# Low-Complexity Joint Temporal-Quality Scalability Rate Control for H.264/SVC

Randa Atta[1] and Mohammad Ghanbari[2], Life Fellow, IEEE

***Abstract*—Rate control in scalable video coding (SVC) is a very challenging problem because of the inter-layer prediction structure which makes developing an efficient rate-control algorithm complex and difficult. Little prior work is available for joint temporal-quality (T-Q) scalability considering the rate-distortion (R-D) dependency among the temporal and quality layers. However, most of the rate control algorithms in SVC suffer from high computational complexity, growing significantly with the number of layers. In this paper, a single-pass joint temporal-quality rate-control algorithm is presented for H.264/SVC. In this algorithm, by analyzing the R-D dependency of joint T-Q scalability, Cauchy distribution-based rate-quantization (R–Q), and distortion-quantization (D–Q) models, a set of empirical values are first derived to estimate the initial values of the R-D model parameters for the joint temporal and quality layers. Then, a novel prediction mechanism to update these model parameters is proposed to allocate the bit budgets efficiently among the temporal and quality layers and hence to improve the performance of the proposed algorithm. Experimental results show that the proposed algorithm achieves better coding efficiency with low computational complexity compared to two other benchmark rate-control algorithms.**

***Index Terms*—H.264/SVC, temporal-quality scalability, joint bit allocation, rate-distortion optimization, video coding.**

## I. Introduction

Rate control (RC) is an important part in video coding that imposes some constraints on video transmission, such as the limited channel bandwidth and transmission delay. Consequently the major task of rate control is to adapt the rate of the bit stream to match the available channel bandwidth with minimal delay while achieving highest possible video quality. Furthermore, it guarantees that the oscillation in bit rate is within the tolerance of the virtual buffer and prevents the buffer from "underflow" or "overflow".

Rate control algorithms are often formulated as an optimal bit allocation problem. The problem can be interpreted as how efficiently one can distribute a given bit budget among different control levels (such as group of pictures (GOP), frame layer and macro-block (MB) layer). The proper quantization parameters (QP) at frame or/and macro-block levels are then estimated to minimize the distortion. Several approaches, ranging from high complexity operational R-D (ORD) approaches [30, 31] to simpler analytical R-D model approaches [4-24], have been proposed to deal with this complex bit allocation problem. Several rate control algorithms based on analytical R-D models have been proposed for non-scalable video coders [4-8]. Some of them have been recommended in video coding standards such as Test Model Near-term 8 (TMN8) [5] for H.263, and JVT-G012 [6] for the advanced video coding (AVC) standard H.264/AVC.

On the other hand, bit allocation in SVC is a very challenging problem because of the inter-layer prediction structure which makes the R-D characteristics of one enhancement layer dependent on its preceding layers. This structure makes developing the rate-control algorithms complex and difficult. Recently, several RC algorithms have been developed for SVC [9-24], including the temporal-, spatial-, and quality-layer RC algorithms. Some of them are based on the algorithms adopted in the previous video coding standards which do not exploit inter-layer dependency among the layers as in [9]. The other RC algorithms, which can be classified into single- and multi-pass algorithms, have considered the inter-layer dependency and hierarchical temporal prediction structure for H.264/AVC scalable extensions [1, 2].

For temporal-layer rate control, Liu et al. [10] proposed a frame level bit allocation algorithm for temporal scalability by utilizing a set of empirically weighted factors for allocating the bits among the temporal layers, and a linear sum bits R-Q model [8] was used to determine the quantization parameter for each coding unit. Even though utilizing a fixed weighting factor at each temporal layer improves the bit allocation strategy, it was not able to maximize coding efficiency. In [11], an adaptive weighting factor was developed for efficient frame level bit allocation among the various temporal layers. Although using adaptive scaling factor scheme improved the performance of the temporal scalability, it cannot be properly justified to represent the dependency among the temporal layers. Cho et al. [12] proposed a multi-pass GOP-based dependent distortion model that takes the inter-dependency among the temporal layers into consideration. Although the

[1] Randa Atta is with the Electrical Engineering Department, Faculty of Engineering, Port Said University, Port Said, Port Fouad 42523, Egypt, (e-mail: r.atta@eng.psu.edu.eg).

[2] Mohammad Ghanbari is Emeritus Professor at the School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK, CO4 3SQ, (e-mail:ghan@essex.ac.uk), as well as Professor at the School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran (e-mail:ghan@ut.ac.ir).

algorithm given in [12] reduces the computational complexity as compared to that in [13], it still requires a number of encoding passes to calculate the model parameters. A practically single-pass rate control algorithm for H.264/SVC hierarchical B-pictures was developed in [14] to reduce the computational complexity.

Regarding spatial- or quality-layer rate control, exploring the dependency of the interlayer R-Q characteristics to improve rate-control performance is required. Recently, in [15] Hu et al. proposed a spatial-layer RC algorithm for SVC by first introducing an adaptive Qp-initialization model to determine the initial Qp value for the base and enhancement layers. Consequently a two-stage Qp estimation strategy based on the Cauchy distribution-based R-Q model [7] was designed to improve rate-control performance by implementing a frame complexity prediction method and an adaptive model-parameter technique. It has been shown that the rate-control performance of this algorithm was superior to the other two RC algorithms in [10] and [16]. Liu et al. [17] proposed a multi-pass model-based spatial layer bit allocation algorithm for H.264/SVC. They investigated the inter-layer dependency in terms of rate and distortion among the spatial layers and derived the analytical rate and distortion models. Subsequently, a single-pass bit allocation algorithm was proposed in [19] for spatial scalability of H.264/SVC.

Most of the existing rate-distortion models are available for temporal or/and spatial scalability coding of H.264/SVC. Little prior work is available for quality and joint T-Q scalability to consider the R-D dependency among the temporal and quality layers. Li et al. [21] developed one-pass multi-layer rate-distortion optimization algorithm for quality scalability. Later, a quality-layer bit allocation algorithm for H.264/SVC was presented in [22] by establishing the rate and distortion models for quality layer of H.264/SVC. Cho et al. [23] proposed a joint temporal-quality layer bit allocation algorithm based on an analytical solution to a Lagrangian equation. This algorithm allocates the assigned bit budget at each quality layer to each coding unit based on their proposed dependent linear R-D models for the joint T-Q scalability of H.264/SVC. Although the performance of this algorithm outperforms that of the Joint Scalable Video Model (JSVM) FixedQPEncoder [3], it still demands multiple pre-encoding passes in order to determine the model parameters. Due to this extra computational requirement, its complexity is still high.

In this paper, a single-pass joint temporal-quality bit allocation algorithm is introduced. The main contributions of this paper can be summarized as follows: First, the work in [23] is extended to simplify the optimization problem by specifying an overall target bit rate to encode all the quality layers rather than predefining a target bit rate at each quality layer as done in [23]. It is also extended to support the quality layers with different temporal resolutions in order to achieve better R-D performance when bit budgets are allocated. The optimal bit allocation problem is then formulated using the Lagrangian multiplier approach and solved numerically to adaptively distribute this overall target bit rate by considering the dependency among the layers. This developed joint T-Q layer dependent bit allocation algorithm still requires calculating the model parameters. Second, instead of performing multiple pre-encoding iterations to decide the model parameters as done in [23], an adaptive model-parameter initialization scheme is proposed for joint temporal-quality layers. In this scheme by analyzing the R-D dependency of joint T-Q scalability and Cauchy distribution-based R-Q and D-Q models, suitable initial values of R-D model parameters are derived. A novel adaptive model-parameter mechanism is also proposed to update these model parameters during the encoding process. These two aspects not only lead to improve the overall bit allocation performance but also to significantly reduce the computational complexity compared to [23]. This will be demonstrated in the experimental section.

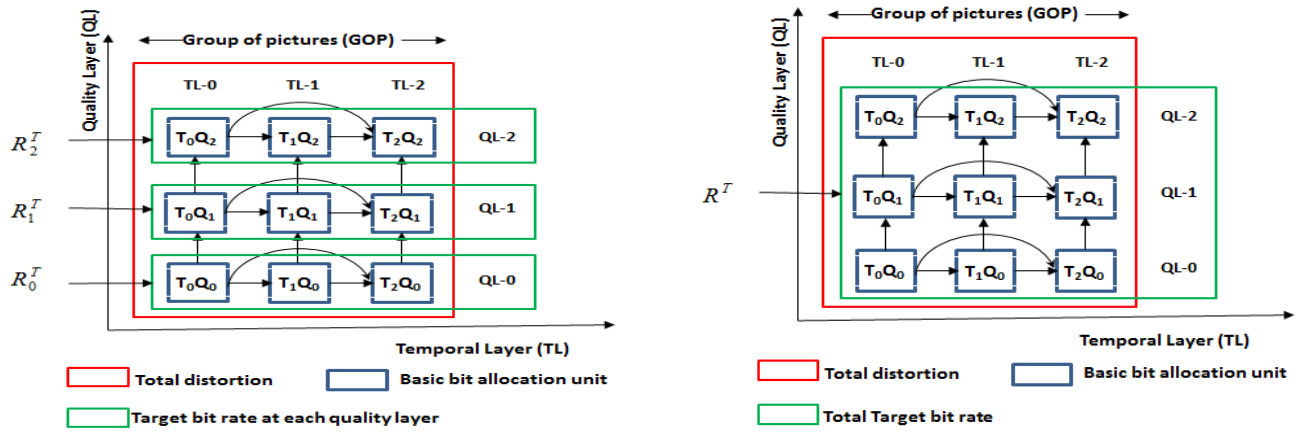
The remainder of the paper is organized as follows: In Section II, the multi-pass joint T-Q layer bit allocation algorithm in [23] is briefly reviewed and simplified such that a total target bit rate is defined and distributed among the quality layers. The proposed single-pass joint T-Q layer bit allocation algorithm is described in Section III, where the adaptive model-parameter initialization scheme is introduced. Experimental results and discussions along with computational complexity are presented in Section IV. Finally, concluding remarks are given in Section V.

## II. R-D Model In A Joint Temporal-Quality SVC

This section briefly reviews the multi-pass joint temporal-quality layer bit allocation algorithm proposed in [23]. It also simplifies the optimization problem by specifying and allocating a total target bit rate among the quality layers.

