_i^0 \right) \qquad (1)$$

under the following three constraints:

$$(i)\ r_i \geq r_i^0,\ i = 0, \ldots, N-1;\quad (ii)\ \sum_{n=0}^{N-1} r_i \leq B;\quad (iii)\ \mathbf{c} \in \mathcal{M} \qquad (2)$$

The visual quality $u_i = u_i(c_i, r_i)$ of the ith frame is a value related to subjective perception, possibly affected by interaction between different media, and it is therefore hardly captured by an analytical relation (see [22,23] for a comprehensive survey on the subject). The approach in [14] imposes a linear relation between the bits assigned to an image area, namely a macroblock, and its resulting visual quality after decoding. This choice has the merit of leading to an analytically tractable solution for the maximization in (1). Here, we recall the relation formerly found in [14] and we extend it in order to take into account also the different coding efficiency corresponding to different frame coding modes. Towards this aim, we relate the visual quality u_i of the ith frame to the root mean square value (RMSV) σ_i of the innovation process between frame i and frame $i-1$. Specifically, we assume

$$u_i(c_i, r_i) = \frac{r_i}{K(c_i) \cdot g(\sigma_i)} \qquad (3)$$

where $g(\sigma_i)$ is a non-decreasing function[c] of σ_i. The factor $K(c_i)$ represents the coding efficiency of the coding mode option associated with the ith frame. The value of $K(c_i)$ affects u_i as a quality penalty reflecting the fact that differently encoded frames will exhibit different quality under the same bit budget. The values for $K(c_i)$ can be directly derived from the rate distortion curves [5] at a typical distortion value, or assigned through an *a priori* criterion. Under such a quality model, the objective function (1) rewrites as follows

$$\mathcal{G}(r_0, \ldots, r_{N-1}; \mathbf{c}) = \prod_{i=0}^{N-1} \frac{\left(r_i - K(c_i) \cdot g(\sigma_i) u_i^0 \right)}{K(c_i) g(\sigma_i)} \qquad (4)$$

Let us consider a general coding mode assignment $\mathbf{c}^{(m)} \in \mathcal{M}$ and let us denote by $g^{(m)}(r_0, \ldots, r_{N-1}) \overset{\text{def}}{=} \mathcal{G}(r_0, \ldots, r_{N-1}; \mathbf{c}^{(m)})$ the form of the objective function in (1) when $\mathbf{c} = \mathbf{c}^{(m)}$. Using the method of Lagrange multi-

pliers, the maximization of $g^{(m)}(r_0, \ldots, r_{N-1})$ with respect to the r_i's leads to the following optimal allocated rates:[d]

$$r_i^{(m)} = K(c_i^{(m)})g(\sigma_i)u_i^0 + \frac{1}{N}\left(B - \sum_{n=0}^{N-1} K(c_n^{(m)})g(\sigma_n)u_n^0\right) \tag{5}$$

Substituting (5) in (4) yields the following optimal value of the objective function corresponding to the coding mode assignment $\mathbf{c}^{(m)}$:

$$\begin{aligned} G^{(m)} &\overset{\text{def}}{=} g^{(m)}(r_0^{(m)}, \ldots, r_{N-1}^{(m)}) \\ &= \frac{\left(B - \sum_{n=0}^{N-1} K(c_n^{(m)})g(\sigma_n)u_n^0\right)^N}{N \prod_i K(c_i^{(m)}) \prod_i g(\sigma_i)} \end{aligned} \tag{6}$$

The maximum of (1) is then found as the supremum of the finite set $\{G^{(0)}, \ldots, G^{(M-1)}\}$, generated from (6) by varying $\mathbf{c}^{(m)}$ in $\mathcal{M}$. The optimal coding mode assignment $\mathbf{c}^{(m_{\text{opt}})}$ can be stated as follows

$$m_{\text{opt}} = \arg \max_{m \in \{0, \ldots, M-1\}} G^{(m)} \tag{7}$$

Hence, the optimal allocated bits are obtained as $r_i^{(m_{\text{opt}})}, i = 0, \ldots, N-1$.

In order to find m_{opt}, let us recall that, due to coding constraints, all the N-tuples $\mathbf{c}^{(m)} \in \mathcal{M}$ are permutations of the same values, let us say $c_i^{(m)}, i = 0, \ldots, N-1$. On account of this observation, the set $\mathcal{K}_m = \{k_i^{(m)} \overset{\text{def}}{=} K(c_i^{(m)})\}, i = 0, \ldots, N-1$, collecting the values of the weight function $K(\cdot)$ over the elements of the N-tuple $\mathbf{c}^{(m)}$ is affected by different choices of $\mathbf{c}^{(m)}$ only in the form of a permutation of its elements. As a consequence, it is easily seen how the denominator in (6) takes the same values for all the possible choices of $\mathbf{c}^{(m)}$, and hence has then no effect in (7).

Moreover, since the numerator of (6) is non-negative because of the constraints (2), the optimal coding mode N-tuple is obtained as the coding mode assignment $\mathbf{c}^{(m)}$ that minimizes

$$\mathcal{C}(\mathbf{c}^{(m)}, \sigma_i, u_i^0) = \sum_{i=0}^{N-1} k_i^{(m)} g(\sigma_i)u_i^0. \tag{8}$$

Since different choices of $\mathbf{c}^{(m)}$ affect the weights $k_i^{(m)}$ only in the form of a permutation of their indexes, minimization of (8) is achieved in correspondence of a certain index permutation of the weights $k_i^{(m)}$. Minimization of

(8) is thus obtained by the coding mode N-tuple $\mathbf{c}^{(m_{\text{opt}})}$ satisfying the following condition

$$k_i^{(m_{\text{opt}})} \leq k_j^{(m_{\text{opt}})} \text{ if } g(\sigma_i)u_i^0 \geq g(\sigma_j)u_j^0, \forall i, j = 0, \ldots, N-1 \tag{9}$$

Under the hypothesis of uniform minimal quality all over the sequence, i.e. $u_i^0 = u_{\min}, i = 0, \ldots, N-1$, the solution provided by (9) corresponds to the choice of progressively assigning the less efficient coding mode (higher values of the coding cost k_i) to the frame with lower amounts of innovation (smaller values of $g(\sigma_i)$).

The condition (9) directly comes from the following Proposition, whose proof is reported in Appendix.

Proposition 1. *Given the finite set* $\mathcal{A} = \{a_1, \ldots, a_n\}$ *such that* $a_i \in \mathbb{R}^+$ *and* $a_i \geq a_{i-1}, \forall i = 1, \ldots, n$, *and the finite set* $\mathcal{P} = \{p_1, \ldots, p_n\}$ *such that* $p_i \in \mathbb{R}^+$ *and* $p_i \leq p_{i-1}, \forall i = 1, \ldots, n$, *then, for all the possible permutations* $f : \{1, \ldots, n\} \to \{1, \ldots, n\}$ *of the index i, it results:*

$$\sum_{i=1}^{n} a_i p_i \leq \sum_{i=1}^{n} a_i p_{f(i)}$$

To recap, our optimal allocation and coding mode selection leads to the following criteria for maximization of the overall quality of the decoded sequence (1):

Coding mode selection criterion: *the coding mode assignment used in the subsequence of N frames is performed by coding with the less efficient coding modes the frames with the smallest amount of innovation.*

Bit budget allocation criterion: *the bit allocation is performed in two steps: at first an initial allocation is performed in order to satisfy the minimum quality constraints, then the remaining bit budget after this first assignment is fairly redistributed among the frames in the subsequence of N frames.*

The optimal coding procedure optimized according to these criteria is summarized in Appendix, where also a few implementation details are discussed.

The above-summarized criteria resulting from the maximization of the objective function (1) basically lead to smoother instantaneous fluctuations of the video bit rate. The smooth traffic behavior is a result of the cooperative game underlying the optimized procedure of the coding mode selection and frame bit allocation. In fact, the constraint on the initial quality u_i^0 is satisfied with minimal initial budget $\sum_{i=0}^{N-1} r_i^0$ when the less efficient SP coding mode is assigned to the frame with the minimum

amount of innovation. This assignment reduces the unbalance of the initial bit budgets r_i^0 required to guarantee the desired initial quality of the different frames. Besides, after the initial allocation, the remaining bit budget is fairly allocated among the frames. Hence, the fair allocation resulting from the joint quality improvement pursued by optimizing the objective function (1) in a cooperative fashion results in a smoother behavior of the video traffic.

