Joint Temporal–Spatial Bit Allocation for Video Coding With Dependency

Shan Liu and C.-C. Jay Kuo, *Fellow, IEEE*

***Abstract*—The joint temporal–spatial bit allocation problem with consideration of dependency arising from motion compensated prediction as well as frame interpolation is investigated in this research. After the problem formulation, several heuristic methods are proposed to provide near-optimal and suboptimal solutions to this problem. First, the selective iteration algorithm (SIA) based on the monotonic property of the rate-distortion (R-D) curve is proposed to achieve the near-optimal solution. Then, the greedy iteration algorithm (GIA) is presented as a suboptimal heuristic to reach an R-D performance close to that obtained using SIA, while at a much lower complexity. Furthermore, by adaptively grouping frames based on the mean of the absolute difference differentials and applying greedy pruning to groups of frames, the suboptimal solution is significantly expedited so that it can be applied in real-time applications. Frames to be skipped in the coding process and quantization parameters (QPs) exploited in coded frames are adaptively and jointly determined to reach a proper tradeoff between temporal and spatial qualities. Experimental results show that the proposed methods can enhance the overall quality of compressed video at various bit rates in comparison with H.263+/TMN8 [1] using fixed frame rates and QPs, as well as the adaptive quantization solution proposed in [2].**

***Index Terms*—Bit-rate budget, frame skipping (FS), quantization parameters (QPs), temporal–spatial bit allocation, rate-distortion (R-D).**

I. Introduction

VIDEO compression is generally processed in both temporal and spatial domains. Temporal compression techniques such as motion compensated prediction (MCP) and frame skipping (FS) (temporal subsampling) are widely used in state-of-the-art video coding standards, e.g., MPEG-1/2/4 [3]–[5] and H.261/3/L [6]–[8], to reduce temporal redundancy among successive frames. Traditional still image compression methods based on DCT and quantization are adopted to reduce spatial redundancy within each frame. Both temporal and spatial compression methods seek a tradeoff between quality and bit rates. In video compression standards, spatial quality is mainly affected by the quantization parameter (QP), where a bigger QP results in a coarser picture with lower bit rates and a smaller QP results in a finer picture with higher bit rates. Sometimes, even with very blurred images, the bit rate is still over the budget if every frame is encoded. In this case, temporal subsampling is needed, where some frames are dropped without being encoded and transmitted. Under a fixed bit rate budget, more skipped frames result in worse temporal quality while each coded frame can have higher spatial quality. Consequently, temporal and spatial resolutions have to be adaptively adjusted according to video characteristics to reach the optimal joint quality. Moreover, temporal quality can be enhanced by using MCI to reconstruct skipped frames [9]. Due to the complicated prediction and interpolation dependency relationship, temporal–spatial quality tradeoff with adaptive FS has to be studied carefully in order to achieve the best overall visual quality for the entire sequence.

There has been previous work on temporal and spatial bit allocation. First, given a frame type structure and a frame rate, Ortega and Ramchandran [2], [10] investigated the optimal bit allocation problem with dependent quantization. In their work, QPs were adaptively selected under the bit rate budget at a constant frame rate, resulting in spatial-quality-oriented bit allocation. Video temporal activity and FS were not considered in [2], [10]. In contrast, Song *et al.* [11] proposed a frame rate control scheme that adjusts the encoding frame rate based on inherent motion activities. Then, the MB-layer rate-distortion (R-D) model such as that adopted in ITU-T H.263+/TMN8 can be used for bit allocation (i.e., QP adjustment) within a frame. Although both the frame rate and QPs were adaptively selected in [11], they were adjusted respectively rather than jointly. In order to jointly optimize the temporal and spatial quality, a more computationally intensive temporal–spatial bit allocation approach was recently proposed by Reed and Lim [12] under the assumption that all frames were intracoded. In this work, FS and QP were simultaneously adjusted for the best overall video quality. However, the I-frame restriction, which was adopted to avoid the computational difficulty in dealing with prediction dependency, imposes a severe constraint on its applicability to motion-compensated video coding schemes such as ISO/IEC MPEG and ITU-T H.26x codec families. Also, FR was used in [12] to reconstruct skipped frames. As a result, interpolation dependency could be greatly simplified while the quality of interpolated frames was sacrificed.

In this work, the dependent temporal–spatial bit allocation problem is investigated for video consisting of both intra- and intercoded frame types. Besides, MCI is exploited for skipped frame reconstruction. The problem is formulated in Section II. Then, the solution to this problem is studied through various approaches. On one hand, although dynamic programming has been widely used to solve the independent R-D optimization problem, it does not provide the optimal solution to dependent

Manuscript received October 11, 2001; revised December 1, 2002. This work has been supported in part by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, under Cooperative Agreement EEC-9529152. This paper was recommended by Associate Editor H. Watanabe.

S. Liu was with the University of Southern California, Los Angeles, CA 90089 USA. She is now with Sony R&D Laboratory, San Jose, CA 95134 USA.

C.-C. J. Kuo is with the Integrated Media Systems Center and Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2564 USA (e-mail: shanl@sipi.usc.edu).

Digital Object Identifier 10.1109/TCSVT.2004.839996

optimization problems due to its solution nature [10], [13]. On the other hand, exhaustive search, which can guarantee the optimality of the solution, is of little practical value due to the extremely high computational complexity.

Therefore, several near-optimal and suboptimal heuristics with reasonable computational costs are proposed in this paper. They are briefly summarized below. First, the selective iteration algorithm (SIA) is presented by extending the monotonic property, which was observed and exploited in the one-dimensional (1-D) R-D optimization method [2], [10], i.e., adaptive quantization in the spatial domain, to the two-dimensional (2-D) case and then applying it to the 2-D R-D optimization problem in the joint temporal–spatial domain. Using this monotonic property to prune out unqualified parameter (QP/FS) combinations, the complexity of exhaustive search is significantly reduced and a near-optimal solution can be reached. Second, the greedy iteration algorithm (GIA) is developed to further reduce the complexity of SIA with more aggressive pruning so that suboptimal results can be achieved in practice. These two algorithms and the 2-D monotonic property are detailed in Section III.

In Section IV, an adaptive frame grouping (AFG) scheme is introduced as a preprocessing tool, where successive frames with constant motion velocity and similar texture complexity are grouped and assigned with the same temporal (FS) and spatial (QP) resolutions. By applying the GIA to a group of frames (GOFs), the frame interpolation dependency can be greatly reduced, and thus the overall complexity. It is worthwhile to emphasize that both skipped and coded frames contribute to the operational R-D function in our framework. The MCI scheme based on the triangular-patch affine warping method [14], [15] is adopted to reconstruct skipped frames.

Experimental results are given in Section V to demonstrate the performance of proposed bit allocation algorithms. It is observed that the proposed algorithms can outperform the H.263+/TMN8 standard codec [1] with fixed FS/QP from 0.3 to 1 dB in average PSNR for different test sequences under various bit rate budgets. Moreover, the PSNR variance is smaller, which implies that the output visual quality is more consistent from frame to frame. This is usually preferred by the human visual system. The performance improvement of the 2-D R-D optimization over the 1-D R-D optimization as proposed in [2] is also reported. Finally, concluding remarks are given in Section VI.

II. Problem Formulation

A. Preliminaries

1) Spatial QP Adaptation: The classical bit allocation problem was examined with or without temporal dependency [2], [10]. Given N dependent coded frames with a fixed frame rate, the problem is to select QPs for all frames, $\{Q_i, i=1,\ldots,N\}$, to achieve the best overall spatial quality. Mathematically, the problem is to determine

$$\min_{Q_1,\ldots,Q_N} \sum_{i=1}^{N} D_i(Q_1,\ldots,Q_i)$$
$$\text{subject to} \quad \sum_{i=1}^{N} R_i(Q_1,\ldots,Q_i) < B \tag{1}$$

where $D_i(\cdot)$ and $R_i(\cdot)$ are the distortion and the rate of the ith frame under a given QP selection, respectively. Also, B stands for the total bit budget. Note that the frame rate is predetermined in (1) so that FS is fixed for the entire sequence. Only the spatial quality of coded frames is optimized while skipped frames are not considered if there are any.