### *A. Problem Formulation*

In the joint T-Q layer bit allocation problem [23], a scalable block defined by a temporal layer (TL-ID) and a quality layer (QL-ID) identification number is used as a basic bit allocation unit. Each scalable block consists of a frame or a set of frames. In general, allocation of bits among the temporal and quality layers can be carried out within a GOP using two simple strategies as illustrated in Fig. 1. In the first strategy the target bit rate for each QL is given according to the requirement of end-users/applications. The bit budget assigned to each quality layer (QL) within a GOP is adaptively allocated to each scalable block within the same quality layer, similar to those defined in [22] and [23]. In this strategy, the optimization problem for dependent bit allocation can be formulated as seeking the optimal quantization step sizes of each scalable block in a GOP such that the total GOP distortion is minimized subject to a target bit rate for each QL. Let $N_Q$ and $N_T$ be the number of quality and temporal layers, respectively. Given the target bit budget, $R_k^T$ , at each quality layer QL-*k*, the constrained bit allocation problem can be mathematically given as:




(a) (b)

Fig. 1. Illustration of two bit allocation strategies within a GOP. (a) Strategy I: multi-Rate strategy. (b) Strategy II: fixed-Rate strategy.

$$\mathbf{Q}^* = \arg\min_{\mathbf{Q}\in\Omega} \sum_{k=0}^{N_Q-1}\sum_{i=0}^{N_T-1} D_{i,k}(q_{i,k}), \text{ subject to } \sum_{i=0}^{N_T-1} R_{i,k}(q_{i,k}) \le R_k^T, \ \forall k \in \{0,\cdots, N_Q-1\}, \quad (1)$$

where $D_{i,k}(q_{i,k})$ and $R_{i,k}(q_{i,k})$ are respectively the distortion and the rate of a scalable block $T_iQ_k$ at temporal layer TL-*i* and quality layer QL-*k*. $\mathbf{Q}$ is an $N_Q \times N_T$ matrix whose elements $(q_{i,k})$ are the quantization step sizes (i.e., *q* values) of all the scalable blocks in a GOP. $\mathbf{Q}^*$ and $\Omega$ are the optimal quantization step size and the set of all possible quantization step sizes, respectively. This constrained optimization problem can be solved using the Lagrangian multiplier method and converted into its equivalent unconstrained form as in [23]

$$J(\mathbf{Q}^*,\Lambda^*) = \arg\min_{\mathbf{Q}\in\Omega,\Lambda\in\Re^{N_Q}} J(\mathbf{Q},\Lambda)$$

$$J(\mathbf{Q},\Lambda) = \sum_{k=0}^{N_Q-1}\sum_{i=0}^{N_T-1} D_{i,k}(q_{i,k}) + \sum_{k=0}^{N_Q-1}\lambda_k\left(\sum_{i=0}^{N_T-1} R_{i,k}(q_{i,k}) - R_k^T\right), \quad (2)$$

where $\Lambda$ is an $N_Q \times 1$ vector whose elements are the Lagrangian multipliers $\lambda_k$'s. This kind of bit allocation strategy, called multi-Rate strategy, may not appropriately assign the bits to each quality layer (i.e. unsuitable $R_k^T$ bounds for the given constraints) and then the dependency among the layers may not be considered well. Thus the overall optimal R-D performance may not be achieved as will be illustrated in Section IV-A.

On the other hand, in the second strategy denoted by fixed-Rate, the overall bit budget ($R^T$) for the full temporal-quality resolution is given. An encoder has still to distribute this bit budget adaptively among the quality and the temporal layers by considering the dependency among these layers for guaranteed optimal coding efficiency. It should be mentioned that in the multi-Rate strategy the target bit rate for each QL is given according to the requirement of end-users/applications while in the fixed-Rate strategy there is no constraints on bit rate per quality layer (i.e., it is variable), but the constraint is on the total budget. Hence the multi-Rate strategy is suitable for video distributions at various known target bit rates but the fixed-Rate strategy is suitable for layered video coding, protecting lower layers with higher priority to higher layers, but resulting higher overall quality. In this paper, the focus is on the second strategy, where the bit rate of every (temporal-) quality- layer is adaptively determined during the rate control process rather than predefining it as in the first strategy. For more discussions about the comparison between these two strategies, refer to Section IV-A. In this bit allocation strategy, the Lagrangian cost function can be expressed as:

$$\mathbf{Q}^*,\lambda^* = \arg\min_{\mathbf{Q}\in\Omega,\ \lambda\in\Re} J(\mathbf{Q},\lambda),$$

$$J(\mathbf{Q},\lambda) = \sum_{k=0}^{N_Q-1}\sum_{i=0}^{N_T-1} D_{i,k}(q_{i,k}) + \lambda\left(\sum_{k=0}^{N_Q-1}\sum_{i=0}^{N_T-1} R_{i,k}(q_{i,k}) - R^T\right), \quad (3)$$

where $\lambda$ is the Lagrangian multiplier. It can be seen that in the unconstrained optimization problem (3) there is an overall bit budget ($R^T$) and one Lagrangian multiplier while in (2) there are $N_Q \times 1$ values of both the bit budgets $R_k^T$ and the Lagrangian multipliers $\lambda_k$'s. The given total bit budget in (3) is distributed among the temporal and quality layers whereas the given bit budget at each quality layer in (2) is distributed among the temporal layers. The unconstrained optimization problem either in (2) or in (3) can be solved using an exhaustive search over all possible combinations of the quantization step sizes for each temporal-quality layer (scalable block) in a GOP. As the number of layers increase, the complexity of this search algorithm increases too. The complexity issue is solved by developing an analytical model-based bit allocation algorithm. This requires $R_{i,k}(q_{i,k})$ and $D_{i,k}(q_{i,k})$ for each scalable block to be first estimated and then the optimal quantization step sizes that minimize the cost function in (2) or in (3) can be calculated. Details of the dependent linear R-D models for the joint T-Q scalability will be discussed in the next subsection.

### B. R-D Models for a Scalable Block in Joint T-Q Scalability

R-D models can be characterized by their rate-quantization step size R–Q and distortion-quantization D–Q functions which have been extensively studied in the literature [7], [8], [25]-[28]. Many of these models have been developed based on observations and analysis. The first step in developing the dependent R-D functions for combined T-Q scalability was introduced in [23] and is used in this paper.

For R-D dependency of a scalable block in temporal scalability studied in [23], first, it was shown that the rate of a dependent scalable block in the temporal scalability $T_i$ and quality scalability $Q_k$ is independent to that of the temporally preceding blocks $T_j Q_k$ for $j< i$. As a result, the relation between the rate of a dependent layer and its own quantization step size can be written as in [23], yielding:

$$R_{i,k}(q_{0,k},\ q_{1,k},\ ...,\ q_{i,k}) \approx R_{i,k}(q_{i,k}). \tag{4}$$

Second, the distortion of a dependent scalable block in TL-*i* and QL-*k* was derived analytically as a linear sum of the distortion functions of its reference layer TL-0 and QL-*k* and can mathematically be expressed as:

$$D_{i,k}(q_{i,0},\ q_{i,1},\ \ldots,\ q_{i,k}) = \sum_{j=0}^{i} \zeta_{i,j}^{k}\ D_{0,k}(q_{j,k}), \tag{5}$$

where $\zeta_{i,j}^{k}$'s are model parameters which show how the TL-*j* contributes to the distortion of the current layer, TL-*i*, where $j \le i$. More details on calculating these parameters can be found in [23].

For the R-D characteristics of a scalable block in quality scalability, it was observed in [23] that the distortion of a scalable block is independent to that of the preceding scalable blocks in the QL references, i.e., $T_i Q_j$ for $j< k$ and is given as: $D_{i,k}(q_{i,0},\ q_{i,1},\ \ldots,\ q_{i,k}) \approx D_{i,k}(q_{i,k})$. Moreover, the QL distortion dependency of a scalable block at a temporal layer TL-0 and a quality layer QL-*k* is strongly correlated with the distortion of its reference quality layer QL-0. Therefore, in this case the distortion function of a scalable block in temporal layer $T_0$ and quality layer $Q_k$ can be simplified to:

$$D_{0,k}(q_{i,0},\ q_{i,1},\ \ldots,\ q_{i,k}) \approx D_{0,k}(q_{i,k}) = \mu_0^k\ D_{0,0}(q_{i,k}), \tag{6}$$

where $\mu_0^k$ is the distortion model parameter of $T_0 Q_k$. Furthermore, the rate function of that scalable block at QL-*k* can be expressed as:

$$R_{i,k}(q_{i,0},\ q_{i,1},\ \ldots,\ q_{i,k}) = \sum_{j=0}^{k} \sigma_i^{k,j}\ R_{i,0}(q_{i,j} + j\,\Delta) + \eta_i^k, \tag{7}$$

where $R_{i,0}$ is a texture (residual) rate of TL-*i* and QL-0, which is a function of $q_{i,j}$ the quantization step size of $T_iQ_j$, and $\Delta$ is a predefined constant that represents the difference between the quantization parameters of two consecutive TLs and QLs. $\Delta$ was set to two in our experiments as recommended in [23]. $\sigma_i^{k,j}$ and $\eta_i^k$ are the rate model parameters of a scalable block at TL-*i* and QL-*k* and are named Cho rate model parameters which can be expressed as:

$$\sigma_i^{k,0} = m_{i,k},\ \ \sigma_i^{k,j} = m_{i,k-j} - m_{i,k-j+1},\ \ \forall j \in \{1,\cdots,\ k\}, \tag{8}$$

$$\eta_i^k \approx R_{i,k}(q_{i,0},\ q_{i,0} - \Delta, \cdots, q_{i,0} - k\,\Delta) - m_{i,0}\ R_{i,0}(q_{i,0}), \tag{9}$$

where $m_i$'s represent the slopes of the rate model lines passing through the pivot points. More details on identifying the pivot points and calculating the slopes of the rate model lines can be found in [23]. The quality layer QL-*k* gives $k+1$ slopes of $m_0, m_1, ..., m_k$. Now, without considering the influences of preceding TL and QL blocks, the rate and quantization relation of a joint temporal-quality scalable block is the same as equation (7) whereas from (5) and (6), the relation between the distortion and quantization can be simplified to:

$$D_{i,k}(q_{i,0},\ q_{i,1},\ \ldots,\ q_{i,k}) = \mu_0^k \sum_{j=0}^{i} \zeta_{i,j}^{k}\ D_{0,0}(q_{j,k}), \tag{10}$$

where $D_{0,0}(q_{j,k})$ is a residual distortion function of $T_0Q_0$, $q_{j,k}$ is the quantization step size and $\zeta_{i,j}^{k}$ is a distortion model parameter of $T_iQ_k$. Based on the rate and distortion models in (7) and (10) respectively, the unconstrained optimization problem in (3) can be rewritten as:

$$\begin{aligned} J(\mathbf{Q},\lambda) = & \sum_{k=0}^{N_Q-1} \sum_{i=0}^{N_T-1} \omega_{i,k}\ D_{0,0}(q_{i,k}) + \\ & \lambda\Big(\sum_{k=0}^{N_Q-1} \sum_{i=0}^{N_T-1} \Big(\sum_{j=0}^{k} \sigma_i^{k,j}\ R_{i,0}(q_{i,j} + j\,\Delta) + \eta_i^k\Big) - R^T\Big), \end{aligned} \tag{11}$$

where $\omega_{i,k}$ is the model parameter which is given by $\omega_{i,k} = \mu_0^k \sum_{j=i}^{N_T-1} \zeta_{j,i}^{k}$. It is named here Cho distortion model parameters.