Remarkably, the smoothness of the traffic of the encoded video resulting from such cooperative optimization is expected to be beneficial in a realistic network scenario, in terms both of transmission delay jitter of the video traffic and of load of network buffers. These benefits will be quantitatively assessed in "Experimental results" section.

Final remarks

The above detailed coding mode selection and bit budget allocation criteria depend on the linear visual quality model in (3). The selection of such a model for the visual quality u_i of the ith frame corresponds to a linear approximation of the peak signal-to-noise ratio (PSNR) empirical curves around a selected value of the bit rate (see for instance Figure 9 in [24]). Interestingly enough, should the following different linear model for the visual quality penalty over u_i due to the coding efficiency $K(c_i)$ be adopted

$$u_i(c_i, r_i) = \frac{r_i}{g(\sigma_i)} - K(c_i)$$

it can easily be proved that such a model results in exactly the same optimal criteria obtained with the quality model in (3).

Let us finally observe that once the optimal bit budget per frame is allocated, any rate control scheme can be employed to encode the sequence; for instance, the algorithm described in [14] can further be applied at a macroblock layer for individual frame coding.

To sum up, while recent literature works refer to within-frame optimization techniques, the scope of our optimization includes different encoded frames, jointly taking into account system constraints and rate-distortion aspects. Besides, we have demonstrated that the optimization problem phrased in (1) can separately be solved w.r.t. the two tuples $c_0, \ldots, c_{N-1}$ and $r_0, \ldots, r_{N-1}$. Moreover, although the finite and discrete nature of the tuple $c_0, \ldots, c_{N-1}$ would allow to find the optimal c_i's using an exhaustive search, we have provided a closed form solution.

The so found novel resource allocation procedure extends to frame-level the macroblock-oriented resource allocation procedure found in [14], also accounting for the unbalance between coding mode efficiency.

Experimental results

In this section, we show some experimental results of the herein analyzed optimized method, obtained using the H.264 codec [25] on different test sequences in QCIF and CIF format at the reference frame rates of $f_0 = 10$ and 30 fps.
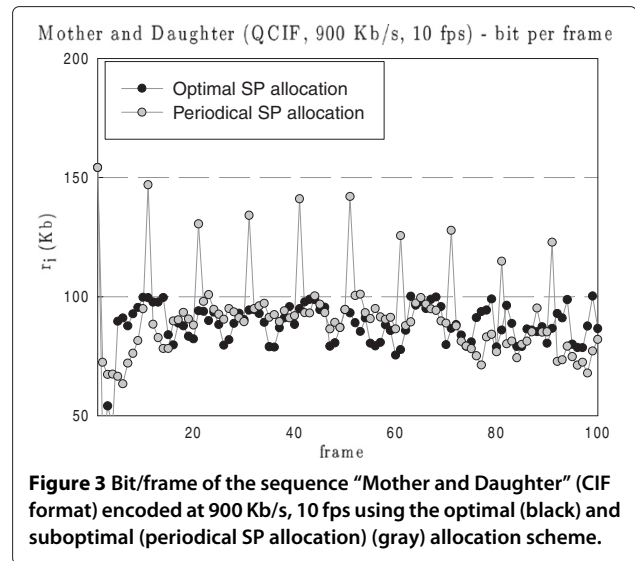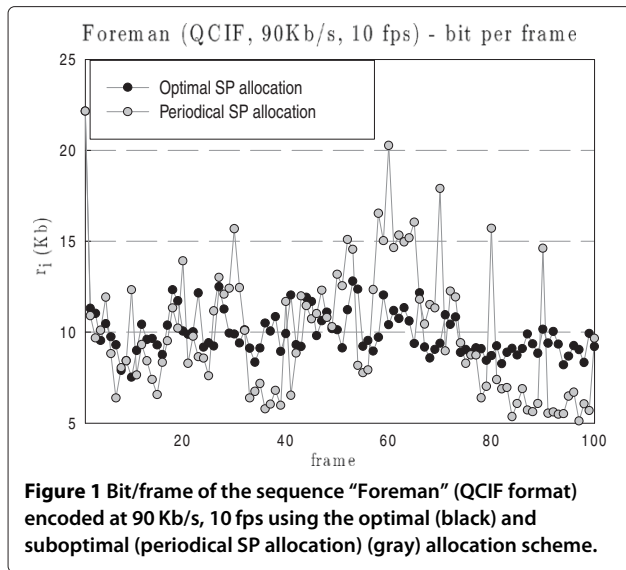
The optimization is performed by first partitioning the sequence in groups of N frames with $N = f_0 \cdot 1$, corresponding to at least one SP frame every 2 s, so as to achieve a good compromise between accessibility and compression efficiency. For every group of N frames, the switching coding mode $c_i = $ SP (SP frame) is assigned to the frame with minimum RMSV of the innovation process; the remaining frames in the subsequence are coded as $c_i = $ P (P frame). The optimization algorithm is then applied by evaluating the bit budget for each frame to be encoded. Following the guidelines in Step IV of "Optimal coding procedure" in Appendix, we compute the initial frame bit budget $r_i^0, i = 0, \ldots, N-1$ by applying a coarse quantization coding stage with quantization parameters fixed so as to assure the desired average initial quality set to $PSNR_0 = 20$ dB.[e] The optimal frame bit budgets $r_i, i = 0, \ldots, N-1$ are then evaluated by fairly redistributing the remaining bit budget according to (5). Finally, the sequence is encoded under the per frame bit budget constraints r_i. A coarse rate control procedure is implemented using a constant within frame QP; such strategy can be further refined using a spatially varying QP as described in [14].

For comparison, we consider also a suboptimal coding approach with fixed SP periodicity (one SP frame each N frames), using quantization parameters chosen according to the analysis in [5].

We first apply the presented coding mode selection and allocation scheme to encode the test sequences "Foreman" and "Coastguard" in QCIF format and the test sequence "Mother and Daughter" in CIF format at $f_0 = 10$ fps. Figures 1, 2 and 3, respectively, plot the bit/frame budgets r_i obtained employing the optimal encoding procedure on the QCIF test sequences "Foreman" and "Coastguard" at the nominal rate of 90 Kb/s and the CIF sequence "Mother and Daughter" at the nominal rate of 900 Kb/s. For comparison, we have evaluated the bit/frame budgets r_i^{PER} of the suboptimal (periodical SP insertion) strategy.
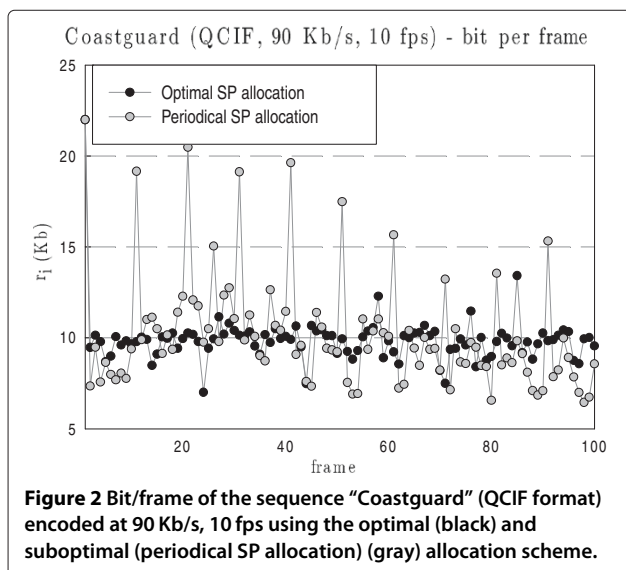
In all these cases, maximization of the objective function (1) leads to smoother instantaneous bit rate fluctuations; the effect is even more noticeable on the sequence in CIF format, showing that the higher the objective bit rate, the more effective the optimal allocation scheme is.

Noteworthily, the fair allocation strategy leading to the maximization of (1) is always achieved under an improvement of the PSNR or of the coding gain as shown in Table 1, summarizing the average bit rate and PSNR for the optimal and suboptimal strategies in different

Figure 1 Bit/frame of the sequence "Foreman" (QCIF format) encoded at 90 Kb/s, 10 fps using the optimal (black) and suboptimal (periodical SP allocation) (gray) allocation scheme.

Figure 3 Bit/frame of the sequence "Mother and Daughter" (CIF format) encoded at 900 Kb/s, 10 fps using the optimal (black) and suboptimal (periodical SP allocation) (gray) allocation scheme.

encoding conditions. We observe that the improvement is more relevant at higher spatial or temporal resolutions.