In contrast, conventional frame-rate control algorithms perform dynamic FS according to the motion activity of underlying video while determining the QP of each frame independently. It is natural that the bit rate can be more efficiently allocated if the temporal–spatial tradeoff is properly exploited.

2) Affine MCI: No matter how a sequence is compressed and transmitted, the final received video quality can be best evaluated at the full frame rate. Thus, frame-rate up-conversion techniques are used to reconstruct skipped frames from neighboring coded frames when FS is adopted. Commonly used methods include FR, frame averaging (FA), and MCI [9]. Better frame interpolation results in improved temporal quality, and thus the overall visual quality. The six-parameter affine warping model [14], [15] is adopted in our proposed system. It is defined as

$$\begin{aligned} x_i' &= a_0 + a_1 \times x_i + a_2 \times y_i \\ y_i' &= b_0 + b_1 \times x_i + b_2 \times y_i \end{aligned} \tag{2}$$

where (x_i', y_i') is the 2-D coordinates of the interpolated pixel in the skipped frame while (x_i, y_i) is that of the corresponding pixel in the reference frame (e.g., the previous or the next coded frame). Once a skipped frame is reconstructed, its distortion (D) can be measured by calculating the PSNR value of the reconstructed picture with respect to the original one. Obviously, the rate (R) of a skipped frame is zero. Hence, the R-D performance of a skipped frame can be applied to the proposed 2-D R-D optimization problem formulated in the next subsection.

B. Dependent Temporal–Spatial Bit Allocation

The overall quality of the full frame-rate video playback is dependent on coded frames as well as skipped frames. The frame set $\mathbf{S}$ is used here to indicate coded/skipped frames

$$\begin{aligned} &\mathbf{S} = [S_1, S_2, \ldots, S_N] \\ &S_i \in [0,1], \quad i = 1,\ldots,N. \end{aligned} \tag{3}$$

In above, S_i takes a binary value (1/0) to denote a coded or a skipped frame among a total of N frames in the input sequence. Similar to that in (1), $\mathbf{Q}$ denotes the set of QP, i.e.,

$$\begin{aligned} &\mathbf{Q} = [Q_1, Q_2, \ldots, Q_N], \\ &Q_i \in [Q_{\min}, Q_{\max}], \quad i = 1,\ldots,N. \end{aligned} \tag{4}$$

Most video coding standards rely on motion-compensated prediction to reduce the temporal redundancy among successive frames, which results in prediction dependency among coded frames. For example, the P-frame is predicted from the previous reference (I or P) frame while the B-frame is predicted from both the previous and the next reference frames. These INTER (P and B) frames contribute to higher compression efficiency. On the other hand, since the skipped frame is interpolated from adjacent coded frames, its quality is dependent upon that of coded frames. Thus, interpolation dependency is also introduced. We call both prediction and interpolation dependency the temporal dependency, which is generally unavoidable in MCP-based video coders, especially for low bit rate applications.

Based on the above discussion, the total distortion is contributed by both coded and skipped frames and can be expressed by

$$\sum_{i=1}^{N} D_i(\mathbf{Q},\mathbf{S}) = \sum_{i=1}^{N} \{D_i(\mathbf{Q},\mathbf{S}) \,|\, (S_i=1) + D_i(\mathbf{Q},\mathbf{S}) \,|\, (S_i=0)\}. \quad (5)$$

Since skipped frames do not cost any bit, the total rate is contributed by coded frames only. That is

$$\sum_{i=1}^{N} R_i(\mathbf{Q},\mathbf{S}) = \sum_{i=1}^{N} R_i(\mathbf{Q},\mathbf{S}) \,|\, (S_i=1). \quad (6)$$

Consequently, the 2-D bit allocation problem is to find $\mathbf{Q}^*$ and $\mathbf{S}^*$ such that

$$[\mathbf{Q}^*,\mathbf{S}^*] = \arg\min_{\mathbf{Q},\mathbf{S}} \sum_{i=1}^{N} D_i(\mathbf{Q},\mathbf{S}) \quad \text{subject to } \sum_{i=1}^{N} R_i(\mathbf{Q},\mathbf{S}) < B \quad (7)$$

where $D_i(\mathbf{Q},\mathbf{S})$ and $R_i(\mathbf{Q},\mathbf{S})$ are the distortion and the rate of the ith frame under the given FS and QP sets, respectively. Usually, R is measured in bits and D in the mean square error (MSE).

This problem can be simplified by forcing all coded frames to be intraframes to avoid prediction dependency and using FR instead of bidirectional interpolation methods (e.g., FA and MCI) to eliminate interpolation dependency. As a result, the temporal dependency can be totally ignored as done in [12]. Following this line of thought, we can rewrite (5) and (6) as

$$\sum_{i=1}^{N} D_i(\mathbf{Q},\mathbf{S}) = \sum_{i=1}^{N} \{D_i(Q_i) \,|\, (S_i=1) + D_i\left(Q_{i_p}\right) \,|\, (S_i=0)\} \quad (8)$$

and

$$\sum_{i=1}^{N} R_i(\mathbf{Q},\mathbf{S}) = \sum_{i=1}^{N} R_i(Q_i) \,|\, (S_i=1) \quad (9)$$

where Q_i is the QP of the ith intracoded frame and Q_{i_p} is the QP of the immediate previous intraframe of the ith skipped frame.

Under such simplifying conditions, both the prediction and the interpolation dependency relations are removed, and dynamic programing can be directly applied for the optimal solution. However, these simplifying conditions do not apply to MCP-based video coders that are widely in use today. Hence, we do not simplify the problem formulation, but propose some suboptimal heuristics to the original problem given by (5), (6), and (7) in this work.

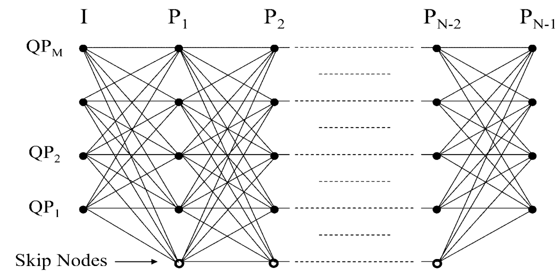


Fig. 1. Trellis for the temporal–spatial bit allocation problem.

C. Complexity

The basic difference between the 1-D (spatial only) and the 2-D (joint temporal–spatial) bit allocation problems is that the frame rate (or FS) is predetermined in the 1-D case, while adaptive in the 2-D case. Therefore, for the 1-D problem, only QP is adaptive and only coded frames are of concern when evaluating the R-D performance. In contrast, FS and QP are both varying in the 2-D case, and the R-D performance evaluation should involve both coded and skipped frames.

The traditional 1-D bit allocation problem can be presented by a "trellis", where each stage of the trellis indicates a coded frame and each node of that stage corresponds to the R-D performance of the frame coded at a certain QP [2], [10]. When the trellis is applied to the 2-D bit allocation problem, an extra node is introduced at each stage to indicate the skipped frame, which is called the "skip node" as shown in Fig. 1. Unlike the coded frame (node), whose rate and distortion values are calculated at the current stage, a skipped frame will be reconstructed by affine MCI [see (2)] after the next coded frame is available in our scheme. Therefore, the R-D performance of a skipped frame cannot be computed at the current stage but at later stages. It is thus clear that skip nodes have to be examined and pruned in later stages, while coded nodes can be pruned in the current stage. Hence, the complexity of the 2-D bit allocation problem is much higher than that of the 1-D problem.