## III. Single-Pass Joint T-Q Layer Bit Allocation

The bit allocation problem of joint temporal-quality scalability in (11) assumes all the quality layers have the same temporal resolution (i.e., $N_T$ is the same for all QLs). However, to control the extra bit rate of the SVC over the single layer encoder one may assign different frame rates at each quality layer as shown in Fig. 2. In this figure, the combined T-Q scalability plane has three quality layers QLs, and each quality layer has different temporal layer numbers. This plane is composed of two sub-planes; square/rectangle plane based on the number of quality layers and triangle plane. In the square/rectangle plane, each quality layer has the same number of temporal layers while the triangle plane contains the remaining number of the temporal layers. Arrows demonstrate the prediction dependency among the coding units. In this case the Lagrangian cost function in (11) should be modified to include the triangle plane. Therefore the global optimal bit allocation problem for joint temporal-quality scalability with the same or with different temporal resolutions at each quality layer can be formulated as:

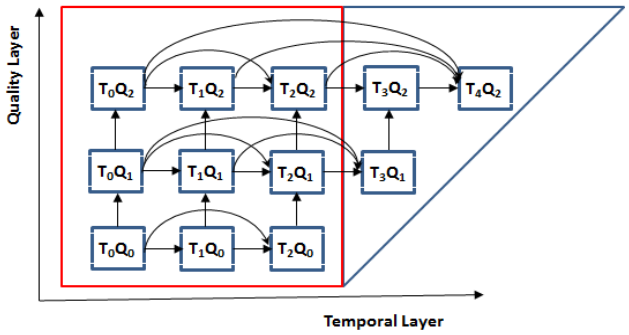



Fig. 2. H.264/SVC layer structure with combined T-Q scalability of three QLs and different temporal resolutions at each QL.

$$J(\mathbf{Q},\lambda)=\sum_{k=0}^{N_Q-1}\sum_{i=0}^{n_T[k]-1}\omega_{i,k}\ D_{0,0}(q_{i,k})+\lambda(R_1(\mathbf{Q})+s\cdot R_2(\mathbf{Q})-R^T),\quad(12)$$

where $n_T[k]$ is the $k$-th element of vector $\mathbf{n_T}$ which has $N_Q\times 1$ elements. Each element represents the number of temporal layers at each quality layer QL-$k$. $s$ is a switching factor set to one when the quality layers have different temporal resolutions and otherwise it is set to zero. $R_1(\mathbf{Q})$ and $R_2(\mathbf{Q})$ represent the rates in the square/rectangle plane and the triangle plane respectively and are given by:

$$R_1(\mathbf{Q})=\sum_{k=0}^{N_Q-1}\sum_{i=0}^{n_T[0]-1}(\sum_{j=0}^{k}\sigma_i^{k,j}\ R_{i,0}(q_{i,j}+j\Delta)+\eta_i^k),\quad\text{and}$$

$$R_2(\mathbf{Q})=\sum_{k=1}^{N_Q-1}\sum_{i=n_T[0]}^{n_T[k]-1}(\sum_{j=l}^{k}\sigma_i^{k-l,j-l}\ R_{i,l}(q_{i,j}+j\Delta)+\eta_i^{k-l}),\quad(13)$$

where factor $l$ is set as follows:

$$l=\begin{cases}\lceil i/2\rceil, & \text{if } k=N_Q-1 \text{ and } i=n_T[k]-1,\\ \lfloor i/2\rfloor, & \text{otherwise,}\end{cases}$$

$\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denote the floor and ceiling functions which map a real number to the largest previous or the smallest following integer, respectively. The rate and distortion in (12) can be described using the Cauchy distribution-based R-Q and D-Q models [7], respectively, due to its reported superior performance to other models. The R-D models for a scalable block at TL-$i$ and QL-$k$ are formalized in [7] as:

$$R_{i,k}(q_{i,k})=a_{i,k}\cdot q_{i,k}^{-\alpha_{i,k}}\quad\text{and}\quad D_{i,k}(q_{i,k})=b_{i,k}\cdot q_{i,k}^{\beta_{i,k}},\quad(14)$$

where $a_{i,k}$, $b_{i,k}$, $\alpha_{i,k}$ and $\beta_{i,k}$ are model parameters. The overall distortion in (12) is based on the distortion of the scalable block at TL-0 and QL-0, which can be formulated using (14) as $D_{0,0}(q_{i,k})=b_{0,0}\cdot q_{i,k}^{\beta_{0,0}}$. Moreover, since the rate of a scalable block in the square plane is based on the rate of the scalable block at TL-$i$ and QL-0 (i.e. $i<n_T[0]$), it can be expressed as $R_{i,0}(q_{i,j})=a_{i,0}\cdot q_{i,j}^{-\alpha_{i,0}}$. The rate of a scalable block in the triangle plane (i.e. $i\ge n_T[0]$) can be expressed as $R_{i,l}(q_{i,j})=a_{i,l}\cdot q_{i,j}^{-\alpha_{i,l}}$. For simplicity, instead of representing both a and $\alpha$ as a two dimensional matrix, they can be represented as 1D vector of length $n_T[N_Q-1]$ (i.e. for example $a_{i,0}$ (for $i<n_T[0]$) and $a_{i,l}$ (for $i\ge n_T[0]$) are replaced by $a_i$ (for $0\le i\le n_T[N_Q-1]-1$). Also, since both of $b_{0,0}$ and $\beta_{0,0}$ are one value, they are represented by $b$ and $\beta$, respectively. Furthermore, $\Delta$ is not indicated in $R_{i,0}(q_{i,j})$ and $R_{i,l}(q_{i,j})$ because the notation $q_{i,j}+j\Delta$ means $q_{i,j}(QP_{i,j}+j\Delta)$ which is the one-to-one mapping between the quantization step-size and the quantization parameter. The rate and distortion can then be rewritten as:

$$R_{i,0}(q_{i,j})\ (\text{or } R_{i,l}(q_{i,j}))=a_i\cdot q_{i,j}^{-\alpha_i}\quad\text{and}$$

$$D_{0,0}(q_{i,k})=b\cdot q_{i,k}^{\beta},\quad(15)$$

where $(a_i,\alpha_i)$ are the Cauchy rate model parameters and $(b,\beta)$ are the Cauchy distortion model parameters. Using (13) and (15) the optimization problem in (12) becomes:

$$\begin{aligned}J(\mathbf{Q},\lambda)=&\sum_{k=0}^{N_Q-1}\sum_{i=0}^{n_T[k]-1}\omega_{i,k}\cdot b\cdot q_{i,k}^{\beta}+\\&\lambda(\sum_{k=0}^{N_Q-1}\sum_{i=0}^{n_T[k]-1}(\sum_{j=0}^{k}\sigma_i^{k,j}\cdot a_i\cdot q_{i,j}^{-\alpha_i}+\eta_i^k)+\\&s\cdot(\sum_{k=1}^{N_Q-1}\sum_{i=n_T[0]}^{n_T[k]-1}(\sum_{j=l}^{k}\sigma_i^{k-l,j-l}\cdot a_i\cdot q_{i,j}^{-\alpha_i}+\eta_i^{k-l}))-R^T).\end{aligned}\quad(16)$$

By taking the partial derivative of the cost function, $J$, with respect to $q_{i,k}$'s and $\lambda$, and setting the result of derivative to zero, it yields $(\sum_{k=0}^{N_Q-1}n_T[k])+1$ nonlinear equations. Mathematically, we have:

$$\frac{\partial J(\mathbf{Q},\lambda)}{\partial q_{i,k}}=\omega_{i,k}\cdot b\cdot\beta\cdot q_{i,k}^{\beta-1}+\lambda(\frac{\partial R_1(\mathbf{Q})}{\partial q_{i,k}}+s\cdot\frac{\partial R_2(\mathbf{Q})}{\partial q_{i,k}})=0\ \text{and}$$

$$\begin{aligned}\frac{\partial J(\mathbf{Q},\lambda)}{\partial\lambda}&=R_1(\mathbf{Q})+s\cdot R_2(\mathbf{Q})-R^T\\&=\sum_{k=0}^{N_Q-1}\sum_{i=0}^{n_T[k]-1}(\sum_{j=0}^{k}\sigma_i^{k,j}\cdot a_i\cdot q_{i,j}^{-\alpha_i}+\eta_i^k)+\\&\quad s\cdot(\sum_{k=1}^{N_Q-1}\sum_{i=n_T[0]}^{n_T[k]-1}(\sum_{j=l}^{k}\sigma_i^{k-l,j-l}\cdot a_i\cdot q_{i,j}^{-\alpha_i}+\eta_i^{k-l}))-R^T\\&=0,\end{aligned}\quad(17)$$

where

$$\frac{\partial R_1(\mathbf{Q})}{\partial q_{i,k}}=-\sum_{j=k}^{N_Q-1}\sigma_i^{k,j}\cdot a_i\cdot\alpha_i\cdot q_{i,j}^{-\alpha_i-1}$$

and

$$\frac{\partial R_2(\mathbf{Q})}{\partial q_{i,k}} = -\sum_{j=l}^{N_Q-1} \sigma_i^{k-l,j-l} \cdot a_i \cdot \alpha_i \cdot q_{i,j}^{-\alpha_i-1} \ . \quad (18)$$

These nonlinear equations can be solved using any numerical method to determine the values of $q_{i,k}$'s. It should be mentioned that the convergence of numerical methods such as Newton's cannot be guaranteed in general, since it depends on many factors such as the nature of the involved objective and constraint functions, the number of variables and the used constraints [32]. In this paper, Newton method was used to determine the values of $q_{i,k}$'s in each GOP. If the method does not converge at a certain GOP, the values of $q_{i,k}$'s are set to those obtained from the previous GOP. Implementing this algorithm requires multiple pre-encoding passes (several iterations) to derive the model parameters for each video sequence. In the following an adaptive model-parameter initialization scheme is proposed for joint temporal-quality scalability in (17) to convert this multi-pass algorithm to a single-pass algorithm.

### *A. Model Parameters initialization*

In this paper, the focus is on single-pass implementation of the bit allocation algorithm in which there is no prior information about the statistical properties of the input video sequence. To solve the nonlinear equations in (17), two categories of R-D model parameters, which are not known, need to be estimated. The first category is the Cauchy rate model parameters $(a_i, \alpha_i)$ and the distortion model parameters $(b, \beta)$. The second category is Cho rate model parameters $(\sigma_i^{k,j}, \eta_i^k)$ and the distortion model parameters $(\omega_{i,k})$. The first stage in the proposed adaptive model-parameter is the estimation of the appropriate initial values of these two categories of R-D model parameters. In order to estimate the suitable initial values of R-D model parameters of Cauchy and Cho, several experiments were conducted on various video sequences in quarter common intermediate format (QCIF), common intermediate format (CIF) and 4CIF. Twenty five video sequences were used in these experiments selected from the databases in [33, 34]. Moreover, two test scenarios were taken into account: Scenario I, two quality layers and the number of temporal layers were equal at each quality layer and was set to three (i.e., $n_T[0] = n_T[1] = 3$ and $s$ was set to zero). Scenario II, two quality layers and the number of temporal layers were different at each quality layer, we set $n_T[0] = 3$, $n_T[1] = 4$ and $s$ was set to one. More details about the simulation parameters used to estimate the model parameters are given in the experimental section (Section IV).