Besides, in order to test the performance of the herein presented allocation scheme over more realistic video contents, we have also considered the encoding of a group of $N = 300$ frames extracted from the movie "Fear and Loathing in Las Vegas" (labeled as "Movie" sequence in the following) encompassing a scene change. The "Movie" sequence is encoded in CIF format at $f_0 = 30$ fps at the nominal rate of 480 Kb/s. Figure 4 reports the bit/frame budgets r_i while Figure 5 reports the PSNR per frame. Inspection of the plots in Figures 4 and 5 confirms how the optimal allocation scheme allows to better share the available bit budget among the frames to be encoded with the aim of maximizing the encoded video quality.

Figure 2 Bit/frame of the sequence "Coastguard" (QCIF format) encoded at 90 Kb/s, 10 fps using the optimal (black) and suboptimal (periodical SP allocation) (gray) allocation scheme.

Let us now show how the smoothness of the instantaneous bit rate due to the cooperative optimization strategy is beneficial in a realistic network scenario. For the sake of concreteness, let us refer to the simplified model of the end-to-end communication link depicted in Figure 6. The scheme comprises a transmission buffer of size B_T, a channel at fixed nominal rate R_C, and a playout buffer of size B_R.

We first show that the adoption of the optimal encoding strategy is beneficial in terms of transmission delay jitter. In order to quantify such benefit, we compute the transmission delays $d_i = r_i/R_C$, associated to the transmission of the frames of the encoded video sequence. In Figure 7, we plot the cumulative distribution of the delays $d_i = r_i/R_C$, in the case of the CIF sequence "Mother and Daughter" encoded at a nominal bit rate of 900 Kb/s with the optimal and suboptimal (periodical SP insertion) allocation schemes. In this case, the channel rate R_C has been set to $R_C = 1000$ Kb/s, corresponding to the nominal source rate plus a gross 10% margin. Results show that, resorting to the game theoretic allocation scheme, the transmission delay is always less than 100 ms; on the contrary, when the periodical SP allocation is employed, up to 10% of the frames suffer a higher delay.

The smoothed nature of the video traffic generated using the described optimized procedure also affects the transmission buffer load. We have compared the frame loss rate observed while filling a buffer with a sequence encoded according to the optimal allocation scheme and the frame loss rate experienced by a sequence with periodic SP frames. The transmission buffer is managed with a first-in first-out (FIFO) policy; a frame is stored in the buffer only if there is available space for it, otherwise it is lost. In Figures 8, 9, and 10, we compare

Table 1 Average bit rate and PSNR measurements for the optimal and suboptimal (periodical SP allocation) strategies

Sequence	Format	fps	Bit rate R (kb/s)	PSNR (dB)	Bit rate R^{PER} (kb/s)	PSNR^{PER}(dB)
Foreman	QCIF	10	89	**35.0**	89	34.8
Coastguard	QCIF	10	**88**	31.9	91	31.9
Mother and Daughter	CIF	10	**799**	46.5	829	46.5
Foreman	QCIF	30	**93**	**28.3**	103	28.1
Movie	CIF	30	**469**	**41.6**	480	41.3

the frame loss rate obtained using the optimal and the periodical SP allocation scheme for the sequences "Coastguard" (QCIF format, 90 Kb/s, $R_C = 100$ Kb/s),"Mother and Daughter" (CIF format, 900 Kb/s, $R_C = 1000$ Kb/s), and "Movie" (CIF format, 480 Kb/s, $R_C = 500$ Kb/s). In all the considered simulations, a reduction of the frame loss rate is observed when the optimal frame layer encoding scheme is adopted.
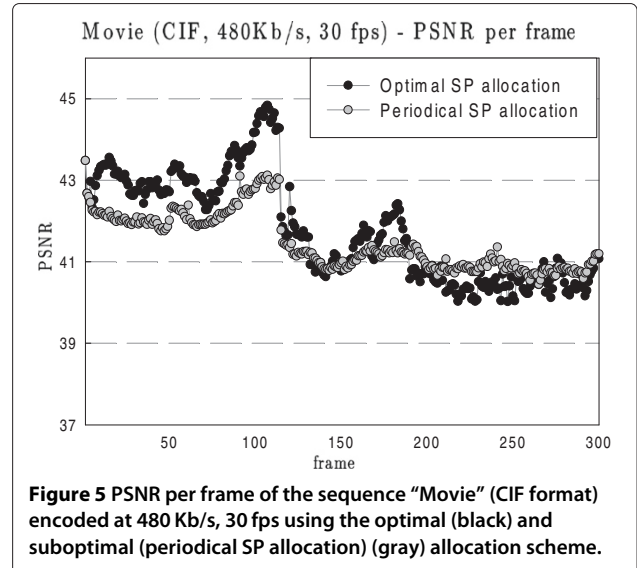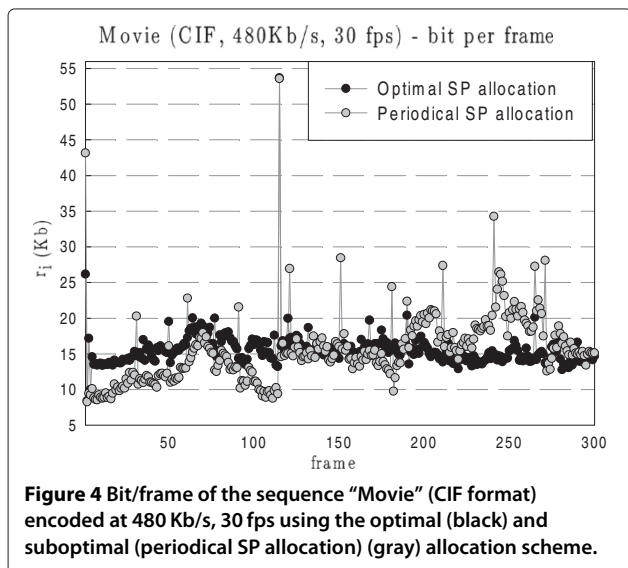
Let us now refer to the same scenario as in Figure 6, when a specific bandwidth reservation scheme, namely the Downstairs Reservation (DR) scheme is employed. As variable bit rate (VBR) video data are likely to exhibit severe bit rate fluctuations on both short and large scales, suitable smoothing procedure are designed to realize the transmission of VBR by means of a series of constant bit rate (CBR) segments. Video server is then required to reserve the correct amount of bandwidth to effectively transmit each segment. Several techniques to achieve such piece-wise CBR reservations have been proposed in recent literature. Among others, the DR scheme exhibits the desired property of avoiding upwards bandwidth reallocations, that is, every CBR segment is characterized by a bit rate equal or less than the previous segments. Such a characteristic deeply simplifies the network admission control procedures.

The DR scheme starts by reserving a channel rate equal to the peak bit rate of the encoded sequence; after the peak occurrence, the reserved rate is reduced to the next peak and so on. Specifically, let us suppose to have encoded the N frames of a given sequence so that the ith frame is assigned with r_i bit. The DR procedure starts by evaluating the following quantities

$$A_j = \frac{f_0}{j} \sum_{i=1}^{j} r_i \quad j = 1, \ldots, N$$

The A_j's represent the average transmission bit rate for subsequences composed of j frames. The largest of these averages indicates the frame for which the maximum instantaneous bit rate is observed in the encoded sequence; such a value is employed for bandwidth reservation of the first CBR segment. Once the largest A_j has been identified, say for $j = \hat{j}$, so that the first segment spans the first $\hat{j}$ frames, new averages are evaluated starting from the $(\hat{j} + 1)$th frame

$$A_l = \frac{f_0}{l - \hat{j}} \sum_{i=\hat{j}+1}^{l} r_i \quad l = \hat{j} + 1, \ldots, N.$$

Figure 4 Bit/frame of the sequence "Movie" (CIF format) encoded at 480 Kb/s, 30 fps using the optimal (black) and suboptimal (periodical SP allocation) (gray) allocation scheme.