As shown in Fig. 1, we assume that there are total N frames in the sequence and the total number of possible QP levels is M. Given that any frame except the first and the last can be skipped, there are totally $\sum_{i=0}^{N-2} M^{N-i} C_{N-2}^{i}$ possible paths, from which the best one should be selected. Hence, the complexity of exhaustive search is $O(M^N \sum_{i=0}^{N-2} C_{N-2}^{i}/M^i)$. In contrast, the complexity of the 1-D exhaustive search method, i.e., without considering FS, is $O(M^N)$ [2].

III. Solutions to the Problem

Dynamic programming was proposed to solve the independent 2-D R-D optimization problem [12]. However, this technique is not applicable to the dependent case. Generally speaking, exhaustive search is required to guarantee the optimal solution by examining all possible paths. However, the complexity of exhaustive search is too high to be practical. Some suboptimal heuristics are proposed in this section to reduce the complexity. Our goal is to obtain near-optimal or suboptimal solutions at reasonable computational costs. In this paper, we focus our discussions on the IPP... structure. That is, the first

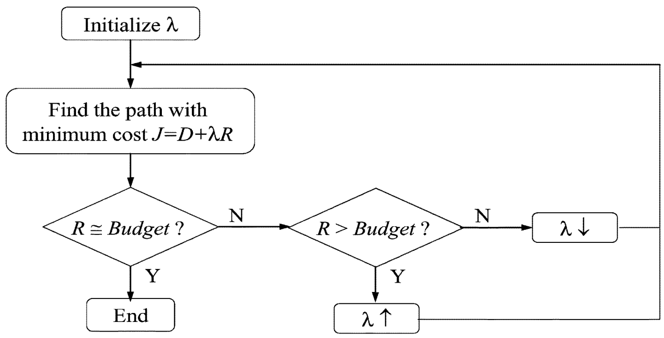


Fig. 2. R-D optimization procedure via iterations on λ.

frame is coded as an intraframe while all following frames are P-frames. However, the solution methodology given below can be easily extended to other video coding structures that may involve multiple I-frames and/or additional B-frames.

A. SIA

The high complexity of exhaustive search is due to the exponentially increasing number of paths from one stage to another. Thus, the key issue here is to develop a pruning mechanism to eliminate unlikely paths from the "trellis" at each stage so that the number of candidate paths can be reduced. Based on this idea, the SIA is proposed.

1) Cost Function: First, a cost function J is defined to measure the R-D performance of a selected path with the Lagrangian parameter λ

$$\begin{aligned} D(i,k) &= \sum_{j=1}^{k} D(i,j), \\ R(i,k) &= \sum_{j=1}^{k} R(i,j), \\ J(i,k) &= D(i,k) + \lambda R(i,k) \end{aligned} \tag{10}$$

where $D(i,k)$ and $R(i,k)$ are the total distortion and the corresponding rate for the ith path up to the kth frame (or the kth stage), respectively.

Any video encoder, especially MCP-based, can be adopted in the proposed optimization mechanism to provide $R(i,j)$ and $D(i,j)$ for coded frames. The baseline H.263+/TMN8 [1] is used as an example in this work. As stated in Section II-A2, the affine MCI scheme is adopted to calculate the distortion $D(i,j)$ for skipped frames while the rates of skipped frames are equal to zero. The optimization procedure is conducted through iterations. During the iteration, the cost parameter λ starts from the high and the low boundary values and converges gradually to a proper value in the middle so that the total bit rate converges to the budget. It is briefly shown in Fig. 2. The trellis based pruning methods, i.e., selective pruning and greedy pruning, are applied to find out the path with minimum total cost, which will be discussed later in this section and Section III-B, respectively.

2) FS and QP Constraint: Skip nodes are introduced to represent skipped frames as shown in Fig. 1. Theoretically, the maximum FS can be $N-2$ in a sequence with N total frames, i.e., any frame except the first and the last frame can be skipped. However, in general, temporal quality would become unacceptable when FS exceeds a certain threshold. Hence, a maximum value of FS, $S_{\max}$ is set to ensure proper temporal quality. FS is checked at every stage. That is, any path that has more than $S_{\max}$ successive skipped frames is pruned out.

Furthermore, the flickering effect is an annoying visual artifact that occurs when the spatial quality of adjacent frames varies significantly. Although most video coding standards such as MPEG-4 and H.263 support a wide QP range, (e.g., from 2 to 31 in a total of 30 levels for H.263) it is observed that, when the QP difference between adjacent frames is above a certain value ΔQ, the flickering effect becomes visible and annoying to human beings. Therefore, by limiting the selective QP range, the flickering effect can be alleviated while the complexity can be reduced at the same time. In terms of mathematics, let the QP of the kth coded frame be $Q(k)$, then the QP range for the coded frame at the $(k+1)$th stage can be limited as

$$Q(k+1) \in [Q(k) - \Delta Q, Q(k) + \Delta Q] \bigcap [Q_{\min}, Q_{\max}] \tag{11}$$

where ΔQ is the maximum of QP difference between adjacent coded frames, while $Q_{\min}$ and $Q_{\max}$ are the minimum and the maximum QP supported by video coding methods, respectively.

3) Monotonic Property: Ramchandran *et al.* [2] pointed out an interesting observation on the R-D curve of the 1-D dependent bit allocation problem, which is called the monotonic property. This property can be stated mathematically as follows. For any $\lambda \geq 0$, it is observed that

$$J(i,j) \leq J(i',j), \quad \text{if } i \leq i' \tag{12}$$

where i and j represent the QP of the ith and the jth coded frames from the lowest (finest) to the highest (coarsest) level, respectively. In words, the 1-D monotonic property means that a "better" (finer quantized) predictor leads to more efficient coding. This property appears to be valid for the 2-D case with FS as well. That is, we observed that a better reference frame results in not only better predicted frames but also more accurate reconstruction of skipped frames. Thus, for any $\lambda \geq 0$, we have

$$\begin{aligned} &J(i,s_{ij},j) \leq J(i',s_{i'j},j), \quad \text{if } i \leq i' \\ &J(i,s_{ij},j) \leq J(i,s_{ij'},j'), \quad \text{if } j \leq j' \\ &J(i,s_{ij},j) \leq J(i',s_{i'j'},j'), \quad \text{if } i \leq i' \;\; j \leq j' \end{aligned} \tag{13}$$

where i and j stand, respectively, for the QP of the ith and the jth coded frame from the lowest (finest) to the highest (coarsest) level and s_{ij} denotes the skipped frame reconstructed from frames i and j. Since this property is not proved but observed from numerous experiments, the solution achieved by exploiting (12), (13) for pruning should not be claimed as strictly optimal. Instead, we say that it is near-optimal at a substantially lower computational complexity compared with exhaustive search.

Pruning conditions for both coded and skipped frames can be derived based on the monotonic property. The pruning rules for coded frames given in [2] are stated as follows.

Rule 1: If

$$J(i) + J(i,j) \le J(i') + J(i',j) \quad \text{for any } i \le i' \tag{14}$$

then branch (i', j) cannot be a part of the optimal path and should be pruned.

Rule 2: If

$$J(i,j) \le J(i,j') \quad \text{for any } j \le j' \tag{15}$$

then branch (i, j') cannot be a part of the optimal path and should be pruned.

Furthermore, we develop additional pruning rules for the 2-D dependent bit allocation problem by considering skipped frames as follows.

Rule 3: If

$$\begin{aligned} J(i) + J(i, s_{ij}, j) + J(i,j) \le J(i') + J(i', s_{i'j}, j) \\ + J(i',j) \quad \text{for any } i \le i', \end{aligned} \tag{16}$$

then branch $(i', s_{i'j}, j)$ cannot be a part of the optimal path and should be pruned.

Rule 4: If

$$J(i, s_{ij}, j) + J(i,j) \le J(i, s_{ij'}, j') + J(i,j') \quad \text{for any } j \le j' \tag{17}$$

then branch $(i, s_{ij'}, j')$ cannot be a part of the optimal path and should be pruned.