First we explain the initialization of Cauchy rate and distortion model parameters. Since the rate of a scalable block at TL-*i* and QL-*k* is dependent on the rate of the block at TL-*i* and QL-0, the parameters $(a_i, \alpha_i)$ are obtained from QL-0. For the distortion, it is based on the distortion generated from TL-0 and QL-0, so there are only two parameters $(b, \beta)$ that need to be defined. The rate model parameter $\alpha_i$ was restricted to two sets of predefined constant values; one set is identified for the reference scalable block at TL-0 (i.e., this set includes the values of $\alpha_0$, $i$=0) and the second set for the dependent scalable blocks at TL-*i*, $i$>0 (i.e., $i \in \{1, \cdots, n_T[N_Q - 1] - 1\}$). The values of these sets were empirically obtained and given as:

$$\alpha_0 = \begin{cases} 1.0, & \text{if } R_{0,0} / N_p > 0.2 \\ 1.2, & \text{if } R_{0,0} / N_p < 0.07 \text{ and} \\ 1.4, & \text{otherwise} \end{cases}$$

$$\alpha_i = \begin{cases} 1.2, & \text{if } R_{i,0} / N_p > 0.05, \\ 1.3, & \text{if } R_{i,0} / N_p < 0.01, \quad i > 0, \\ 1.8, & \text{otherwise,} \end{cases} \quad (19)$$

where $N_p$ is the number of pixels per frame. The distortion model parameter $\beta$ was set to 1.4. After encoding the first GOP using the initial quantization parameters defined at each temporal and quality layer, the output bits $R_{i,0}$ and distortion $D_{0,0}$ resulting from encoding the scalable blocks at QL-0 can be obtained. The values of $\alpha_i$ are then chosen from the sets given in (19). According to the Cauchy R-Q and D-Q models in (15), the complexity measure $a_i$ for a scalable block in the $i$th temporal layer and QL-0 and the parameter $b$ for the scalable block in TL-0 and QL-0 can be respectively derived as:

$$a_i = R_{i,0}(q_{i,0}) \cdot q_{i,0}^{\alpha_i} \text{ and } b = D_{0,0}(q_{0,0}) \cdot q_{0,0}^{-\beta}. \quad (20)$$

Second, the initial values of Cho rate and distortion model parameters were also estimated. The Cho distortion model parameters $(\omega_{i,k})$ represent the contribution of each TL scalable block distortion on the overall GOP distortion. The empirical values of the $\omega$'s parameters for two quality layers that we were concerned in the experiments are given for the first- and second- quality layer respectively as:

$$\omega_{i,0} = \begin{cases} (2.753,\ 0.223,\ 0.086), & \text{for QCIF} \\ (1.901,\ 0.435,\ 0.177), & \text{for CIF} \\ (2.592,\ 0.342,\ 0.157), & \text{for 4CIF} \end{cases}$$

and

$$\omega_{i,1} = \begin{cases} (2.590,\ 0.253,\ 0.116,\ 0.201), & \text{for QCIF} \\ (2.378,\ 0.430,\ 0.227,\ 0.206), & \text{for CIF} \\ (2.299,\ 0.458,\ 0.242,\ 0.193), & \text{for 4CIF.} \end{cases} \quad (21)$$

Finally, Cho rate model parameters $(\sigma_i^{k,j}, \eta_i^k)$ were also estimated from the experiments mentioned above. The initial values of $\sigma_i^{k,j}$ at each TL were estimated as: $\sigma_0^{0,0} = 1$,

$\sigma_0^{1,0} = -0.747$, $\sigma_0^{1,1} = 2.092$, $\sigma_1^{0,0} = 1$, $\sigma_1^{1,0} = 0.267$, $\sigma_1^{1,1} = 1.218$, and $\sigma_2^{0,0} = 0.359$, $\sigma_2^{1,0} = 0.359$, $\sigma_2^{1,1} = 1.252$. From these values, it can be seen that $\sigma_i^{k,j}$ has values greater than one for $j > 0$ and equal to one for $j = k = 0$ (i.e., $\sigma_i^{0,0} = 1$). To find $\eta_i^k$, it is worth mentioning that the rate of a certain temporal layer TL is the sum of the rates of all QL blocks and that rate is dependent on the rate of the reference QL block $R_{i,0}$. According to (7), $\sum_{j=0}^{k} \eta_i^k$ represents the amount of overhead bit rate due to having QLs. Since parameters $\eta_i^k$'s are evaluated based on the rate slopes $m_{i,0}$'s as in (9), the initial values of $m_{i,0}$'s are given and set to $m_{0,0} = 1.378$, $m_{1,0} = 1.724$, and $m_{2,0} = 2.005$. In other words, after encoding the first GOP, the output bits $R_{i,0}$ and $R_{i,1}$ resulting from encoding the scalable blocks at each temporal and quality layer can be accessible. Substituting the above initial values of $m_{i,0}$'s, $R_{i,0}$ and $R_{i,1}$ in (9) $\eta_i^1$ can be calculated and used for encoding the next GOP.

To further verify the accuracy of these estimated initial model-parameters, the proposed single-pass RC with these initial model-parameters is compared with Cho fixed-Rate multiple-pass RC algorithm. Fig. 3 shows the performance of the proposed algorithm using the estimated rate and distortion parameters described in this subsection during encoding of the video sequences. It can be seen that the proposed algorithm with the suggested parameters in encoding a video sequence with only one iteration provides comparable PSNR performance to Cho fixed-Rate algorithm using the parameters obtained from encoding each video sequence with several iterations.

It is worth noting that rate and distortion models parameters of Cauchy and Cho are empirically estimated by encoding each video sequence several times using the multi-pass RC algorithm in (17). To get rate model parameters for Scenario I, three rate pivot points ($(q_{0,0}, q_{0,0} - \Delta)$, $(q_{0,1} + \Delta, q_{0,1})$ and $(q_{0,0}, q_{0,1})$) are generated by actual encoding each video sequence three times, one at each pivot point. Using the output bits resulting from encoding a video sequence at the pivot points, the slopes of the rate model lines for each quality layer ($m_{i,0}$ and $m_{i,1}$) are calculated as in [23]. Cho rate model parameters ($\sigma_i^{k,j}$, $\eta_i^k$) are evaluated from these slopes by using (8) and (9). The averaged values of the rate slopes $m_{i,0}$'s and $\sigma_i^{k,j}$ for all video sequences yield the initial values of those parameters. The parameters $(\alpha_i)$ can also be calculated and their initial values can then be estimated by analysing and classifying the obtained values of $\alpha_i$ for all sequences into two sets based on the output bits as indicated in (19). To get the parameters $(\omega_{i,k})$, the slopes of the distortion model lines for each quality layer are calculated at three distortion pivot points as in [23]. These slops are used to calculate $\zeta_{i,j}^k$ and $\omega_{i,k}$. The obtained values of $\omega_{i,k}$ for all video sequences are classified into two sets as in (21) based on the format of video sequences. The initial values of some model parameters may not be close to those obtained from multiple-pass RC algorithm for some test sequences. This drawback is compensated by adaptively updating some of the model parameters during the encoding process.

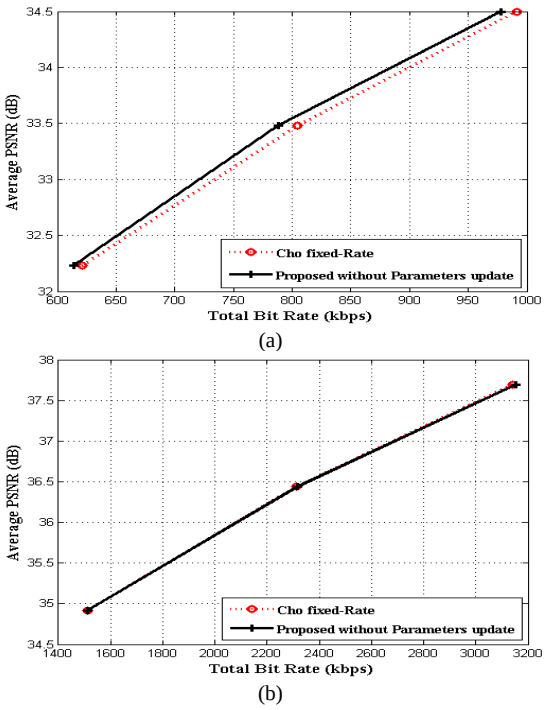



Fig. 3. Coding efficiency comparison (Average Y-PSNR versus Rate) of the proposed RC with initial model-parameter determination method and Cho fixed-Rate, multiple-pass RC algorithm in Scenario I for different spatial complexity sequences. (a) "*Football*". (b) "*Soccer*".

### *B. Updating the Model Parameters*

Using constant model parameters during encoding a video sequence cannot reflect the changes that may occur from GOP to GOP in a video sequence. For better bit allocation strategy it is desired to adapt the model parameters $a_i$, $b$ and $m_{i,0}$'s for each GOP. In the second stage of the adaptive model-parameter initialization scheme, these model parameters are predicted from the parameters of the previously encoded GOPs. $a_i$ and $b$ can be updated in the encoding process using the following linear or weighted average of $\hat{a}_i(n-1)$ and $\hat{b}(n-1)$ with $\hat{a}_i(n-1)$ and $\hat{b}(n-1)$ as:

$$a_i(n) = w \cdot \hat{a}_i(n-1) + (1-w) \cdot \hat{a}_i(n-1), \quad (22)$$

$$b(n) = w \cdot \hat{b}(n-1) + (1-w) \cdot \hat{b}(n-1), \quad (23)$$

where $w$ is the weighting parameter which was set to 0.5 in our experiments. $\hat{a}_i(n-1)$ and $\hat{b}(n-1)$ are the actual values

obtained from (20) for the last coded $(n-1)$ th GOP. $\bar{a}_i(n-1)$ and $\bar{b}(n-1)$ are average values of the $a_i$ 's and $b$ 's predicted so far from the previously encoded GOPs using recursive form as:

$$\bar{a}_i(n-1) = [(n-2)\cdot \bar{a}_i(n-2) + a_i(n-2)]/(n-1).$$
$$\bar{b}(n-1) = [(n-2)\cdot \bar{b}(n-2) + b(n-2)]/(n-1). \quad (24)$$

Since the rate model parameters $\eta_i^k$ 's are determined based on the slopes $m_{i,0}$ and the values of $R_{i,k}(q_{i,0},\ q_{i,0}-\Delta,\cdots,q_{i,0}-k\Delta)$ and $R_{i,0}(q_{i,0})$, steeper slope with the value of $R_{i,k}$ greater than $R_{i,0}$ indicates higher parameter values of $\eta_i^k$ 's while steeper slope with the value of $R_{i,k}$ smaller than $R_{i,0}$ indicates the absolute values of $\eta_i^k$ 's are smaller. Considering the distribution of the actual DCT coefficients of various frames in different sequences or even of different quality layers in the same sequence significantly varies, it is required to update the $m_{i,0}$'s from a GOP to the next. According to (9), $m_{i,0}(n)$ of the $n$th GOP can be derived after encoding the corresponding GOP as:

$$m_{i,0}(n) = (R_{i,k}(q_{i,0},\ q_{i,0}-\Delta,\cdots,q_{i,0}-k\Delta) - \eta_i^k(n))/R_{i,0}(q_{i,0}). \quad (25)$$

However, the actual values of $m_{i,0}(n)$ 's cannot be directly derived from (25) since $R_{i,k}$, $R_{i,0}$ and $\eta_i^k(n)$ are inaccessible until the encoding of the $n$th GOP is completed. Thus the values of $m_{i,0}(n)$ 's are predicted as follows:

$$m_{i,0}(n) = w\cdot m_{i,0}(n-1) + (1-w)\cdot \hat{m}_{i,0}(n-1), \quad (26)$$

where $\hat{m}_{i,0}(n-1)$ is the actual value obtained from (25) for the last coded $(n-1)$ th GOP. $m_{i,0}(n)$ and $m_{i,0}(n-1)$ are the current and the previous prediction values of the slopes of the rate model lines. Also $\eta_i^k(n)$ is calculated as $\eta_i^k(n) = \left|R_{i,k}(q_{i,0},\ q_{i,0}-\Delta,\cdots,q_{i,0}-k\Delta) - m_{i,0}(n\text{-}1)\cdot R_{i,0}(q_{i,0})\right|$. Once a GOP is coded, the actual values of rate and distortion for that GOP are calculated. The actual values of $a_i$, $b$ and $m_{i,0}$'s parameters are also calculated and used to update their prediction values (and to calculate the quantization step sizes) for encoding the remaining GOPs. The corresponding quantization parameters $QP_{i,k}(n)$'s for the $n$th GOP are determined using the one-to-one relationship between the quantization step-size and the quantization parameter [29].