Figure 5 PSNR per frame of the sequence "Movie" (CIF format) encoded at 480 Kb/s, 30 fps using the optimal (black) and suboptimal (periodical SP allocation) (gray) allocation scheme.
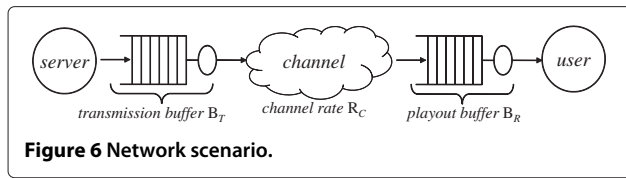
Figure 6 Network scenario.

and the largest values among the A_l's is employed for the reservation of the following segment. The procedure is iterated for the entire sequence.

According to the DR guidelines reported in [10], we modify the network scenario in Figure 6 so that the channel rate R_C varies according to the DR criterion. The transmission buffer is managed with a FIFO policy. The playout buffer is emptied at the nominal sequence frame rate, and the frames are extracted from the buffer according to their decoding order. The playout process starts with a playout delay D. A frame correctly transmitted is considered lost at the receiver side either if there is not enough space in the buffer for its storage or if it is received after its decoding deadline.[f] All the received frames exhibit a different delay at the playout buffer, due to the transmission buffer queue and the random delay introduced by the channel. We model such random delay following the channel model described in [26]. We compare the frame loss rate for the optimal coding mode assignment and bit budget allocation and the suboptimal method with fixed SP periodicity. For comparison, we also consider the DR-oriented allocation strategy presented in [10]. We consider, for this experiment, the test sequence "Foreman" in QCIF format, encoded at a nominal bit rate of 100 Kb/s. Setting $f_0 = 30$ fps, we obtain 9 steps of the DR rate within the 300 frames long sequence.

Figure 8 Transmission buffer frame loss rate for the sequence "Coastguard" (QCIF format) encoded at 90 Kb/s, 10 fps using the optimal (black) and suboptimal (periodical SP allocation) (gray) algorithm.

Within the approach in [10], the number of inserted SP frames equals the number of steps of the downstairs reservation function; for fair comparison, we have employed the same number of SP frames, namely nine SP frames for the settings of this experiment, also for the presented approach and for the suboptimal scheme with fixed SP periodicity. Table 2 summarizes the averaged bit rate and the PSNR attained, under these settings, by using the optimal strategy, the suboptimal one with fixed SP periodicity, and the approach in [10]. Figure 11 reports the frame loss rate at the transmitter buffer, while Table 3 reports the overall end-to-end frame loss rate for different sizes of both the transmitter and the playout buffer, for a playout delay $D = 5$ s. Figure 12 reports the overall end-to-end
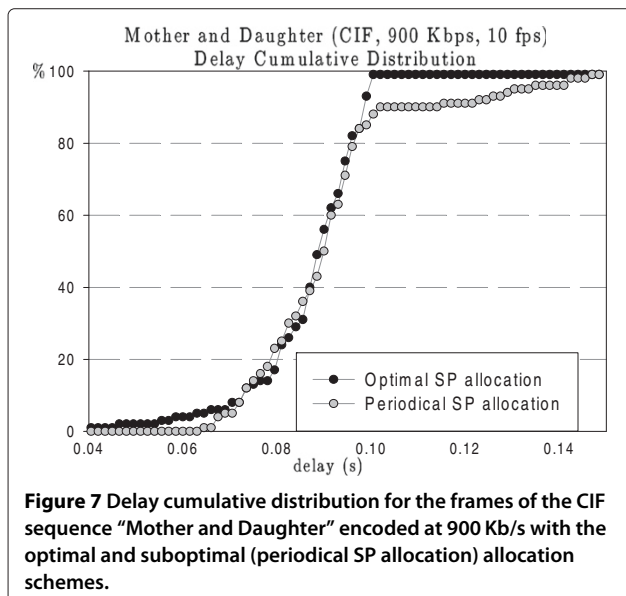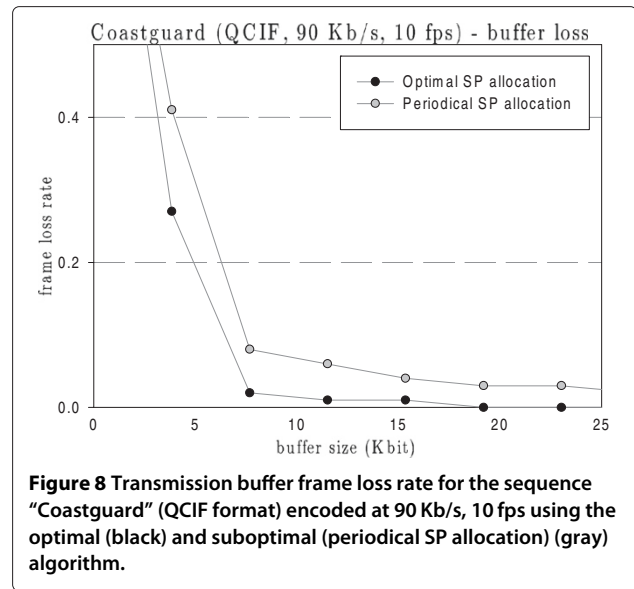

Figure 7 Delay cumulative distribution for the frames of the CIF sequence "Mother and Daughter" encoded at 900 Kb/s with the optimal and suboptimal (periodical SP allocation) allocation schemes.
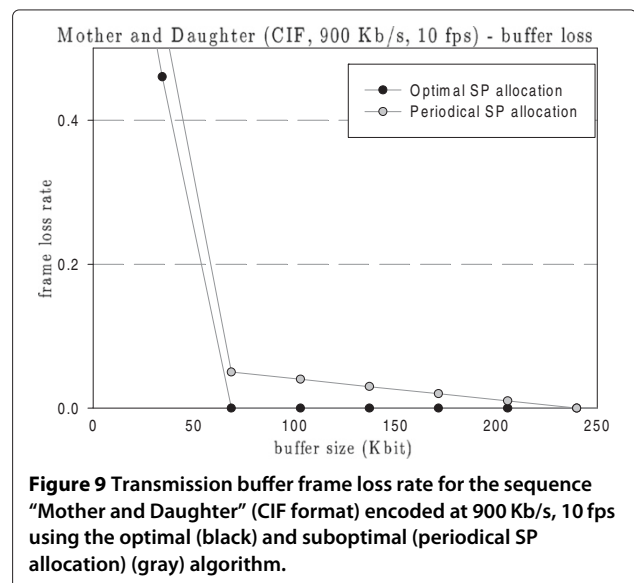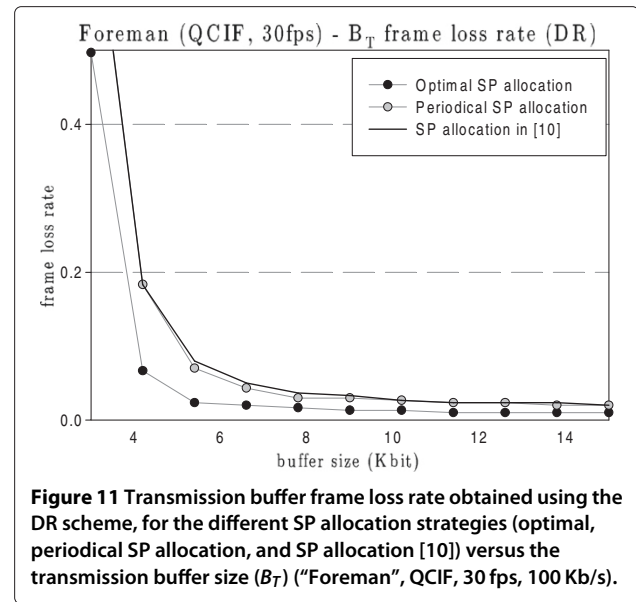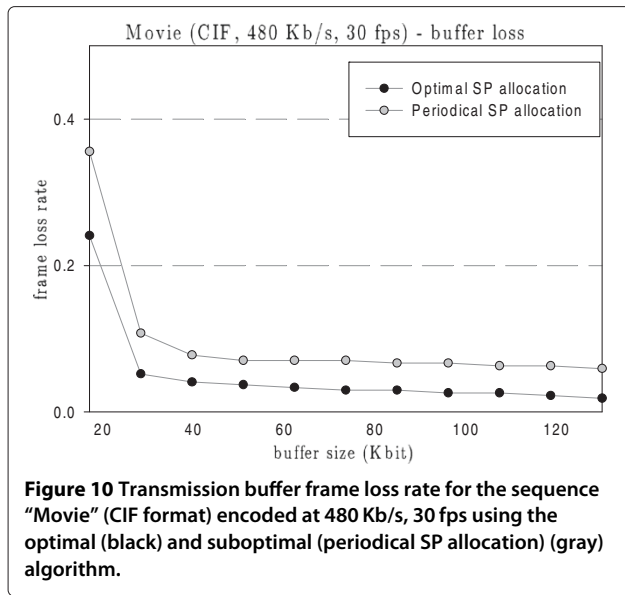

Figure 9 Transmission buffer frame loss rate for the sequence "Mother and Daughter" (CIF format) encoded at 900 Kb/s, 10 fps using the optimal (black) and suboptimal (periodical SP allocation) (gray) algorithm.