The property stated in Rule 3 can be proved using contradiction. That is, let us assume branch $(i', s_{i'j}, j)$ be a part of the optimal path and $(i', s_{i'j}, j, s_{i'jk}, k, \ldots, n)$ be the optimal path. Based on the monotonic property (12), we have

$$J(i,j) \le J(i',j), \quad \text{if } i \le i' \tag{18}$$
$$J(i,j,k) \le J(i',j,k) \tag{19}$$
$$\cdots$$
$$J(i,j,k,\ldots,n) \le J(i',j,k,\ldots,n) \tag{20}$$

where k is the QP level of the coded frame following the jth frame and n is the QP level of the last coded frame.

Based on the monotonic property (13), we have

$$J(i,j,s_{ijk},k) \le J(i',j,s_{i'jk},k) \tag{21}$$
$$\cdots$$
$$J(i,j,k,\ldots,s_{ijk\ldots n},n) \le J(i',j,k,\ldots,s_{i'jk\ldots n},n) \tag{22}$$

Summing up (18)–(22), we conclude that the total cost $\Sigma\mathbf{J}$ of path $(i,j,k,\ldots,n)$ is less than that of path $(i',j,k,\ldots,n)$, which contradicts to the assumption that $(i',j,k,\ldots,n)$ is the optimal path. Therefore, branch $(i', s_{i'j}, j)$ cannot be a part of the optimal path, and should be pruned. Rule 4 can be proved using similar arguments. The above four pruning rules are used together to eliminate unlikely branches so that the complexity is reduced.

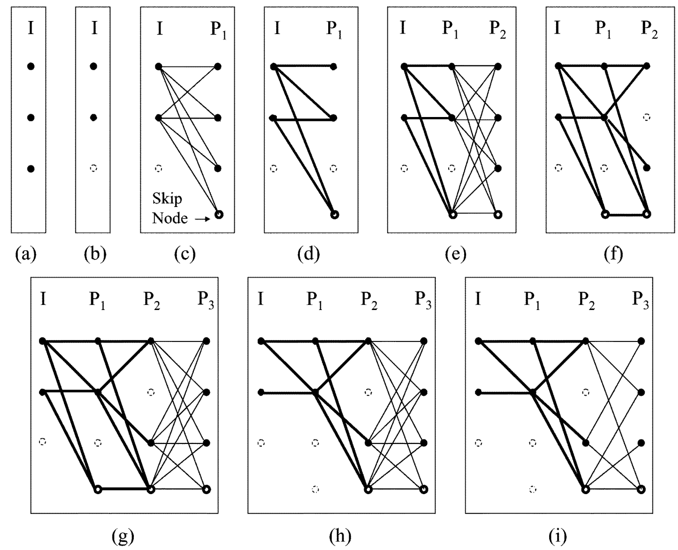


Fig. 3. Illustrative example of the SIA.

4) Description of SIA: The SIA is described below with an illustrative example given in Fig. 3.

Step 1) Initialize the value of λ.
Step 2) Calculate $J(i,1)$ for the first frame, which is an I-frame, for every QP value within the range $i \in [Q_{\min}, Q_{\max}]$, as shown in Fig. 3(a).
Step 3) Prune unqualified I-nodes according to the monotonic property as shown in Fig. 3(b).
Step 4) Grow the trellis to Stage 2 by coding the first P-frame with all QP values. The skip node is reserved as shown in Fig. 3(c).
Step 5) Prune at Stage 2 with Rules 1 and 2. Note that the skip node should be kept as shown in Fig. 3(d).
Step 6) Grow the trellis to one more stage. The skipped frame in the previous stage is reconstructed by the neighboring reference frames coded with selected QPs as shown in Fig. 3(e).
Step 7) Prune at Stage 3 based on the monotonic property, i.e., Rules 1 and 2 for pruning the third coded frames, Rules 3 and 4 for pruning previous skipped frames. The skip node at Stage 3 is reserved as shown in Fig. 3(f).
Step 8) Similar to Step 4, grow trellis to Stage 4 as shown in Fig. 3(g).
Step 9) Prune paths that have more successive skipped frames than $S_{\max}$ as shown in Fig. 3(h). We set $S_{\max} = 2$ in this example.
Step 10) Similarly to Step 7, pruning is performed based on the monotonic property as shown in Fig. 3(i).
Step 11) Repeat Step 8–10 until the last frame. Update λ and return to Step 2.
Step 12) Stop when λ converges.

The complexity of selective iteration is much lower than that of exhaustive search. Thus, it is a feasible solution and can be

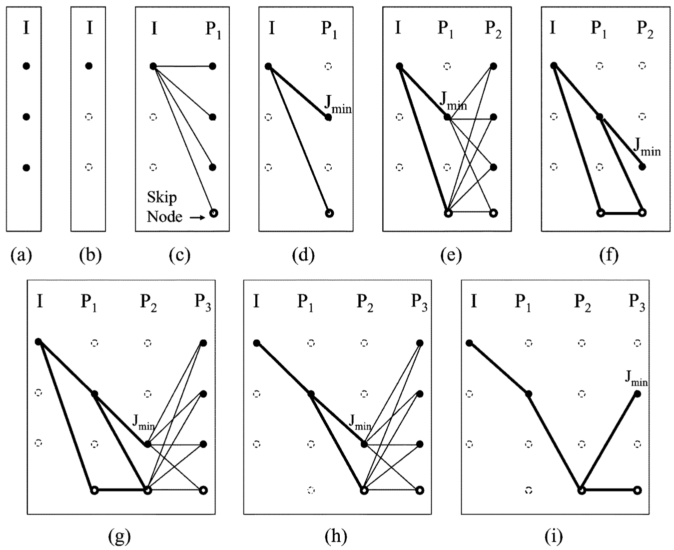


Fig. 4. Illustrative example of the GIA.

used to provide a near-optimal benchmark. However, due to the fact that many skip nodes within the range of $S_{\max}$ need to be reserved, the complexity of the 2-D dependent problem is still significantly higher than that of the 1-D counterpart. Hence, we propose a GIA in the following section to further reduce the complexity by more aggressive pruning

B. GIA

In this approach, instead of pruning out nodes that violate the monotonic property, only nodes that give the best R-D performance and pending nodes are kept while all others are pruned out. The best nodes correspond to those frames coded with certain QP values, through which the path with the minimum cost J passes. The pending nodes are skip nodes through which more than one path may be reserved. The GIA shares the same cost function and FS/QP constraints with SIA, and the number of successively skipped frames has to be less than $S_{\max}$ as well. The operation procedure is similar to SIA but with more aggressive pruning. An example is shown in Fig. 4.

Description of GIA

Step 1) Initialize λ.

Step 2) Calculate $J(i,1)$ for the first frame, which is an I-frame, with each QP within $i \in [Q_{\min}, Q_{\max}]$ as shown in Fig. 4(a).

Step 3) Select the I-frame with the lowest cost J as shown in Fig. 4(b).

Step 4) Grow the trellis to Stage 2 by coding the first P-frame with all QP values. The skip node is reserved as shown in Fig. 4(c).

Step 5) Keep the node with the lowest cost J, and prune out all others except the skip node as shown in Fig. 4(d).

Step 6) Grow the trellis to one more stage. The skipped frame in the previous stage is reconstructed by the neighboring reference frames coded with selected QPs as shown in Fig. 4(e).

Step 7) Prune out all nodes except the best node and the skip node as shown in Fig. 4(f).

Step 8) Grow trellis to Stage 4 as shown in Fig. 4(g).

Step 9) Prune out paths that have more successively skipped frames than $S_{\max}$ ($S_{\max} = 2$ in this example) as shown in Fig. 4(h).

Step 10) Prune nodes in the same way as described in Step 7 as shown in Fig. 4(i).

Step 11) Repeat Step 8–10 until the last frame. Update λ and return to Step 2.

Step 12) Stop when λ converges.