## IV. Experimental Results

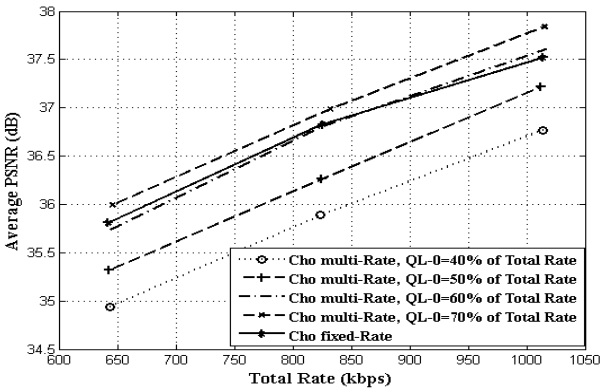
The proposed single pass bit allocation algorithm was implemented in the SVC reference software JSVM 9.19.14 [3]. To evaluate the performance of the algorithm, several experiments were performed on various video sequences in QCIF, CIF, and 4CIF. In these experiments, two test scenarios are considered. For Scenario I, two quality layers and three temporal layers (i.e., GOP size is four) are utilized whereas for Scenario II, two quality layers having different number of temporal layers are used, where the GOP sizes at QL-0 and QL-1 are four and eight respectively. For both scenarios, at QL-0, every TL-0 frame is encoded as a P-frame except for the first frame of video sequences which are coded as I-frame. Furthermore, QL-1 is encoded using adaptive inter-layer prediction from QL-0. The initial values of quantization parameters are set to 32 and 30 at the QL-0 and QL-1, respectively. Some of the simulation parameters are given in Table I and the other parameters are set as defaults of the reference software.

TABLE I
SUMMARY OF SOME SIMULATION PARAMETERS.

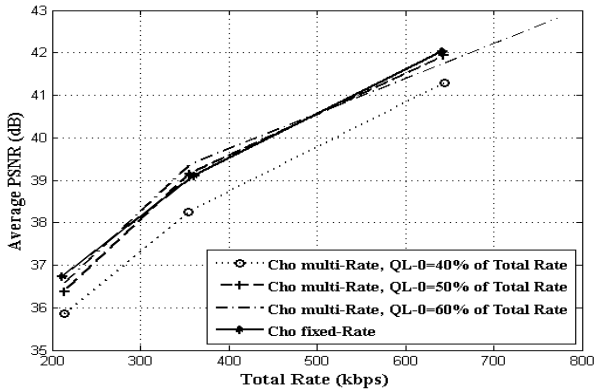
| | | QCIF | CIF | 4CIF |
|---|---|---|---|---|
| Base Layer Mode | | AVC Compatible | | |
| Intra Period | | -1 | | |
| Reference no. | | 1 | | |
| Symbol Mode | | CABAC | | |
| Resolution | | 176×144 | 352×288 | 704×576 |
| Scenario I, frame rate | QL-0, 3TLs | 15 | 30 | 30 |
| | QL-1, 3TLs | 15 | 30 | 30 |
| Scenario II, frame rate | QL-0, 3TLs | 7.5 | 15 | 30 |
| | QL-1, 4TLs | 15 | 30 | 60 |

### A. Comparison Between Multi-Rate and Fixed-Rate Bit Allocation Strategies

Before assessing the performance of the proposed bit allocation algorithm, we compare between the two bit allocation strategies discussed in Section II. It should be mentioned that the rate control algorithm in [23] here is named Cho multi-Rate strategy. To compare the coding performance of these two bit allocation strategies for the joint temporal-quality scalability of H.264/SVC, the algorithm in [23] was modified to employ the fixed-Rate strategy and is named Cho fixed-Rate. Here, we apply Scenario I on two video sequences with low to high spatial details, "*News*" and "*Crew*". In the multi-Rate strategy, various percentages of the overall bit budgets, which were given to the fixed-Rate strategy, were allocated to QL-0 and the remaining to QL-1. The effect of distributing various percentages of the constrained overall bit budget on the performance of Cho multi-Rate as compared with that of Cho fixed-Rate is shown in Fig. 4. It can be seen from this figure, in the Cho multi-Rate scheme as the percentage of bits assigned to QL-0 increases, the overall performance of the Cho multi-Rate improves up to a certain point, and any increase in QL-0 bit rate beyond this point will be wasted. This is because the assigned bit rate budget to QL-1 will decrease and hence the overall quality will not be improved further. This point is clear for the "*News*" sequence when the percentage of the total bit budget assigned to QL-0 increases especially at 50 and 60 percentages, the averaged PSNR does not increase significantly. This means when the bit budgets are not appropriately assigned among the quality layers, the overall optimal R-D performance may not be achieved. For Cho fixed-Rate, the total bit budget is adaptively distributed among the quality and the temporal layers by




(a)




(b)

Fig. 4. Performance comparison between multi-Rate and fixed-Rate bit allocation schemes for the averaged Y-PSNR values of QL-0 and QL-1 as a function of the obtained total bit rate. (a) "*Crew*" and (b) "*News*".

taking into account the dependency among these layers and the characteristics of the video sequences, that is, the residual information in each scalable block. Also, it can be seen that the performance of the Cho fixed-Rate for the two sequences is comparable to that of Cho multi-Rate when 60 percentage of the total bit budget is allocated to QL-0. This is due to the constraint regarding delta quantization parameter ( $\Delta$ ) that we used in the experiments and was set to two.

Furthermore, Fig. 5 illustrates the percentage of bit rate distributed to QL-0 over different bit rates for various video sequences. This figure demonstrates the importance of the fixed-Rate strategy for allocating the total bit budget among the layers, where bit rate allocated to QL-0 varies depending on the total bit budget and the characteristics of the test video sequence. From the above discussion, we concluded to use the fixed-Rate strategy in this paper.

### *B. RD Performance*

The performance of the proposed algorithm was compared with the benchmark rate-control algorithms of the reference JSVM FixedQPEncoder tool and Cho fixed-Rate which are multiple-pass algorithms. To employ the FixedQpEncoder tool, an initial QP and a target bit rate are first assigned to each quality layer. Then, the encoder uses different values of QP to encode a sequence, a QP value in each iteration. The value of generated bit rate is then fed back to the next iteration to adapt the quantization parameter. This search algorithm

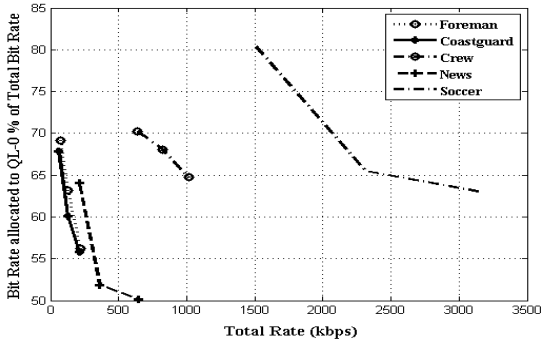



Fig. 5. The percentage of the target bit rate distribution to QL-0 using Cho fixed-Rate bit allocation for various test video sequences.

terminates when the obtained bit rate falls within an acceptable mismatch range of the target bit rate or the number of encoding iterations exceeds the maximum number of iterations ($N_{\max}$). In the experiments, $N_{\max}$ was set to 15 and the maximum negative and positive mismatch were set to 2%. Since the FixedQpEncoder is a multi-Rate RC tool, the target bit rate for each quality layer should be predefined before the encoding process. For both of these two RC algorithms, including Cho fixed-Rate and the proposed algorithm, a total bit budget $R^T$ is given and distributed adaptively among the temporal and quality layers as explained in Section IV-A. Consequently regarding the FixedQpEncoder, the obtained bit rates ( $R^O$ ) from QL-0 using Cho fixed-Rate and the proposed algorithm are averaged and this average value ( $R^{avg}$ ) after rounding it to nearest integer is assigned to QL-0 of the FixedQPEncoder while the total bit budget $R^T$ is assigned to QL-1.

The R-D results of the proposed algorithm, Cho fixed-Rate, and JSVM FixedQPEncoder algorithms in terms of average Y-PSNR and the obtained bit rates ( $R^O$ ) at each QL are summarized in Tables II-VI for Scenario I and Scenario II, where "**QL**" and "PSNR" indicate the quality layer number and the average Y-PSNR obtained at each QL. "Avg" and "Average" represent the average value of the results obtained at each quality layer and from the test sequences, respectively. In these tables, the obtained bit rate at QL-0 of the proposed and Cho fixed-Rate algorithms indicates the bit rate results from the distribution of a percentage of $R^T$ to that layer while the obtained bit rate ( $R^O$ ) at QL-1 indicates the total bit rate resulting from encoding both layers (the full T-Q resolution). For FixedQPEncoder algorithm, the obtained bit rates at QL-0 and QL-1 represent the bit rates resulting from encoding these layers using the allocated bit rates ( $R^{avg}$ ), and $R^T$ to QL-0 and QL-1, respectively. As seen the proposed bit allocation algorithm provides better performance than the two algorithms of Cho fixed-Rate and JSVM for the most test sequences. It achieves an averaged Y-PSNR gain of about 0.28-0.39 dB over Cho's algorithm. The reason behind the good performance of the proposed bit allocation algorithm is due to the proposed adaptive model-parameter initialization mechanism to initialize and update the model parameters for temporal and quality layers. Unlike the Cho fixed-Rate

TABLE II.
PERFORMANCE COMPARISON OF THE PROPOSED ALGORITHM, JSVM, AND CHO FIXED-RATE RC ALGORITHMS IN **SCENARIO I FOR QCIF SEQUENCES**.