Figure 10 Transmission buffer frame loss rate for the sequence "Movie" (CIF format) encoded at 480 Kb/s, 30 fps using the optimal (black) and suboptimal (periodical SP allocation) (gray) algorithm.

Figure 11 Transmission buffer frame loss rate obtained using the DR scheme, for the different SP allocation strategies (optimal, periodical SP allocation, and SP allocation [10]) versus the transmission buffer size (B_T) ("Foreman", QCIF, 30 fps, 100 Kb/s).

frame loss rate for various sizes of the playout buffer when the transmission buffer size is fixed in order to obtain a transmission frame loss rate equal to 5% for all of the approaches (4.2 Kb for the optimal approach and 6.6 Kb for the periodical SP allocation and for the approach in [10]). Simulation results show that also in this specific scenario, under which the scheme in [10] has been designed, the coding mode assignment and the resource allocation obtained via the game theoretic approach allow to significantly reduce the buffer losses, and hence to enhance the quality of the received stream.

Also in this case, we have assessed the performance of herein presented approach over a real video content. Specifically, we have considered a portion of $N = 600$ frames extracted from the recording of a soccer match (from now on labeled as "Sport" sequence) in CIF format at a frame rate of $f_0 = 30$ fps at a nominal rate of 450 Kb/s. The said "Sport" sequence encompasses a scene change. We have run the DR scheme over 6 windows made up by 100 frames in order to obtain the number of SP frames introduced by the approach in [10], and then we have employed the same number of SP frames, namely 20 SP frames for the settings of this experiment, also for

the presented approach and for the suboptimal scheme with fixed SP periodicity. Table 4 summarizes the averaged bit rate and the PSNR attained, for the "Sport" sequence, by using the optimal strategy, the suboptimal one with fixed SP periodicity, and the approach in [10]. Figure 13 reports the frame loss rate at the transmitter buffer, while Figure 14 reports the overall end-to-end frame loss rate for various sizes of the playout buffer when the transmission buffer size is fixed in order to obtain a transmission frame loss rate equal to 5% for all of the approaches (20 Kb for the optimal approach and 55 Kb for the periodical SP allocation and for the approach in [10]).
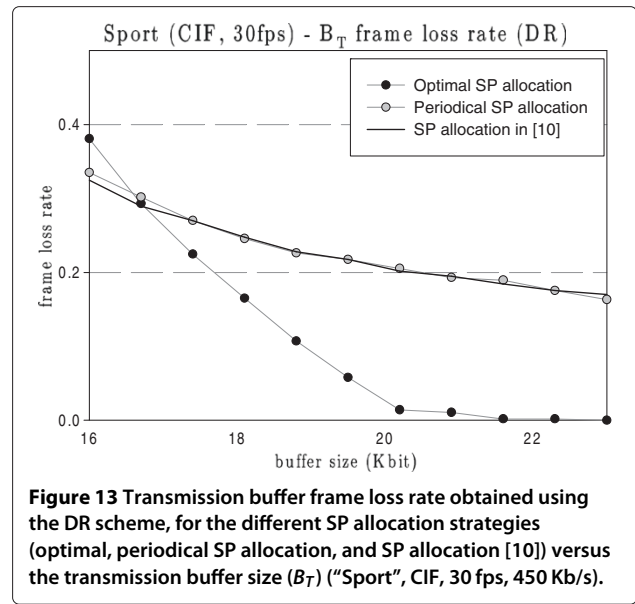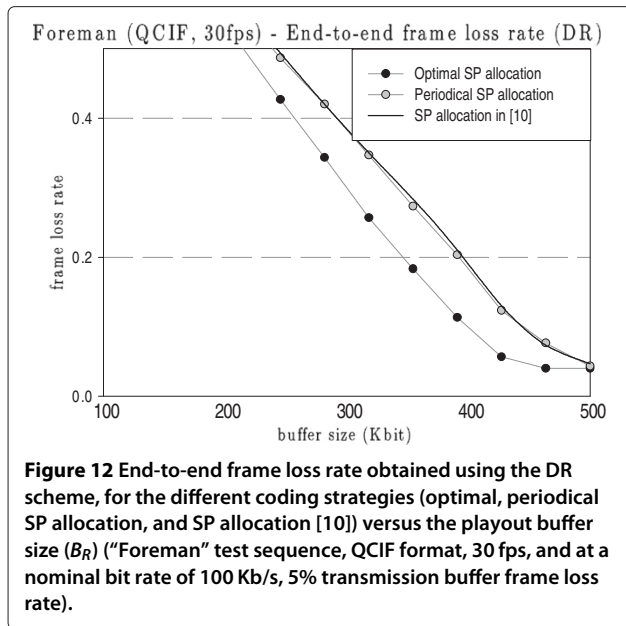
Finally, for completeness, we also report the results obtained using the same settings in absence of the DR scheme. Specifically, the transmission buffer is emptied at a constant rate equal to the nominal sequence bit rate and the channel rate R_C is constant and equal to $R_C = 100$ Kb/s for the "Foreman" sequence, and to $R_C = 500$ Kb/s for the "Sport" sequence. Figures 15 and 16 show, for both of the sequences, the frame loss rate evaluated

Table 2 Average bit rate and PSNR measurements for the optimal presented allocation scheme, the suboptimal strategy (periodical SP allocation) and the work in [10] ("Foreman" test sequence, 30 fps, QCIF format)

Allocation strategy	Bit rate (Kb/s)	PSNR (dB)
Optimal	**93**	**28.3**
Suboptimal (fixed SP periodicity)	103	28.1
Work in [10]	103	28.1

Table 3 End-to-end frame loss rate for different transmission and playout buffer under DR scheme ("Foreman", QCIF, 30 fps, 100 Kb/s)

SP frames Allocation scheme with DR	tx buffer (Kbit) Playout buffer (Kbit)			
	10.2	10.2	15	15
	460	500	390	430
Optimal	**0.02**	**0.01**	**0.09**	**0.02**
Periodical SP allocation	0.06	0.03	0.17	0.09
Work in [10]	0.05	0.03	0.22	0.12

Figure 12 End-to-end frame loss rate obtained using the DR scheme, for the different coding strategies (optimal, periodical SP allocation, and SP allocation [10]) versus the playout buffer size (B_R) ("Foreman" test sequence, QCIF format, 30 fps, and at a nominal bit rate of 100 Kb/s, 5% transmission buffer frame loss rate).

Figure 13 Transmission buffer frame loss rate obtained using the DR scheme, for the different SP allocation strategies (optimal, periodical SP allocation, and SP allocation [10]) versus the transmission buffer size (B_T) ("Sport", CIF, 30 fps, 450 Kb/s).

at the transmission side, i.e. caused only by the transmission buffer overflow, for different sizes of the transmission buffer (expressed in Kb). Simulation results show how the described approach outperforms both the periodical SP insertion and the allocation criterion introduced in [10] for all the buffer sizes. As previously stated, this is explained by the smoothed nature of the video traffic generated using the described optimized procedure. Tables 5, 6, and 7 report, for both of the sequences, the overall end-to-end frame loss rate at various sizes of the transmission and the playout buffers. Also for the overall frame loss rate our allocation method suffers a minor number of losses with respect to the other approaches. Finally, Figures 17 and 18 report the overall end-to-end frame loss rate for various sizes of the playout buffer when the transmission buffer size is fixed in order to obtain a transmission frame loss rate equal to 5% for all of the approaches. In any case, the optimal strategy exhibits the best performance in terms of transmission buffer load.
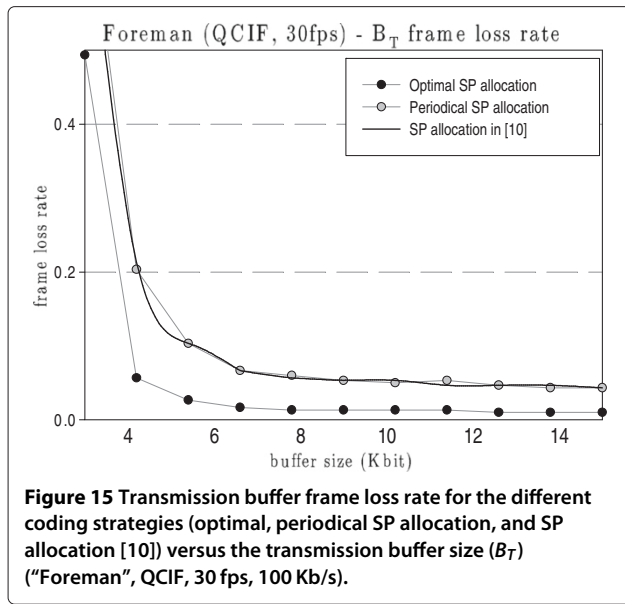
All the presented results clearly highlight the impact of the optimal criteria for coding mode assignment and bit allocation with respect to state-of-the-art approaches.