Since the quality of an I-frame will normally affect the quality of the following P-frames and thus the overall quality of the whole sequence, we slightly modify the above algorithm to keep all possible QPs of the first I-frame, and apply the monotonic pruning rule to the first I-P group only. Experiments show that this modification may slightly enhance the optimization speed by facilitating the convergence of λ. The complexity of GIA is tremendously reduced compared with either exhaustive search or SIA, while the solution is still close to the optimal. Thus, it is a more practical approach. Let the total number of frames be N, the total number of candidate QP levels be M and the maximum FS be S, the number of reserved paths at each stage is fixed $(S+1)$. Then, the complexity in one iteration of GIA is $O(SMN)$.

It will be shown in Section V that, compared with H.263+/TMN8 codec with fixed QP and FS, around 0.3–1.0 dB improvement in PSNR is achieved by applying GIA for joint temporal–spatial bit allocation, with affine MCI for skipped frame reconstruction. In fact, any frame interpolation method can be adopted in association with the proposed method to result in selected QP/FS combinations and various degrees of quality enhancement. It is observed that bidirectional frame interpolation methods (such as MCI and FA) generally provides better reconstructed frames than unidirectional methods (such as FR). Thus, more frames can be skipped when MCI is adopted in comparison with FR.

IV. Fast Approach With Adaptive Frame Grouping

FS significantly affects the complexity of the 2-D dependent bit allocation problem. As discussed before, the complexity of 2-D dependent GIA is $O(SMN)$ while that of 1-D dependent GIA (with fixed FS) is $O(MN)$, where N is the total number of frames, M is the total QP levels and S is the maximum FS. To reduce the complexity caused by FS while keeping a certain degree of temporal flexibility, the concept of GOF is introduced below.

A. Adaptive Frame Grouping

Definition: A GOF is a group of frames, including both coded and skipped frames, which have similar temporal–spatial characteristics so that a uniform FS/QP can be applied to them without sacrificing much optimality.

Based on the above definition, we can calculate the cost function at the GOF level via

$$D(i,j) = \sum_{n=1}^{N} D(i,n)$$
$$R(i,j) = \sum_{n=1}^{N} R(i,n) \qquad (23)$$

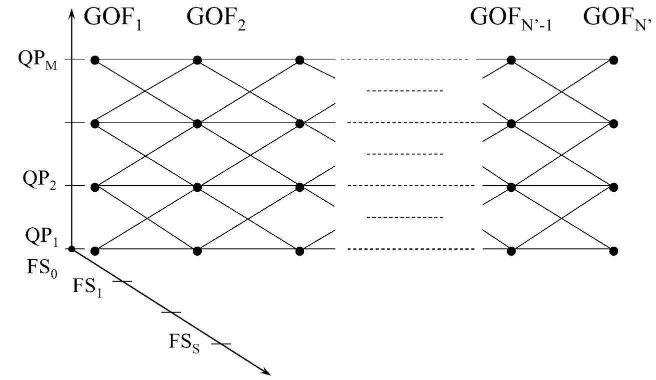


Fig. 5. Trellis for fast temporal–spatial bit allocation with a certain FS.

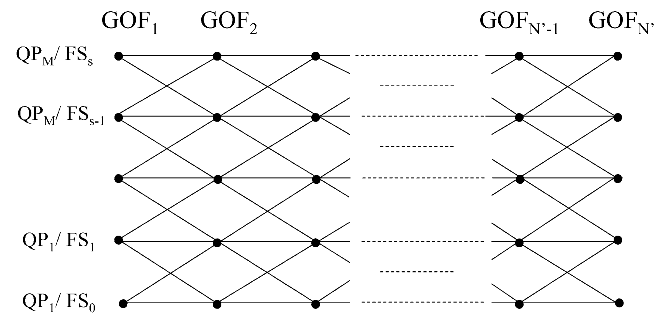


Fig. 6. Simplified trellis for fast temporal–spatial bit allocation.

where $D(i,j)$ and $R(i,j)$ are the distortion and the rate of the jth GOF along the ith path, respectively, and N is the total number of frames in that GOF. The sums of rates and distortion measures from all frames in one GOF are taken as the group rate and the distortion measure, respectively. Note that each GOF does not contribute to the cost equally. Instead, a GOF with more frames weighs more (reflected by larger R and D values resulted from the summation) based on the fact that each single frame is equally weighted in the overall visual quality. By applying (23) to (10), the cost function $J(i,k)$ for the kth stage (GOF) can be calculated in a similar way.

With the GOF concept, skip nodes in Figs. 1, 3, and 4 do not exist in the modified trellis. Each node in Fig. 5 represents a GOF. The temporal–spatial flexibility in this scenario is not at the frame level but at the GOF level.

Let the total number of GOF be N', the number of QP level be M and the maximum FS be S, then the complexity of one iteration is $O(SMN')$. Compared to the single frame greedy iteration, the complexity is reduced by a factor N/N'. Obviously, a constant motion velocity results in a larger number of frames in one GOF and thus less complexity.

One main issue here is how to group frames for efficient bit allocation. A frame-level rate control scheme with frame grouping was proposed by Song *et al.* [11]. In their work, a GOP is defined as a group of pictures starting with an I-frame followed by a long sequence of predicted frames until the next I frame. One GOP is split into sub-GOPs of equal length, where the length is determined a priori. FS and QP can be adjusted among sub-GOPs, while kept fixed within each sub-GOP. This implies that the motion velocity among frames in each sub-GOP should be nearly constant, which might not always be true. In contrast, GOFs defined in this paper are not equally long. Instead, their lengths are adaptively determined to ensure that the same FS and QP can be applied to frames with similar temporal–spatial characteristics.

One widely accepted criterion to measure distinct motion and texture characteristics is the mean of the absolute difference (MAD), which is defined as

$$\mathrm{MAD}(i) = \frac{\sum_{j=1}^{N} |p(i,j) - p(i-1,j)|}{N} \tag{24}$$

where $p(i,j)$ indicates the jth pixel value in the ith frame with a total number of N pixels in one frame. MAD gives the average pixel by pixel difference between two adjacent frames and reflects the motion and residue activities. A larger MAD value implies faster motion and more residues from the previous frame to the current frame so that less FS and QP should be assigned, and vice versa. Frames with similar MAD can be grouped into one GOF. The differential of MAD is used here to detect the MAD value change, i.e.,

$$\mathrm{MAD}'(i) = \frac{d(\mathrm{MAD}(i))}{di} \approx \mathrm{MAD}(i) - \mathrm{MAD}(i-1). \tag{25}$$

When $\mathrm{MAD}'(i)$ exceeds a predetermined threshold, we start a new GOF from the ith frame. Since a skipped frame in one GOF can only be reconstructed after the next coded frame is available (by assuming a general frame interpolation method is used, e.g., bidirectional MCI or FA), the last frame in a GOF must be a coded frame. A GOF can start with a skipped frame except for the first GOF that contains only one I frame. For example, if we detect $\mathrm{MAD}' >$ threshold at $1, 2, 5, 14, \ldots$, then the first GOF includes Frame 1, the second GOF contains Frames 2–4, the third contains 5–13, and so on.

B. GIA Applied to GOF

We explain how GIA is applied to GOFs with an example as shown in Fig. 6, where the 2-D trellis given in Fig. 5 is redrawn in 1-D for simplification. Let the maximum FS be S and the number of QP levels be M.

- Step 1) Initialize the value of λ.
- Step 2) Encode the first GOF (i.e., the I-frame) with all possible QPs.
- Step 3) The second GOF is coded with all QP/FS combinations from all nodes generated in Step 2, where skipped frames are reconstructed by their previous and next coded frames. Because all nodes (M) of the first GOF (I-frame) are reserved, the total number of possible paths at this stage is SM^2.
- Step 4) The R-D cost is calculated for every path up to the current stage, which is contributed by both coded and skipped frames in each GOF, and the path with the lowest cost is selected.
- Step 5) The next (i.e., the third) GOF is generated from the selected current (i.e., the second) GOF, with all QP/FS combinations. The total number of candidate paths in this stage is SM.
- Step 6) Greedy pruning is performed in a way similar to that in Step 4, and only the path with the lowest cost is kept.