| Seq. | $R^T$ (kb/s) | QL | JSVM 9.19.14 [3] | | | | Cho fixed-Rate | | | | Proposed | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $R^O$ (kb/s) | PSNR (dB) | E (%) | Iter | Rate (kb/s) | PSNR (dB) | E (%) | Iter | Rate (kb/s) | PSNR (dB) | E (%) | Iter |
| Coastguard | 128 | 0 | 81.63 | 34.23 | 0.79 | 2 | 75.40 | 33.95 | -2.0 | 7 | 85.93 | 34.46 | -2.0 | 1 |
| | | 1 | 124.01 | 35.40 | -3.1 | 7 | 125.38 | 35.40 | | | 125.42 | 35.53 | | |
| | | Avg | | 34.81 | | | | 34.68 | | | | 34.99 | | |
| | 216 | 0 | 146.0 | 37.04 | -1.9 | 9 | 119.35 | 35.90 | -0.9 | 7 | 178.62 | 37.78 | -2.1 | 1 |
| | | 1 | 213.5 | 38.22 | -1.1 | 5 | 214.00 | 37.98 | | | 211.40 | 38.36 | | |
| | | Avg | | 37.63 | | | | 36.94 | | | | 38.07 | | |
| Foreman | 64 | 0 | 46.12 | 32.81 | -1.8 | 3 | 49.02 | 33.05 | 10.7 | 7 | 44.81 | 32.72 | -2.2 | 1 |
| | | 1 | 67.93 | 33.78 | 6.14 | 8 | 70.89 | 34.18 | | | 62.61 | 33.52 | | |
| | | Avg | | 33.29 | | | | 33.62 | | | | 33.12 | | |
| | 128 | 0 | 79.38 | 36.11 | -1.9 | 5 | 80.19 | 36.00 | -0.8 | 7 | 81.82 | 36.10 | -2.9 | 1 |
| | | 1 | 131.17 | 37.80 | 2.4 | 7 | 126.97 | 37.54 | | | 124.18 | 37.44 | | |
| | | Avg | | 36.95 | | | | 36.77 | | | | 36.77 | | |
| Carphone | 32 | 0 | 21.60 | 34.00 | -1.8 | 3 | 23.50 | 33.74 | -5.9 | 7 | 20.04 | 33.59 | -1.3 | 1 |
| | | 1 | 31.04 | 34.54 | -2.9 | 10 | 30.12 | 34.12 | | | 31.59 | 34.52 | | |
| | | Avg | | 34.27 | | | | 33.93 | | | | 34.05 | | |
| | 64 | 0 | 39.16 | 37.41 | 0.4 | 3 | 38.67 | 36.82 | -6.6 | 7 | 39.79 | 37.43 | -2.7 | 1 |
| | | 1 | 62.95 | 38.50 | -1.6 | 5 | 59.75 | 37.92 | | | 62.28 | 38.53 | | |
| | | Avg | | 37.95 | | | | 37.37 | | | | 37.98 | | |
| **Average** | | | | **35.82** | **2.87** | **11.2** | | **35.55** | **4.48** | **7** | | **35.83** | **2.20** | **1** |

TABLE III
PERFORMANCE COMPARISON OF THE PROPOSED ALGORITHM, JSVM, AND CHO FIXED-RATE RC ALGORITHMS IN **SCENARIO I FOR CIF SEQUENCES**.

| Seq. | $R^T$ (kb/s) | QL | JSVM 9.19.14 [3] | | | | Cho fixed-Rate | | | | Proposed | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Rate (kb/s) | PSNR (dB) | E (%) | Iter | Rate (kb/s) | PSNR (dB) | E (%) | Iter | Rate (kb/s) | PSNR (dB) | E (%) | Iter |
| Football | 648 | 0 | 510.58 | 31.79 | 1.9 | 5 | 496.88 | 31.87 | -4.1 | 7 | 507.44 | 31.95 | -3.1 | 1 |
| | | 1 | 629.95 | 32.50 | -2.7 | 15 | 621.42 | 32.59 | | | 627.87 | 32.66 | | |
| | | Avg | | 32.14 | | | | 32.23 | | | | 32.31 | | |
| | 832 | 0 | 640.1 | 33.05 | 1.1 | 3 | 623.54 | 33.03 | -3.3 | 7 | 643.61 | 33.20 | -2.7 | 1 |
| | | 1 | 810.6 | 33.81 | -2.5 | 15 | 804.53 | 33.93 | | | 809.73 | 33.99 | | |
| | | Avg | | 33.43 | | | | 33.48 | | | | 33.59 | | |
| News | 360 | 0 | 205.80 | 39.01 | -0.57 | 3 | 187.38 | 38.37 | 0.18 | 7 | 227.07 | 39.13 | -2.2 | 1 |
| | | 1 | 365.81 | 40.20 | 1.6 | 4 | 360.65 | 39.85 | | | 351.96 | 40.13 | | |
| | | Avg | | 39.60 | | | | 39.11 | | | | 39.63 | | |
| | 648 | 0 | 376.64 | 42.26 | 0.17 | 4 | 321.78 | 41.26 | -1.0 | 7 | 430.98 | 42.38 | -1.8 | 1 |
| | | 1 | 621.61 | 43.29 | -4.1 | 15 | 641.49 | 42.81 | | | 636.02 | 43.27 | | |
| | | Avg | | 42.75 | | | | 42.04 | | | | 42.83 | | |
| Crew | 832 | 0 | 600.29 | 36.70 | 1.7 | 4 | 561.11 | 36.25 | -0.9 | 7 | 618.13 | 36.63 | -0.3 | 1 |
| | | 1 | 788.13 | 37.36 | -5.2 | 15 | 824.89 | 37.41 | | | 829.12 | 37.56 | | |
| | | Avg | | 37.03 | | | | 36.83 | | | | 37.09 | | |
| | 1024 | 0 | 689.10 | 37.31 | -0.1 | 3 | 655.99 | 36.86 | -1.0 | 7 | 723.36 | 37.24 | -0.1 | 1 |
| | | 1 | 1070.80 | 38.60 | 4.5 | 15 | 1013.59 | 38.19 | | | 1023.29 | 38.30 | | |
| | | Avg | | 37.95 | | | | 37.52 | | | | 37.77 | | |
| **Average** | | | | **37.15** | **3.43** | **16.8** | | **36.87** | **1.75** | **7** | | **37.20** | **1.70** | **1** |

algorithm where the model parameters are constant during encoding of the video sequences, the prediction mechanism employed in the proposed bit allocation algorithm is used to adjust the model parameters to reflect the changes that may occur from GOP to GOP in a video sequence and hence to properly represent the temporal dependency among the temporal layers. Consequently the coding efficiency of the proposed algorithm is improved compared to the two algorithms.

On the other hand, the performance of the proposed algorithm is comparable to that of JSVM FixedQPEncoder algorithm. This is because only the total bit budget $R^T$ is given to the proposed algorithm which distributes it among the quality layers whereas in the FixedQPEncoder algorithm $R^{avg}$ is allocated to QL-0 and $R^T$ is assigned to the QL-1. That means $R^{avg}$ bits more are assigned to encode a video sequence with FixedQPEncoder than those given to the proposed algorithm. However, when $R^T - R^{avg}$ is allocated to QL-1 of FixedQPEncoder (all three algorithms are allocated the same bit budget), in this case a drop in PSNR can be obtained at QL-1 of FixedQPEncoder and hence the proposed

TABLE IV
PERFORMANCE COMPARISON OF THE PROPOSED ALGORITHM, JSVM, AND CHO FIXED-RATE RC ALGORITHMS IN **SCENARIO I FOR 4CIF SEQUENCES**.

| Seq. | $R^T$ (kb/s) | QL | JSVM 9.19.14 [3] | | | | Cho fixed-Rate | | | | Proposed | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Rate (kb/s) | PSNR (dB) | E (%) | Iter | Rate (kb/s) | PSNR (dB) | E (%) | Iter | Rate (kb/s) | PSNR (dB) | E (%) | Iter |
| City | 1600 | 0 | 878.72 | 34.15 | -1.8 | 2 | 860.99 | 33.29 | -6.6 | 7 | 927.96 | 34.18 | -0.9 | 1 |
| | | 1 | 1660.51 | 35.25 | 3.8 | 14 | 1494.50 | 34.51 | | | 1585.29 | 35.16 | | |
| | | Avg | | 34.70 | | | | 33.9 | | | | 34.67 | | |
| | 2400 | 0 | 1210.58 | 35.09 | 2.5 | 15 | 1163.91 | 34.11 | -6.0 | 7 | 1196.23 | 35.00 | -1.2 | 1 |
| | | 1 | 2301.43 | 36.30 | -4.1 | 15 | 2255.81 | 35.80 | | | 2370.08 | 36.26 | | |
| | | Avg | | 35.69 | | | | 34.96 | | | | 35.63 | | |
| Soccer | 2400 | 0 | 1544.66 | 35.95 | 1.4 | 4 | 1515.16 | 35.82 | -3.6 | 7 | 1530.82 | 35.89 | -1.6 | 1 |
| | | 1 | 2406.78 | 37.20 | 0.2 | 3 | 2312.54 | 37.05 | | | 2361.78 | 37.12 | | |
| | | Avg | | 36.57 | | | | 36.44 | | | | 36.51 | | |
| | 3200 | 0 | 2024.96 | 37.12 | 1.2 | 4 | 1980.34 | 37.06 | -1.8 | 7 | 2016.79 | 37.06 | -0.58 | 1 |
| | | 1 | 3398.22 | 38.55 | 6.1 | 15 | 3140.46 | 38.31 | | | 3181.35 | 38.32 | | |
| | | Avg | | 37.83 | | | | 37.69 | | | | 37.69 | | |
| **Average** | | | | **36.19** | **3.55** | **18** | | **35.74** | **4.5** | **7** | | **36.13** | **1.07** | **1** |

TABLE V
CODING EFFICIENCY COMPARISON OF THE PROPOSED ALGORITHM AND CHO FIXED-RATE RC ALGORITHMS IN **SCENARIO II FOR CIF SEQUENCES**.

| Seq. | $R^T$ (kb/s) | QL | Cho fixed-Rate-II | | | Proposed | | |
|---|---|---|---|---|---|---|---|---|
| | | | Rate (kb/s) | PSNR (dB) | E (%) | Rate (kb/s) | PSNR (dB) | E (%) |
| Football | 648 | 0 | 366.88 | 31.97 | -6.6 | 396.43 | 32.32 | -0.46 |
| | | 1 | 604.65 | 32.48 | | 645.00 | 32.80 | |
| | 832 | 0 | 461.63 | 33.31 | -7.1 | 480.29 | 33.57 | -1.06 |
| | | 1 | 772.66 | 33.85 | | 823.17 | 34.19 | |
| News | 360 | 0 | 198.67 | 38.84 | -4.7 | 179.43 | 39.18 | -1.7 |
| | | 1 | 343.15 | 39.87 | | 354.00 | 40.59 | |
| | 648 | 0 | 396.42 | 42.12 | 3.0 | 355.77 | 41.86 | -2.4 |
| | | 1 | 667.63 | 43.09 | | 632.13 | 42.85 | |
| Bus | 648 | 0 | 394.49 | 31.20 | -7.3 | 402.73 | 31.40 | -2.2 |
| | | 1 | 600.84 | 31.68 | | 633.98 | 32.01 | |
| | 832 | 0 | 473.54 | 32.28 | -6.0 | 492.28 | 32.47 | -1.8 |
| | | 1 | 782.01 | 33.15 | | 817.49 | 33.32 | |
| Average | | 0 | | 34.95 | **5.8** | | 35.13 | **1.6** |
| | | 1 | | **35.68** | | | **35.96** | |

TABLE VI
CODING EFFICIENCY COMPARISON OF THE PROPOSED ALGORITHM AND CHO FIXED-RATE RC ALGORITHMS IN **SCENARIO II FOR 4CIF SEQUENCES**.