Table 4 Average bit rate and PSNR measurements for the optimal presented allocation scheme, the suboptimal strategy (periodical SP allocation) and the work in [10] ("Sport" test sequence, 30 fps, CIF format)

Allocation strategy	Bit rate (Kb/s)	PSNR (dB)
Optimal	**437**	**36.9**
Suboptimal (fixed SP periodicity)	489	36.8
Work in [10]	491	36.8

Until now, we have considered optimization of primary SP frames, i.e. the random access frames of the encoded video sequence. When a switching is requested during a streaming session, the server sends a different version of the access frame, namely the secondary SP frame, for decoder buffer synchronization purposes. Since also secondary SP frames are encoded by motion compensation, optimization of primary SP allocation is beneficial for secondary SP bit allocation too. Numerical simulations have shown a variable gain of the optimal allocation scheme over the suboptimal one; in the case of bitstream
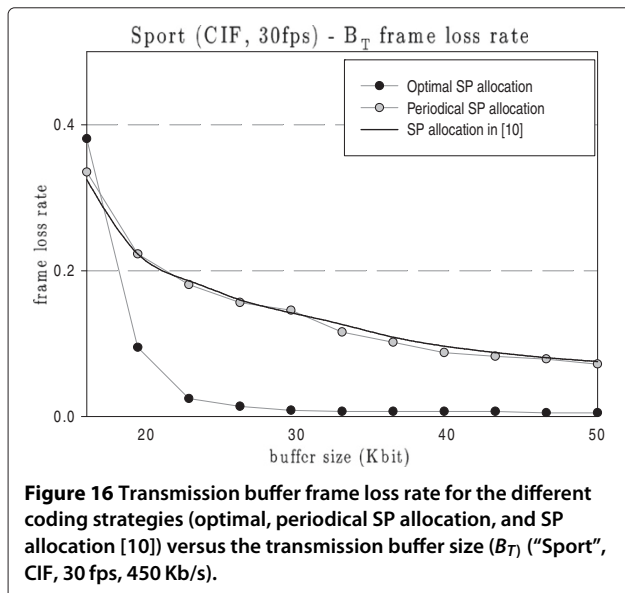

Figure 14 End-to-end frame loss rate obtained using the DR scheme, for the different coding strategies (optimal, periodical SP allocation, and SP allocation [10]) versus the playout buffer size (B_R) ("Sport" test sequence, CIF format, 30 fps, and at a nominal bit rate of 450 Kb/s, 5% transmission buffer frame loss rate).
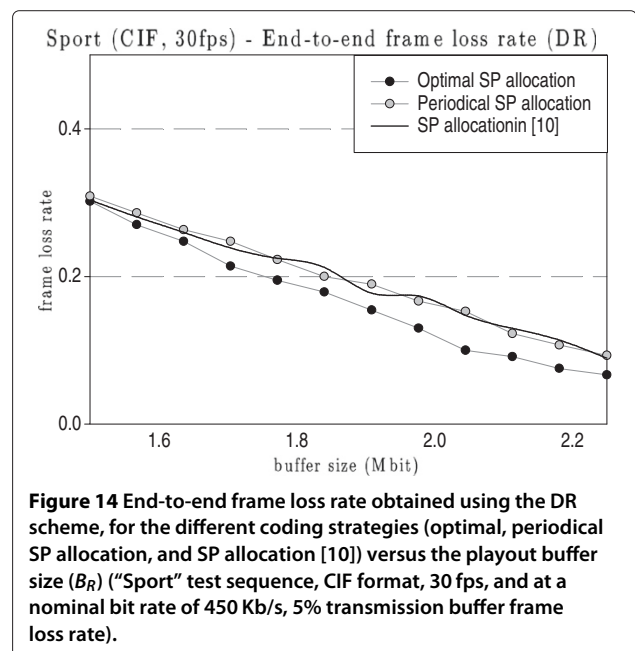
Figure 15 Transmission buffer frame loss rate for the different coding strategies (optimal, periodical SP allocation, and SP allocation [10]) versus the transmission buffer size (B_T) ("Foreman", QCIF, 30 fps, 100 Kb/s).

switching between 70 and 100 Kb/s version of the QCIF sequence "Foreman", we have observed a reduction up to 20%, with an average value of 10%, of the bits allocated to the SP secondary frames.

Conclusion

In this study, we have presented a procedure for optimal frame-level coding mode selection and bit budget allocation, with application to mobile H.264 video streaming. The optimization procedure is here derived via a game theoretic approach. The cooperative game underlying the optimized procedure of the coding mode selection and frame bit allocation basically leads to smoother

Figure 16 Transmission buffer frame loss rate for the different coding strategies (optimal, periodical SP allocation, and SP allocation [10]) versus the transmission buffer size (B_T) ("Sport", CIF, 30 fps, 450 Kb/s).

Table 5 End-to-end frame loss rate for different transmission and playout buffer under DR scheme ("Sport", CIF, 30 fps, 450 Kb/s)

SP frames Allocation scheme	tx buffer (Kbit) Playout buffer (Mbit)			
	20.3	21.2	22.1	23
	2.57	2.37	2.27	2.47
Optimal	**0.01**	**0.01**	**0.03**	**0.01**
Periodical SP allocation	0.21	0.20	0.17	0.16
Work in [10]	0.20	0.19	0.17	0.17

instantaneous fluctuations of the video bit rate. Numerical simulation results show that the encoding optimization procedure is beneficial in different respects, ranging from rate/distortion of the encoded sequences, to network-related issues such as equalization of the transmission delay, and transmission, playout buffer load.

Appendix

Proof of Proposition 1

Let us consider the finite set $\mathcal{A}^{(n)} = \{a_1, \ldots, a_n\}$ with $a_i \geq 0, i = 1, \ldots, n$, sorted in descending order, i.e. $a_i \geq a_{i-1}$, and the finite set $\mathcal{P}^{(n)} = \{p_1, \ldots, p_n\}$ with $p_i \geq 0, i = 1, \ldots, n$, sorted in ascending order, i.e. $p_i \leq p_{i-1}$. Moreover, let us denote by $F^{(n)}$ the set of all the possible permutations $f : \{1, \ldots, n\} \to \{1, \ldots, n\}$ of the first n integers.

We wish to prove that

$$\sum_{i=1}^{n} a_i p_i \leq \sum_{i=1}^{n} a_i p_{f(i)} \quad \forall f(i) \in F^{(n)}. \tag{10}$$

The proof of (10) will be carried out by induction. We will show that

(i) (10) is true for $n = 2$;
(ii) if (10) is true for $n = m - 1$, then it is true also for $n = m$.