Step 7) Repeat Steps 5 and 6 until the whole sequence is finished. Update λ according to the rate and the budget.

Step 8) Repeat Steps 2–7 until λ converges.

Some approximations can be used to further reduce the complexity and make the approach faster. In the standard codec with fixed FS and QP, QP can be determined by FS and the budget. Based on this fact, a specific empirical QP can be associated with a certain FS under the bit rate budget. Therefore, in the above procedure, for each FS value, encoding can be done only once with the empirical QP value so that the complexity of each iteration can be reduced to SN'. It is also observed that λ converges faster with empirical QP and FS pairs.

V. Experimental Results

Experiments were conducted to compare the performance of the proposed heuristic methods (i.e., SIA, GIA applied to single frames and GOFs), the spatial domain adaptive quantization method [2] and the standard H.263+/TMN8 codec [1] with various FS and QP combinations. In our simulations, the average PSNR was calculated based on both coded and skipped frames, i.e., the comparison was conducted at the full frame rate. Skipped frames in both proposed and reference approaches were reconstructed by affine MCI (2), and the distortion of a skipped frame was measured by the PSNR of the reconstructed image with respect to the original. The initial lower and upper boundary values of λ were set to 0.0 and 99.0, respectively, and the threshold to determine the convergence of λ was set to 0.0001. λ was updated using the slope of the R-D curve during iterations until it converged or the maximum number of iterations was met. The IPP… structure was adopted for simplicity. All test sequences were of the QCIF format, if not specified.

A. Frame-Based SIA versus GIA

Since exhaustive search is practically infeasible, we use the result obtained by SIA as the performance benchmark. Because the high complexity of SIA, only a short sequence was simulated. The first five frames (IPPPP) of the "Suzie" sequence were simulated using SIA and GIA for the purpose of comparison. We see from Fig. 7 that GIA (-∗-) can achieve an average PSNR close to that of SIA (-o-) under the same rate while the computational complexity is significantly reduced. Therefore, we can conclude that GIA provides a good heuristic solution to the temporal–spatial dependent bit allocation problem.

B. Frame-Based GIA versus Adaptive Quantization [2]

With fixed FS, the temporal adaptivity of the proposed GIA is eliminated while only QPs of coded frames can be adjusted, which is the same as the method given in [2]. Thus, we set the maximum FS to zero (i.e., $S_{\max} = 0$) in the proposed GIA to simulate the adaptive quantization (greedy) approach in [2] at the full frame rate.

In Fig. 8, we compare the overall R-D performance of the proposed 2-D GIA (frame based, -o-) with that of the spatial domain adaptive quantization [2] (-▷-) applied to Frames 1–120 of the "Suzie" sequence. It can be seen that the proposed GIA achieves a higher average PSNR (luminance) value than its 1-D counterpart [2] since GIA takes advantage of adaptive FS as well as MCI for skipped frame reconstruction when motion activities are low. Thus, more bits can be saved to enhance the picture quality of coded frames, which benefits the reconstruction of skipped frames in return.

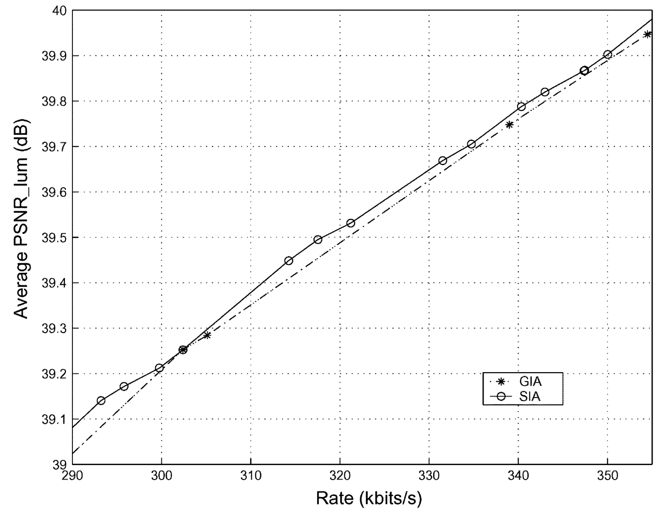


Fig. 7. R-D performance comparison of SIA (-o-) and GIA (-∗-) applied to Frames 1–5 of the "Suzie" sequence.

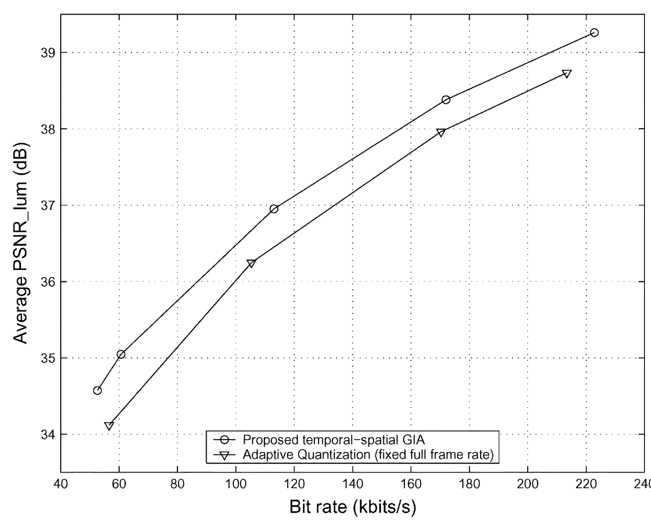


Fig. 8. R-D performance comparison of proposed GIA (-o-) and spatial domain adaptive quantization (-▷-) applied to Frames 1–120 of the "Suzie" sequence.

Fig. 9 illustrates the frame to frame quality (PSNR of luminance) comparison of the proposed GIA (frame based, solid) with the spatial domain adaptive quantization method [2] (dashed) under a bit budget of 60 kb/s. We see that, for most frames, the picture quality using the proposed GIA is better than that obtained by [2], especially when motion activities are low, e.g., before the 40th frame and after the 80th frame. The reason is the same as that stated earlier. When motion activities are high, e.g., from the 40th to the 80th frame, the picture quality obtained by both methods is similar since the full frame rate is preferred by the proposed method so that it is reduced to the 1-D adaptive quantization. The consumed bit rates by the proposed and the 1-D reference methods were 60.66 and 59.58 kb/s, and the corresponding average PSNR values were 35.05 and 34.25, respectively.