| Seq. | $R^T$ (kb/s) | QL | Cho fixed-Rate-II | | | Proposed | | |
|---|---|---|---|---|---|---|---|---|
| | | | Rate (kb/s) | PSNR (dB) | E (%) | Rate (kb/s) | PSNR (dB) | E (%) |
| City | 1600 | 0 | 746.83 | 31.31 | -7.7 | 708.16 | 31.47 | -6.8 |
| | | 1 | 1477.12 | 32.44 | | 1491.56 | 32.66 | |
| | 2400 | 0 | 1074.16 | 32.49 | -14.1 | 1134.90 | 33.35 | -8.3 |
| | | 1 | 2061.38 | 33.88 | | 2201.20 | 34.55 | |
| Soccer | 2400 | 0 | 1410.67 | 34.30 | -13.2 | 1284.18 | 34.28 | -3.08 |
| | | 1 | 2081.34 | 34.78 | | 2326.83 | 35.27 | |
| | 3200 | 0 | 1743.45 | 35.41 | -11.5 | 1995.70 | 36.07 | -1.3 |
| | | 1 | 2831.28 | 35.99 | | 3157.76 | 36.50 | |
| Average | | 0 | | 33.38 | **11.63** | | 33.79 | **4.87** |
| | | 1 | | **34.27** | | | **34.75** | |

TABLE VII
AVERAGED Y-PSNR OF THE QL-0 AND QL-1 OF THE PROPOSED ALGORITHM AND JSVM ALGORITHM WHEN $R^T - R^{avg}$ IS ALLOCATED TO QL-1 IN SCENARIO I.

| Seq. | $R^T$ (kb/s) | Method | Average PSNR (dB) | PSNR Gain Over JSVM |
|---|---|---|---|---|
| Foreman | 128 | Proposed | 36.77 | 0.88 |
| | | JSVM [3] | 35.89 | |
| Carphone | 64 | Proposed | 37.98 | 0.76 |
| | | JSVM [3] | 37.22 | |
| Crew | 832 | Proposed | 37.09 | 0.99 |
| | | JSVM [3] | 36.10 | |
| City | 2400 | Proposed | 35.63 | 0.56 |
| | | JSVM [3] | 35.07 | |

algorithm achieves better averaged Y-PSNR performance than that of FixedQPEncoder as shown in Table VII.

For further illustration, the averaged Y-PSNR value of QL-0 and QL-1 versus the frame number is presented in Fig. 6 to illustrate the comparison between the proposed algorithm and the two algorithms. The proposed algorithm shows better frame quality than both Cho and FixedQPEncoder when $R^T - R^{avg}$ is allocated to QL-1 and it shows comparable quality to FixedQPEncoder when $R^T$ is assigned to QL-1. It can also be seen that the FixedQp tool mostly achieves a consistent video quality throughout the frames of all video sequences among the two algorithms. This is due to the fact that FixedQpEncoder tool uses a constant quantization parameter value to encode the frames within a GOP in each temporal layer.

For Scenario II, the algorithm in [23] was also modified not only to employ the fixed-Rate strategy but also to support the quality layers with different number of temporal layers as in (17) and is named Cho fixed-Rate-II. In this case, the performance of the proposed algorithm was compared with only Cho fixed-Rate-II, since FixedQPEncoder tool supports only encoding the quality layers that have the same temporal resolution (i.e., the number of temporal layers are equal for all QLs). The results are summarized in Tables V and VI, where $s$ in eq. (17) was set to one. The proposed algorithm achieves an averaged PSNR gain of about 0.28-0.48 dB at QL-1 over the Cho fixed-Rate's algorithm. Tables III and V also compare Scenario I and Scenario II for CIF sequences which have equal frame rates at QL-1. These results indicate that coding

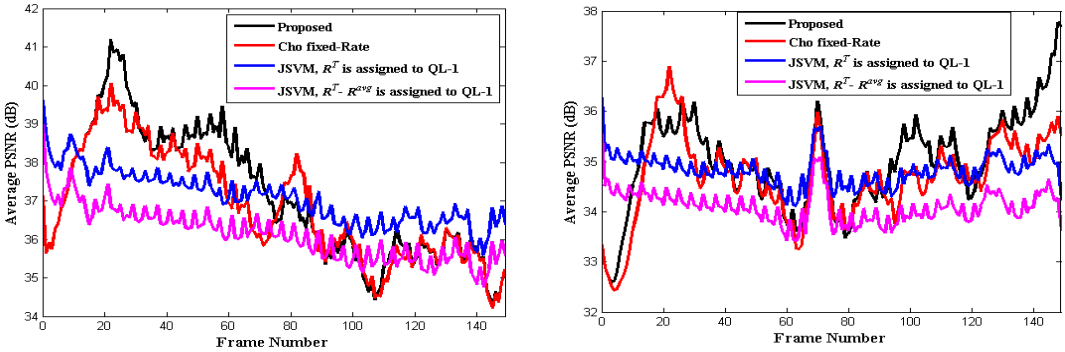



Fig. 6. Frame-by-frame averaged Y-PSNR value of the QL-0 and QL-1 of the three algorithms in Scenario I. (a) "*Crew*", CIF, $R^T$ = 832 kb/s. (b) "*Coastguard*", QCIF, $R^T$ = 128 kb/s.

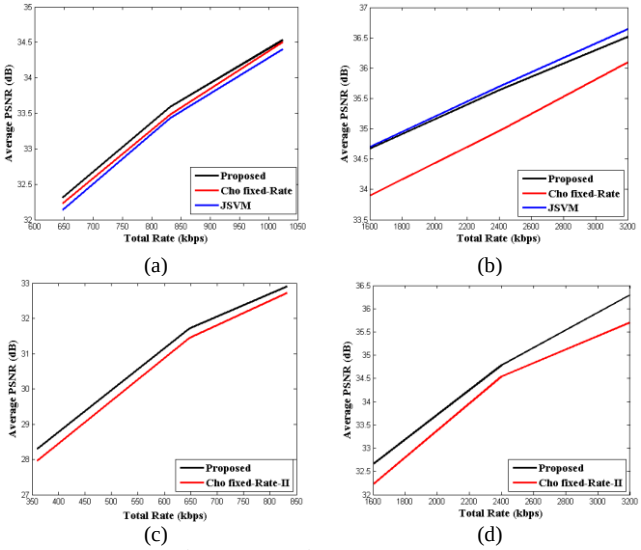



Fig. 7. Illustration of the R-D performance comparison among the three algorithms in Scenario I and Scenario II. (a) "*Football*". (b) "*City*". (c) "*Bus*". (d) "*Soccer*".

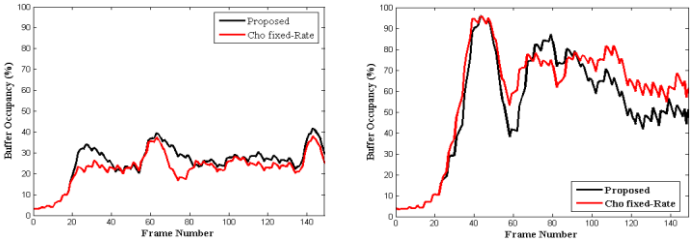



Fig. 8. Demonstration of buffer occupancy as a function of the frame number for the proposed and Cho fixed-Rate algorithms in Scenario I. (a) "*Crew*", CIF, $R^T$ = 832 kb/s. (b) "*Foreman*", QCIF, $R^T$ = 216 kb/s.

efficiency can be improved by using different temporal resolutions at each quality layer. Due to the space limitation, only four results of average PSNR performance versus the bit rates in full T-Q resolution for both scenarios are provided in Fig. 7. It can be seen from both scenarios that the proposed algorithm outperforms the Cho fixed-Rate's algorithm. Moreover, it achieves comparable quality to FixedQPEncoder since $R^T$ (higher bit rate) is assigned to QL-1 as discussed above.

### *C. Accuracy of BR Achievement and Buffer Regulation*

The accuracy of bit rate achievement in full T-Q resolution is evaluated in terms of mismatch error $E$ between the target bit rate $R^T$ and the obtained bit rate $R^O$ , which is given by:

$$E = \frac{(R^O - R^T)}{R^T} \times 100\ \%. \tag{27}$$

Tables II-IV also demonstrate the mismatch error $E$ (%) for the compared algorithms conducted on various test sequences for Scenario I. Since FixedQpEncoder tool is a multi-Rate strategy, a target bit rate is allocated to each quality layer such that $R^{avg}$ is allocated to QL-0 of the FixedQPEncoder and the total bit budget $R^T$ is assigned to QL-1. The mismatch $E$ at each quality layer is then calculated. As shown in these tables, the gap between the obtained bit rate at the full T-Q resolution by the proposed algorithm and the target bit rate is small. The proposed method achieves more accurate bit rate match compared to the Cho fixed-Rate and FixedQPEncoder algorithms. It achieves the overall bit rate absolute mismatch error within the range of 1.1% to 2.2% on average whereas the bit rate mismatches achieved with the Cho fixed-Rate and FixedQPEncoder algorithms are within the range from 1.7% to 4.5% and from 2.8% to 3.5% on average, respectively. Moreover, for the FixedQPEncoder tool, the bit rate accuracy depends on parameters such as the maximum number of iterations and the maximum mismatch. Usually, a configuration of less mismatch error may result in more number of iterations and thus more encoding computational time. Furthermore, for Scenario II the overall bit rate mismatch errors of the proposed and Cho fixed-Rate algorithms are quite larger than those for Scenario I that it is more than 5.5% on average for the Cho fixed-Rate algorithm and less than 5% on average for the proposed algorithm as shown in Tables V-VI. That is due to the fact that, usually the larger is the size of the basic unit, the better video quality can the rate control algorithm achieve, but at the cost of degradation of bit rate accuracy. In other words, for Scenario II the bit rate mismatch errors are quite larger than those for Scenario I since the GOP size at QL-1 is larger than that for Scenario I and hence the size of the basic units becomes

larger. However for both scenarios, it can be seen that the bit rate is precisely controlled using the proposed algorithm.

The performance of the proposed algorithm on buffer status management was also investigated and compared to Cho fixed-Rate algorithm. The buffer size was set to 0.5×bit rate (i.e., the maximal buffer delay is restricted to 500 ms) to satisfy the low delay requirement. Fig. 8 compares the results of buffer occupancy by Cho and the proposed algorithm. As seen the proposed algorithm is able to maintain the buffer status in a stable level and is slightly better than Cho. The buffer occupancy is around 50% when each frame has been encoded. Moreover, it is obvious that the proposed rate control algorithm could efficiently control the buffer status to prevent it from overflow and underflow. On the other hand, FixedQPencoder finds the optimum value of QP after performing a number of iterations with lack of buffer management.

### *D. Complexity considerations*

The computational complexity in terms of the number of iterations required to encode a video sequence is provided in Table VIII. Tables II-IV also show the actual number of iterations of the three algorithms conducted on various test sequences for Scenario I. It should be mentioned that the complexities of encoders are variable depending on the coding conditions such as sequence type, bit rate and inter layer coding relationships. Since the proposed and the Cho fixed-Rate bit allocation algorithms were implemented in the SVC reference software JSVM and each has the same encoding conditions, the computational cost of each encoding iteration is approximately constant. Each encoding iteration not only includes the processing costs such as motion estimation, motion compensation and macro-blocks types decision, but also the determination of the quantization parameters for each coding unit as described in Section III. Although all these are video content dependent and the actual value of the cost can vary, but its overall cost in our method is carried out in only one encoding iteration. While this for the Cho fixed-Rate algorithm requires not only six encoding passes to calculate the rate and distortion model parameters but also one additional encoding pass is demanded to encode the whole video sequence. Thus, the total number of passes is equal to seven for Scenario I. Therefore, although video content can vary, seven passes incurs more costs than one pass, no matter the complexity of video.