The induction basis (i) is easily proved by simple algebra. In fact, when $n = 2$ we have that

Table 6 End-to-end frame loss rate for different transmission and playout buffer ("Foreman", QCIF, 30 fps, 100 Kb/s)

SP frames Allocation scheme	tx buffer (Kbit) Playout buffer (Kbit)			
	4.2	7.8	10.2	10.2
	420	390	390	418
Optimal	**0.07**	**0.08**	**0.07**	**0.05**
Periodical SP allocation	0.21	0.17	0.17	0.11
Work in [10]	0.20	0.17	0.17	0.11

Table 7 End-to-end frame loss rate for different transmission and playout buffer ("Sport", CIF, 30 fps, 450 Kb/s)

SP frames Allocation scheme	tx buffer (Kbit) Playout buffer (Mbit)			
	19.4	33	36.4	43.2
	2.11	2.27	2.35	2.11
Optimal	**0.09**	**0.02**	**0.01**	**0.05**
Periodical SP allocation	0.22	0.11	0.10	0.12
Work in [10]	0.22	0.12	0.09	0.11

$$a_1 p_2 + a_2 p_1 = a_1 p_2 + a_2 p_1 + a_1 p_1 + a_2 p_2 - a_1 p_1 - a_2 p_2$$

$$= (a_1 - a_2)(p_2 - p_1) + a_1 p_1 + a_2 p_2 \tag{11}$$

Given the ordering of the elements of the sets $\mathcal{A}^{(2)}$ and $\mathcal{P}^{(2)}$, the term $(a_1 - a_2)(p_2 - p_1)$ is always non-negative, and hence (11) proves (i).

Having proved the induction basis (i), let us then assume that

$$\sum_{i=1}^{m-1} a_i p_i \leq \sum_{i=1}^{m-1} a_i p_{f(i)} \quad \forall f(i) \in F^{(m-1)} \tag{12}$$

for the sets $\mathcal{A}^{(m-1)}$ and $\mathcal{P}^{(m-1)}$ ordered as stated in the hypothesis. Let us then consider the ordered sets $\mathcal{A}^{(m)}$ and $\mathcal{P}^{(m)}$ of cardinality m. We show that, under these settings, (12) implies that

$$\sum_{i=1}^{m} a_i p_i \leq \sum_{i=1}^{m} a_i p_{h(i)} \quad \forall h(i) \in F^{(m)} \tag{13}$$

Let us rewrite the right-hand side of (13) as

$$\sum_{i=1}^{m} a_i p_{h(i)} = a_m p_{h(m)} + \sum_{i=1}^{m-1} a_i p_{h(i)}$$

If $h(m) = m$, i.e. for all the permutations $h(i)$ of the first m integers whose last element is m, then (13) directly follows from (12). On the other hand, if $h(m) \neq m$, there exists an index j satisfying $h(j) = m$, so that we can write

$$\sum_{i=1}^{m} a_i p_{h(i)} = a_m p_{h(m)} + a_j p_m + \sum_{\substack{i=1 \\ i \neq j}}^{m-1} a_i p_{h(i)}$$

$$= a_m p_{h(m)} + a_j p_m + a_j p_{h(m)} - a_j p_{h(m)} + \sum_{\substack{i=1 \\ i \neq j}}^{m-1} a_i p_{h(i)}$$
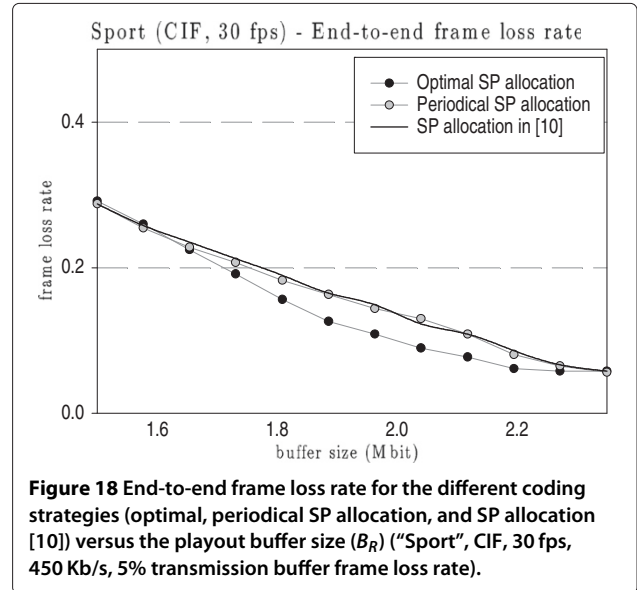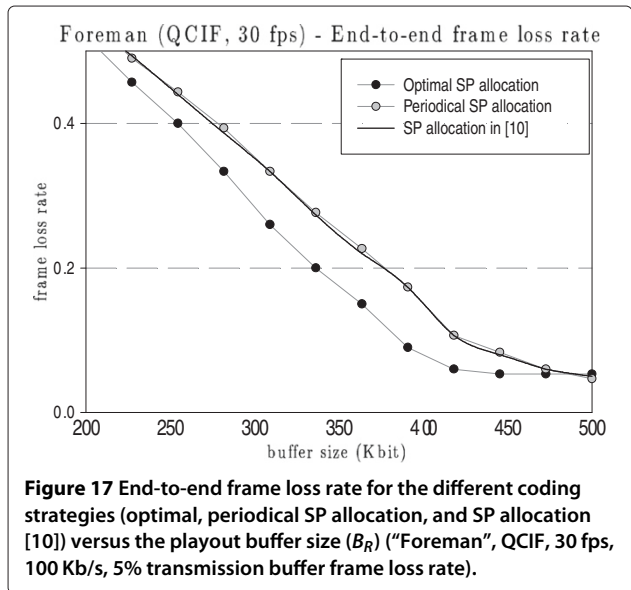
which, in turn, is rewritten as follows

$$\sum_{i=1}^{m} a_i p_{h(i)} = a_m p_{h(m)} + a_j p_m - a_j p_{h(m)} + \sum_{i=1}^{m-1} a_i p_{\tilde{h}(i)} \tag{14}$$

where $\tilde{h}(i) \in F^{(m-1)}$ is an auxiliary permutation defined as follows:

$$\tilde{h}(i) = \begin{cases} h(i) & \text{for } i \neq j \\ h(m) & \text{for } i = j \end{cases} \tag{15}$$

By adding and subtracting $a_m p_m$ to the right-hand side of (14), we have

Figure 17 End-to-end frame loss rate for the different coding strategies (optimal, periodical SP allocation, and SP allocation [10]) versus the playout buffer size (B_R) ("Foreman", QCIF, 30 fps, 100 Kb/s, 5% transmission buffer frame loss rate).

Figure 18 End-to-end frame loss rate for the different coding strategies (optimal, periodical SP allocation, and SP allocation [10]) versus the playout buffer size (B_R) ("Sport", CIF, 30 fps, 450 Kb/s, 5% transmission buffer frame loss rate).

$$\sum_{i=1}^{m} a_i p_{h(i)} = a_m p_{h(m)} + a_j p_m + a_m p_m - a_m p_m$$

$$- a_j p_{h(m)} + \sum_{i=1}^{m-1} a_i p_{\tilde{h}(i)}$$

which, in turn, can be expressed as follows

$$\sum_{i=1}^{m} a_i p_{h(i)} = (a_j - a_m)(p_m - p_{h(m)}) + a_m p_m + \sum_{i=1}^{m-1} a_i p_{\tilde{h}(i)}$$

Since $\tilde{h}(\cdot) \in F^{(m-1)}$, and because of (12), we have

$$\sum_{i=1}^{m} a_i p_{h(i)} \geq (a_j - a_m)(p_m - p_{h(m)}) + a_m p_m + \sum_{i=1}^{m-1} a_i p_i$$

$$= (a_j - a_m)(p_m - p_{h(m)}) + \sum_{i=1}^{m} a_i p_i$$

Finally, since $(a_j - a_m)(p_m - p_{h(m)})$ is non-negative given the ordering of the sets $\mathcal{A}^{(m)}$ and $\mathcal{P}^{(m)}$, we have

$$\sum_{i=1}^{m} a_i p_{h(i)} \geq \sum_{i=1}^{m} a_i p_i$$

which proves (ii).

Optimal coding procedure

Here, we summarize the coding algorithm steps, optimized according to the criteria exposed in "Optimal SP selection and bit budget allocation" section.

Step I: Sequence Partitioning—The coding optimization algorithm is applied by first partitioning the overall sequence in subsequences of equal length N. In each and every subsequence exactly one SP frame shall be introduced.

Step II: Innovation process RMSV estimation—According to the guidelines provided by Proposition 1, the RMSVs $\sigma_i, i = 0, \ldots, N-1$ of the innovation process of the N frames in each subsequence are estimated as the RMSV of the motion–compensation residuals and are sorted in ascending order. We observe that, during the coding process, the motion compensation residuals are generated with respect to the decoded reference frame. Here, we estimate the RMSV of the motion–compensation residual with respect to the original reference frame. This design choice is well suited to be implemented in streaming systems, since it leads to allocate the primary SP frames at the same time index in all the encoded bitstreams. This circumstance enables streaming server rate adaptation by seamless switching among pre-encoded bitstreams.