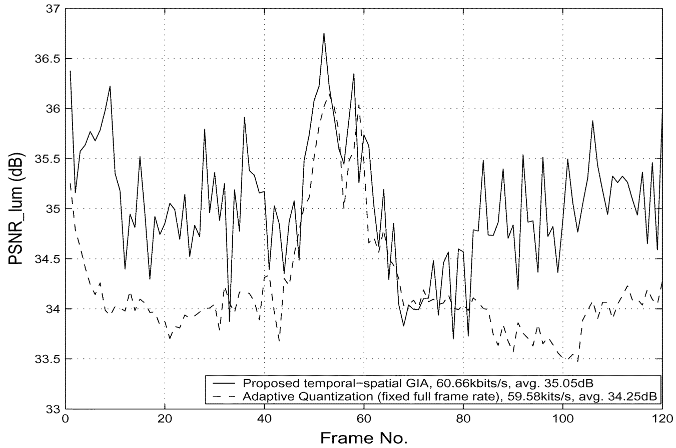
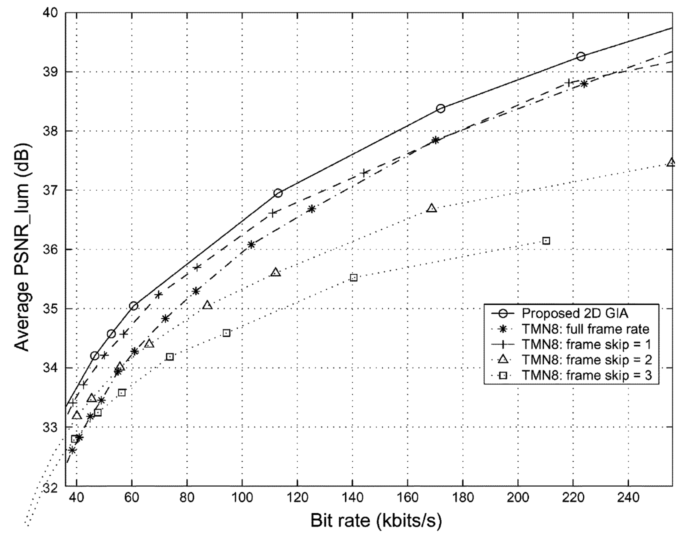
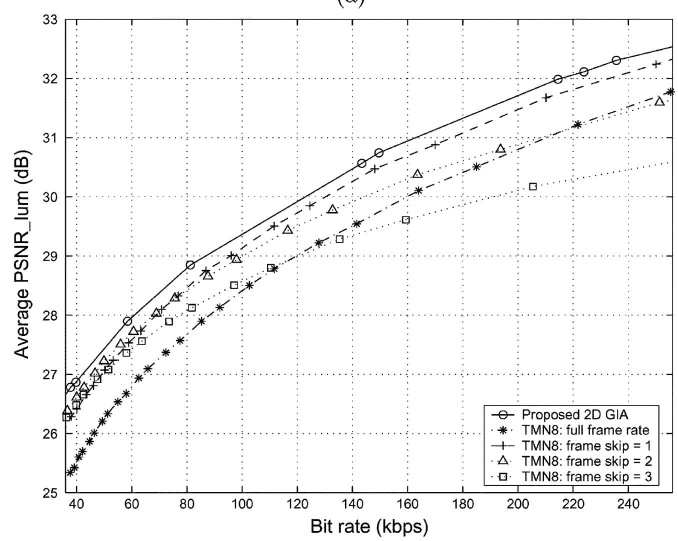


Fig. 9. Frame-to-frame quality comparison of the proposed GIA (solid) and spatial adaptive quantization (dashed) under the bit budget of 60 kb/s applied to Frames 1–120 of the "Suzie"sequence.



(a)



(b)

Fig. 10. R-D performance comparison of proposed GIA (-○-) and standard H.263+/TMN8 [1] with fixed FS/QP combinations ($\mathrm{FS} = 0$: -∗-, $\mathrm{FS} = 1$: -+-, $\mathrm{FS} = 2$: -◁-, and $\mathrm{FS} = 3$: -□-) applied to (a) Frames 1–120 of the "Suzie" sequence and (b) Frames 51–150 of the "Coastguard" sequence.

C. Frame-Based GIA versus H.263+/TMN8 With Fixed FS/QP

Fig. 10 provides the R-D performance comparison of the proposed GIA (-○-) and TMN8 with fixed FS and QP for the "Suzie" sequence (which is a talking head sequence with low temporal and spatial complexities) and the "Coastguard" sequence (with higher complexity). Affine MCI was used to reconstruct skipped frames in both cases. It can be seen that GIA achieves the higher average PSNR value than TMN8 with any fixed FS/QP combination. For the TMN8 coder with fixed FS/QP, reasonable FS results in better overall quality than the full frame rate encoded at low bit rates due to the contribution from affine MCI.

The frame to frame quality comparison between GIA and TMN8 with various fixed FS/QP combinations is given in Fig. 11. Frames 1–120 of the "Suzie" sequence were tested under the bit rate budget of 60 kb/s and affine MCI was used to reconstruct skipped frames in both cases. It can be seen that GIA adaptively selected FS/QP among frames with various temporal and spatial complexities so that better bit allocation was achieved. Besides significantly improving the average PSNR, the proposed method can also maintain constant overall quality as evidenced by a low PSNR variance value, which is preferred by the human visual system. More PSNR variance comparisons are shown in Table I.

Adaptive FS/QP selection by the proposed GIA applied to Frames 1–120 of the "Suzie" sequence under the bit budget of 60 kb/s is given in Fig. 12. Note that $QP = 0$ in the figure simply means that the frame is skipped (instead of actually setting $\mathrm{QP} = 0$). It shows that the full frame rate $(\mathrm{FS} = 0)$ is preferred when motion activities are high (around Frames 40–80) while finer QP with more FS is preferred when motion activities are low, e.g., in the beginning and the ending parts of this sequence.

Tables I and II present the performance in terms of the average PSNR and the PSNR variance for GIA and TMN8 with fixed FS/QP combinations under various budgets on test sequences "Suzie" and "Coastguard", respectively. From these results, we conclude that GIA always achieves higher overall quality (i.e., higher average PSNR) while maintaining constancy (i.e., lower PSNR variance).

D. GOF-Based GIA, Frame-Based GIA, and TMN8 With Fixed FS/QP

Fig. 13 presents the MAD curve of the "Suzie" sequence from the 1st to the 120th frame, where MAD is calculated via (24). The motion velocity is constant among frames with similar MAD, where the same FS/QP can be applied. The MAD differential (25) is calculated and shown in Fig. 14 to determine the variation of the motion velocity as well as spatial complexities. If the absolute value of MAD differential exceeds a predetermined threshold at a certain frame, a new GOF has to start from there. Figs. 15 and 16 present results of adaptive frame grouping on "Suzie" with threshold equal to 0.45 and 0.48. It shows that the smaller the threshold, the finer the grouping is. Comparing the adaptive frame grouping results with adaptive FS/QP selection achieved by frame-based GIA as given in Fig. 12, we see that adaptive frame grouping and adaptive FS/QP selection match each other well.

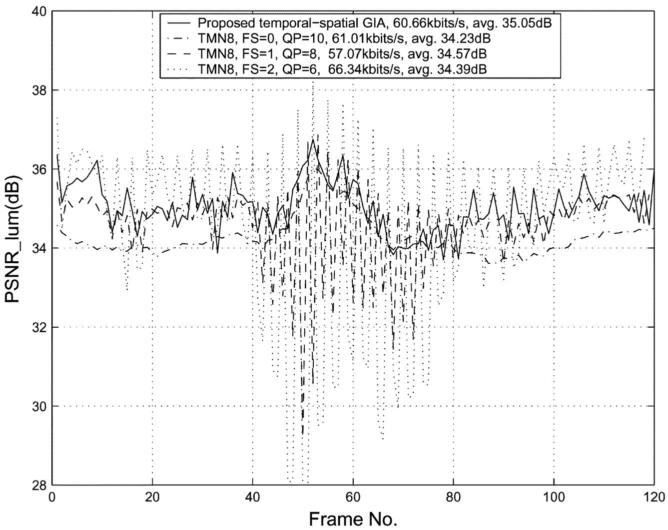


Fig. 11. Frame-to-frame quality comparison of the proposed GIA (solid) and TMN8 with fixed FS/QP combinations, ($\mathrm{FS} = 0/\mathrm{QP} = 10$: dotdash, $\mathrm{FS} = 1/\mathrm{QP} = 8$: dashed, $\mathrm{FS} = 2/\mathrm{QP} = 6$: dotted) under the bit rate budget of 60 kb/s applied to Frames 1–120 of the "Suzie" sequence.