For the reference JSVM FixedQPEncoder tool, it iterates the coding process until the iteration stopping criteria mentioned in Section IV-B is reached (i.e., the number of iterations at each quality layer $N_{ite}[k]$ is variable). That means in the worst case 15 iterations, which is the maximum number of iterations $N_{\max}$, are required to encode a video sequence at each quality layer. Thus the total number of encoding passes in this case is equal to 30 which implies a higher computational complexity than the proposed algorithm. It is observed from Tables II-IV that the average number of iterations with the proposed, Cho fixed-Rate and FixedQPEncoder algorithms are 1, 7 and 15 iterations respectively. Since FixedQPEncoder algorithm does not consider the interlayer dependency among the layers, the computational cost of each encoding iteration is different than the proposed and Cho fixed-Rate algorithms. To investigate the complexity of the proposed algorithm, the CPU times consumed by the three encoders are shown in Table IX, where the simulations were performed on a 2.20 GHz processor with 8 GB of RAM personal computer. In Table IX, the consumed CPU time saving ratio is calculated by:

$$\left(1-\frac{\text{time (proposed algorithm)}}{\text{time (other algorithm)}}\right)\times 100\ \%, \tag{28}$$

where the time (proposed algorithm) and time (other algorithm) represent the CPU times consumed to encode the base and the enhancement layers by the proposed and other algorithms which is either Cho fixed-Rate or FixedQPEncoder, respectively. It is clear from the table that the proposed algorithm saves about 85% and 70% of the time used by Cho fixed-Rate and FixedQPEncoder algorithms to encode a video sequence, respectively. For Scenario II, this percentage of saving is increased by up to 87% of the time used by the Cho fixed-Rate. It can be seen from these results that the proposed algorithm exhibits significant improvement in the reduction of computational time compared to Cho fixed-Rate and FixedQPEncoder algorithms.

TABLE VIII
COMPUTATIONAL COMPLEXITY IN TERMS OF THE NUMBER OF ITERATIONS FOR THREE RATE CONTROL ALGORITHMS.

| | FixedQPEncoder | Cho fixed-Rate | Proposed |
|---|---|---|---|
| No. of itérations | $\sum_{k=0}^{N_Q-1} N_{ite}[k],\ N_{ite}[k] <= N_{\max}$ | $\sum_{k=0}^{N_Q-1} n_T[k]+1$ | 1 |

TABLE IX
AVERAGED CPU TIME (MS/FRAME) CONSUMED BY THE THREE RATE CONTROL ALGORITHMS FOR ENCODING THE BASE AND ENHANCEMENT LAYERS.

| $R^T$ (kb/s) | Seq. | JSVM | Cho fixed-Rate | Proposed | CPU Time Saving ratio | |
|---|---|---|---|---|---|---|
| | | | | | JSVM | Cho fixed-Rate |
| 64 | Foreman | 8303 | 12228 | 1770 | 78.68 | 85.52 |
| | Carphone | 4859 | 11390 | 1503 | 69.08 | 86.81 |
| 832 | Football | 34587 | 52763 | 7221 | 79.12 | 86.31 |
| | Crew | 36142 | 49205 | 7101 | 80.35 | 85.57 |
| 2400 | City | 141245 | 151970 | 21903 | 84.49 | 85.59 |
| | Soccer | 31828 | 163374 | 22619 | 28.93 | 86.16 |
| **Average** | | | | | **70.11** | **85.99** |

## V. CONCLUSION

In this paper, an efficient single-pass joint temporal-quality rate control algorithm was introduced to H.264/SVC. In the proposed algorithm, an overall target bit rate is adaptively distributed among the quality layers with equal and different temporal resolutions instead of predefining a target bit rate at each quality layer used in the existing RC algorithms. Moreover, in order to achieve a single pass RC algorithm, an adaptive model-parameter initialization scheme was proposed

for joint temporal-quality layers. In this scheme, a set of empirical values are first derived to estimate the initial values of the R-D model parameters. An effective prediction mechanism to update these model parameters during the encoding process is then presented to further improve RC performance. Experimental results demonstrated that with the proposed algorithm the mismatch error is reduced and the coding efficiency is improved compared to the two benchmark rate-control algorithms. Further, the proposed RC algorithm provides higher coding efficiency without requiring any pre-encoding process to estimate the model parameters Overall, while the proposed algorithm has a better performance to that of Cho fixed-Rate and FixedQPEncoder algorithms, it significantly reduces the computational time.

## REFERENCES


[1] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[2] M. Wien, R. Cazoulat, A. Graffunder, A. Hutter, and P. Amon, "Real time system for adaptive video streaming based on SVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1227–1237, Sep. 2007.

[3] "Joint Scalable Video Model JSVM 9.19.14 Software Package, CVS server for the JSVM software, June 2011.

[4] ISO-IEC/JTC1/SC29/WG11, "MPEG-2 Video Test Model 5," Draft, Apr. 1993.

[5] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 172–185, Feb. 1999.

[6] Z. Li, F. Pan, K. P. Lim, G. Feng, X. Lin, and S. Rahardja, "Adaptive Basic Unit Layer Rate Control for JVT," Doc. JVT-G012-r1, 7th JVT meeting, Pattaya, Thailand, Mar. 2003.

[7] N. Kamaci, Y. Altinbasak, and R. M. Mersereau, "Frame bit allocation for H.264/AVC video coder via Cauchy-density-based rate and distortion models," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 994–1006, Aug. 2005.

[8] Y. Liu, Z. G. Li, and Y. C. Soh, "A novel rate control scheme for low delay video communication of H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 1, pp. 68–78, Jan. 2007.

[9] L. Xu, S. Ma, D. Zhao, and W. Gao, "Rate control for scalable video model," in *Proc. SPIE*, Jul. 2005, vol. 5960, pp. 525-534.

[10] Y. Liu, Z. Li, and Y. C. Soh, "Rate control of H.264/AVC scalable extension," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 116–121, Jan. 2008.

[11] S. Hu, H. Wang, S. Kwong, T. Zhao, and C.-C.J. Kuo, "Rate control optimization for temporal-layer scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 8, pp. 1152-1162, Aug. 2011.

[12] Y. Cho, J. Liu, D.-K. Kwon, and C.-C. J. Kuo, "H.264/SVC temporal bit allocation with dependent distortion model," in *Proc. IEEE ICASSP*, Apr. 2009, pp. 641–644.

[13] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Trans. on Image Process.*, vol. 3, no. 5, pp. 533–545, Sep. 1994.

[14] J. Liu, Y. Cho, and Z. Guo, "Single pass dependent bit allocation for H.264 temporal scalability," in *Proc. IEEE ICIP*, Sep. 2012, pp. 705-7098.

[15] S. Hu, H. Wang, S. Kwong, and C.-C.J. Kuo, "Novel rate-quantization model based rate control with adaptive initialization for spatial scalable video coding," *IEEE Trans. Ind. Electron.*, vol. 59, no. 3, pp. 1673-1684, Mar. 2012.

[16] A. Leontaris and A.M. Tourapis, "Rate control for the joint scalable video model (JSVM)," Doc. JVT-W043, California, Apr. 2007.

[17] J. Liu, Y. Cho, Z. Guo, and C.-C.J. Kuo, "Bit allocation for spatial scalability coding of H.264/SVC with dependent rate-distortion analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 7, pp. 967-981, Jul. 2010.

[18] X. Jing, J. Y. Tham, Y. Wang, K. H. Goh, and W. S. Lee "Efficient Rate-Quantization Model for Frame Level Rate Control in Spatially Scalable Video Coding," in *Proc. IEEE ICON*, Dec. 2012, pp.339-343.

[19] R. Atta, R. Abdel-Kader, and A. Abd-AlRahem, "Single pass dependent bit allocation for spatial scalability coding of H.264/SVC," in *Proc. EUSIPCO*, Sep. 2014, pp. 251-255.

[20] X. Lu and G. R. Martin, "Rate control for scalable video coding with rate-distortion analysis of prediction modes," in *Proc. IEEE MMSP*, Sep. 2013, pp. 289–294.

[21] X. Li, P. Amon, A. Hutter, and A. Kaup, "One-pass multi-layer rate-distortion optimization for quality scalable video coding," in *Proc. IEEE ICASSP*, Apr. 2009, pp. 637-640.

[22] W. Bo, L. Teng, S. Sun, X. Jing, and H. Huang, "Bit allocation for quality scalability coding of H.264/SVC," in *Proc. IEEE AVSS*, Aug. 2014, pp. 165-170.

[23] Y. Cho, D.-K. Kwon, J. Liu, and C.-C.J. Kuo, "Dependent R/D modeling techniques and joint T-Q layer bit allocation for H.264/SVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 6, pp. 1003-1015, Jun. 2013.

[24] R. Atta, "Optimal bit allocation for subband video coding", *IET Image Process.*, vol. 4, no. 5, Oct. 2010.

[25] S. Ma, W. Gao, and Y. Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1533–1544, Dec. 2005.

[26] H. J. Lee, T. Chiang, and Y.-Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 6, pp.878–894, Sep. 1999.

[27] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 246–250, Feb. 1997.

[28] H. Wang and S. Kwong, "A rate-distortion optimization algorithm for rate control in H.264," in *Proc. IEEE ICASSP*, Apr. 2007, pp. 1149–1152.

[29] M. Ghanbari, *Standard Codecs: Image Compression to Advanced Video Coding*, 3rd ed., IET London, 2011.

[30] Z. Chen and K. N. Ngan, "Recent advances in rate control for video coding," *Signal Process. Image Commun.*, vol. 22, no. 1, pp. 19-38, Jan. 2007.

[31] H. Wang, G. M. Schuster, and A. K. Katsaggelos, "Rate-distortion optimal bit allocation for object-based video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 9, pp. 1113–1123, Sep. 2005.

[32] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1st ed., Cambridge University Press, UK, 2004.

[33] http://trace.eas.asu.edu/yuv/

[34] http://www-itec.uni-klu.ac.at/ftp/datasets/svc/


**Randa Atta** received the B.Sc. and M.Sc. degrees in Electrical Engineering from Suez Canal University, Port Said, Egypt, in 1991 and 1996, respectively. She received the Ph.D. degree in Electronic Systems Engineering from the University of Essex, England in 2004. Currently, she is an Associate Professor at Port Said University. She has authored or co-authored of two books. Her research interests are image/video processing/coding and pattern recognition.

**Mohammad Ghanbari** (M'78–SM'97–F'01, LF'14) is an Emeritus Professor at the School of Computer Science and Electronic Engineering, University of Essex, United Kingdom and a Professor at the school of Electrical and Computer Engineering, University of Tehran, Tehran, Iran. He is internationally best known for the pioneering work on layered video coding, which earned him IEEE Fellowship in 2001 and he was also promoted IEEE Life Fellow in 2014. He has registered for thirteen international patents and published more than 650 technical papers on various aspects of video networking, many of which have had fundamental influences in this field. These include: video/image compression, layered/scalable video coding, video transcoding, motion estimation, and video quality metrics. He is the author and co-author of 9 books, and his book video coding: an introduction to standard codecs, published by IET press in 1999, received the Rayleigh prize as the best book of year 2000 by IET.