Step III: Coding Mode Assignment—Once the RMSV has been evaluated, the SP coding mode is assigned to the frame with the minimum RMSV of the innovation process. The values for $K(c_i)$ can be directly derived from the rate distortion curves at a typical distortion value, or assigned through an *a priori* criterion. In the specific case of only two possible coding modes (P or SP), it is sufficient to establish an ordering between $K(c_i = \text{SP})$ and $K(c_i = \text{P})$, according to the hypothesis of Proposition 1, regardless of their numerical values.

Step IV: Rate Evaluation—After the choice of the coding mode of each frame, the preliminary assignment of the initial rates r_i^0 is performed, based on the assignment of the qualities $u_i^0, i = 1, \ldots, N-1$. Recent investigations on the theoretical and experimental rate-distortion performance of SP and P frames have highlighted that a given level of distortion is achieved by higher rate for SP frames than for P frames [5]. Hence, to avoid initial quality fluctuations, a larger initial bit budget r_i^0 is assigned to the SP frame. The $r_i, i = 0, \ldots, N-1$, are then straightforwardly evaluated using (5).

Step V: Frame Coding—Once the bit budget per frame r_i has been assigned, the subsequence is ready to be encoded. The optimal frame coding under an assigned bit budget per frame can be performed according to different rate control techniques. For instance the optimal approach presented in [14] can be applied; according to this algorithm, the quantization parameter is properly chosen for each macroblock, in order to meet the fairest bit allocation among macroblocks satisfying the bit budget constraint. If the whole frame is encoded by a single quantization parameter, this latter shall be chosen equal to the minimum value compatible with the assigned value r_i.

Endnotes

[a] Extension to the case where B frames are also considered is straightforward, provided the number of allowed B frames in the subsequence is fixed.

[b] For instance, if one and only one out of the N frames in each subsequence is allowed to be an SP frame and the other frames are set as P frames, then it results that $M = N$ and all the N-tuples **c** will exhibit the form **c** = $[\text{P P} \cdots \text{SP} \cdots \text{P}]$, thus differing one from the other only in the location assigned to the SP frame.

[c] As in [14], here we set $g(\sigma_i) = \sigma_i^\alpha$ with $\alpha = 0.8$.

[d] As the optimal allocated bits evaluated according to (5) are real values, they must be quantized to provide an input to the encoder. For instance, the closest integer to $r_i^{(m)}$ can be considered as the assigned rate. The quantization loss thus introduced by this approximation comprises the effect of a single bit on the whole frame and it is therefore negligible.

[e] Let us remark that, when the objective function is maximized, all the N frames overcome this limit. The minimum initial quality can be instead regarded as a parameter that allows to determine initial bit budget,

which is assigned in an unbalanced way and, by complement, the residual bit budget, which is assigned on a fair basis.

[f] In-network losses are neglected in this test.

Competing interest
The authors declare that they have no competing interests.

References

1. T Stockhammer, G Liebl, M Walter, Optimized H.264/AVC-based bit stream switching for mobile video streaming. EURASIP J. Appl. Signal Process. **1**, 1–19 (2006)
2. IF Akyildiz, T Melodia, KR Chowdhury, Wireless multimedia sensor networks: applications and testbeds. Proc. IEEE. **96**(10), 1588–1605 (2008)
3. L Qiong, Y Andreopoulos, M van der Schaar, Streaming-viability analysis and packet scheduling for video over in-vehicle wireless network. IEEE Trans. Veh. Technol. **56**(6), 3533–3549 (2007)
4. W Tan, G Cheung, SP-frame selection for video streaming over burst-loss networks. in *Proc. of IEEE International Symposium on Multimedia,* Vol.1 (Irvine, CA, Palo Alto, CA, USA, 12–14 December 2005)
5. E Setton, B Girod, Rate-distortion analysis and streaming of SP and SI frames. IEEE Trans. Circuits Syst. Video Technol. **16**(6), 733–743 (2006)
6. K-K Lai, Y-L Chan, W-C Siut, Quantized transform-domain motion estimation for SP-frame coding in viewpoint switching of multiview video. IEEE Trans. Circuits Syst. Video Technol. **20**(3), 365–381 (2010)
7. BP Poor, M Fleury, M Altaf, M Ghanbari, Adaptive video stream switching for an IEEE 802.16 channel. in *Wireless Advanced (WiAd) 2011* (London, UK, IEEE, 20–22 June 2011)
8. C-P Chang, C-W Lin, R-D optimized quantization of H.264 SP-frames for bitstream switching under storage constraints. in *IEEE International Symposium on Circuits and Systems,* Vol.2 (Kobe, Japan, 23–26 May 2005), pp. 1242-1245
9. G Cheung, W Tan, Low-latency error control of H.264 using SP-frames and streaming agent over wireless networks. in *Proc. of IEEE International Conference on Communications,* Vol.1 (Glasgow, UK, 24–28 June 2007), pp. 1790-1796
10. M Altaf, E Khan, M Ghanbari, NN Qadri, Efficient bitstream switching for streaming of H.264/AVC coded video. Eurasip J. Image Video Process. **7**, 1–12 (2011)
11. T Maugey, P Frossard, Interactive multiview video system with non-complex navigation at the decoder. IEEE Trans. Multimed. arXiv:1201.0598, (2012) (submitted)
12. K-K Lai, Y-L Chan, C-H Fu, W-C Si, Viewpoint switching in multiview videos using SP-frames. in *IEEE International Conference on Image Processing,* Vol.1 (San Diego, USA, 12–15 October 2008), pp. 1776–1779
13. M Karczewicz, R Kurceren, The SP- and SI-frames design for H.264/AVC. IEEE Trans. Circuits Syst. Video Technol. **13**(7), 637–644 (2003)
14. I Ahmad, J Luo, On using game theory to optimize the rate control in video coding. IEEE Trans. Circuits Syst. Video Technol. **16**(2), 209–219 (2006)
15. S Colonnese, G Panci, S Rinauro, G Scarano, Optimal video coding for bit rate switching applications: a game-theoretic approach. in *Proc. of IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks,* Vol.1 (Espoo, Finland, 18–21 June 2007–15), pp. 1–4
16. X Sun, S Li, F Wu, J Shen, W Goo, The improved SP frame coding technique for the JVT standard. in *Proc. of IEEE International Conference on Image Processing,* Vol.2 (Barcelona, Catalonia, Spain, 14-18 September 2003), pp. 297–300
17. W Tan, B Shen, Method to improve coding efficiency of SP frames. in *Proc. of IEEE International Conference on Image Processing,* Vol.1 (Atlanta, GA, USA, 8–11 October 2006), pp. 1361–1364
18. J Ascenso, F Pereira, Low complexity intra mode selection for efficient distributed video coding. in *Proc. of IEEE International Conference on Multimedia and Expo,* Vol.1 (New York, NY, June 28–July 3, 2009), pp. 101–104
19. I Choi, J Lee, B Jeon, Fast coding mode selection with Rate-Distortion optimization for MPEG-4 Part-10 AVC/H.264. IEEE Trans. Circuits Syst. Video Technol. **16**(12), 1557–1561 (2006)
20. J Nash, Two-person cooperative games. Econometrica. **21**, 128–140 (1953)
21. A Stefanescu, MW Stefanescu, The arbit rated solution for multi-objective convex programming. Rev. Roum. Math. Pure Appl. **29**, 593–598 (1984)
22. J You, U Reiter, MM Hannuksela, M Gabbouj, A Perkis, Perceptual-based quality assessment for audio-visual services: a survey. Elsevier Signal Process.: Image Commun. **25**(7), 482–501 (2010)
23. K Seshadrinathan, R Soundararajan, AC Bovik, LK Cormack, Study of subjective and objective quality assessment of video. IEEE Trans. Image Process. **19**(6), 1427–1441 (2010)
24. S Ma, W Gao, Y Lu, Rate-distortion analysis for H.264/AVC video coding and its application to rate control. IEEE Trans. Circuits Syst. Video Technol. **15**(12), 1533–1544 (2005)
25. H.264/AVC Codec Software Archive [Online], ftp://ftp-imtc-files.org/jvt-experts/reference_software
26. PA Chou, Z Miao, Rate-distortion optimized streaming of packetized media. IEEE Trans. Multimed. **8**(2), 390–404 (2006)