TABLE I
TARGET AND ACTUAL BIT RATES, AVERAGE PSNR VALUES, AND PSNR VARIANCES OF THE PROPOSED GIA AND TMN8 WITH FIXED FS/QP COMBINATIONS FOR THE "SUZIE" SEQUENCE

Budget = 40kbits/s	Rate (kbits/s)	Average PSNR (dB)	PSNR Variance
2-D GIA (frame)	39.27	33.68	0.38
FS=0, QP=15	38.39	32.61	0.22
FS=1, QP=11	38.69	33.41	0.85
FS=2, QP=9	40.10	33.18	3.72
FS=3, QP=8	39.37	32.80	7.90
Budget = 60kbits/s	Rate (kbits/s)	Average PSNR (dB)	PSNR Variance
2-D GIA (frame)	60.66	35.05	0.38
FS=0, QP=10	61.01	34.23	0.25
FS=1, QP=8	57.07	34.57	1.39
FS=2, QP=6	66.34	34.39	6.14
FS=3, QP=6	56.37	33.58	10.78
Budget = 113kbits/s	Rate (kbits/s)	Average PSNR (dB)	PSNR Variance
2-D GIA (frame)	113.04	36.95	0.48
FS=0, QP=7	103.34	36.08	0.27
FS=1, QP=5	111.00	36.61	3.43
FS=2, QP=4	112.12	35.60	10.39
FS=3, QP=3	140.42	35.52	22.74

According to Fig. 16, the initial frame segmentation set was chosen to be $\{1, 2, 3, 5, 46, 47, 49, 54, 55, 56, 57, 75\}$. By eliminating frames whose locations are close to each other, the final frame segmentation set became $\{1, 2, 5, 46, 54, 75\}$. Thus, the six GOFs were:

- GOF 1: 1;
- GOF 2: 2, 3, 4;
- GOF 3: $5, 6, \ldots, 45$;
- GOF 4: $46, 47, \ldots, 53$;
- GOF 5: $54, 55, \ldots, 74$;
- GOF 6: $75, 76, \ldots, 120$.

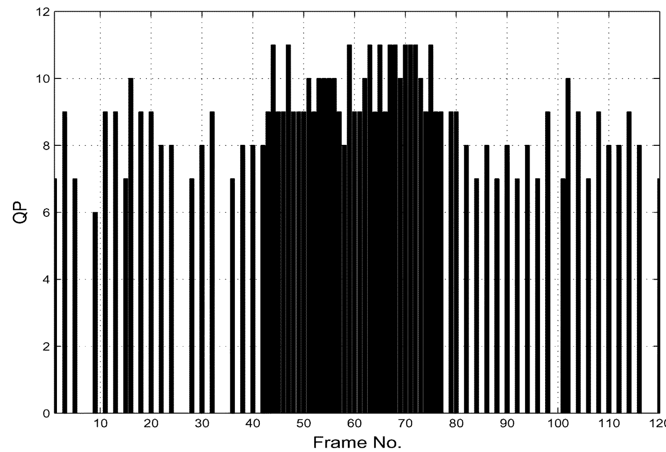


Fig. 12. Adaptive FS/QP allocation by GIA applied to Frames 1–120 of the "Suzie" sequence.

TABLE II
TARGET AND ACTUAL BIT RATES, AVERAGE PSNR VALUES, AND PSNR VARIANCES OF THE PROPOSED GIA AND TMN8 WITH FIXED FS/QP COMBINATIONS FOR THE "COASTGUARD" SEQUENCE

Budget = 40kbits/s	Rate (kbits/s)	Average PSNR (dB)	PSNR Variance
2-D GIA (frame)	40.80	26.78	0.42
FS=0, QP=29	40.80	25.60	0.28
FS=1, QP=22	40.10	26.41	0.62
FS=2, QP=19	40.01	26.61	1.90
FS=3, QP=17	39.92	26.48	3.85
Budget = 61kbits/s	Rate (kbits/s)	Average PSNR (dB)	PSNR Variance
2-D GIA (frame)	61.40	27.90	0.45
FS=0, QP=21	62.50	26.94	0.25
FS=1, QP=17	58.84	27.53	0.94
FS=2, QP=14	60.60	27.72	2.92
FS=3, QP=12	63.70	27.56	5.82
Budget = 84kbits/s	Rate (kbits/s)	Average PSNR (dB)	PSNR Variance
2-D GIA (frame)	84.20	28.85	0.39
FS=0, QP=17	85.30	27.90	0.22
FS=1, QP=13	86.84	28.75	1.50
FS=2, QP=11	87.62	28.66	4.12
FS=3, QP=10	81.80	28.13	7.14

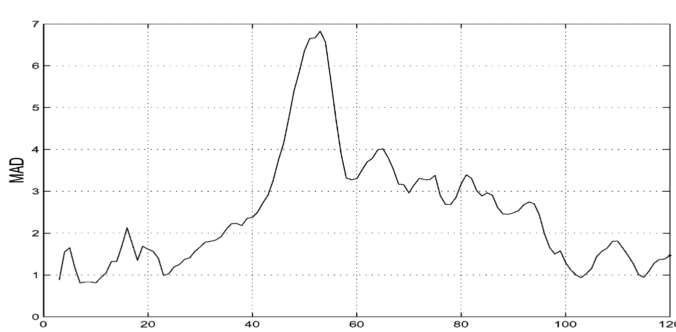


Fig. 13. MAD analysis of the "Suzie" sequence from Frames 1 to 120.

The GOF-based GIA is applied to the above six GOFs and MCI is used to reconstruct skipped frames. It is shown in Fig. 17 that the R-D performance of GOF-based GIA (-○-)is

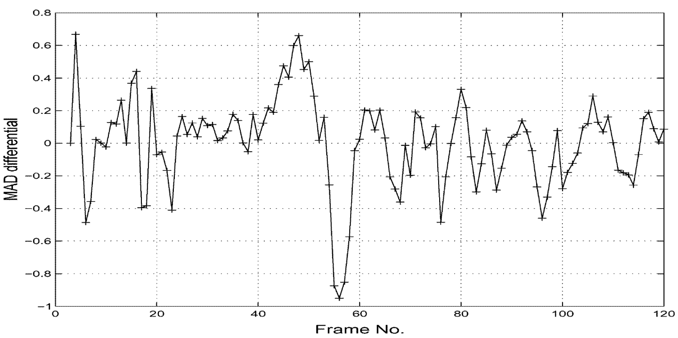


Fig. 14. MAD differential analysis of the "Suzie" sequence from Frames 1 to 120.

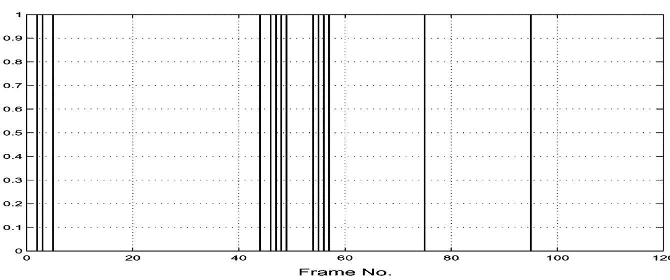


Fig. 15. Adaptive frame grouping result for the "Suzie" sequence based on MAD differential with threshold equal to 0.45.

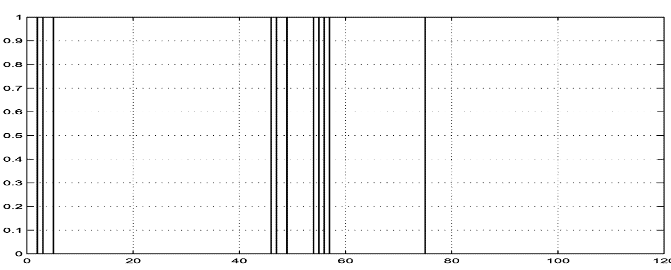


Fig. 16. Adaptive frame grouping result for the "Suzie" sequence based on MAD differential with threshold equal to 0.48.

slightly worse than that of frame-based GIA (-◇-), while much better than that of TMN8 without FS/QP adaptation (-⋆-). Its complexity is significantly reduced compared with frame-based GIA, and at the mean time, visual quality remains constant as evidenced by the low PSNR variance. An example is given in Table IV. The adaptively selected FS/QP for the six GOFs are given in Table IV.

The complexities of our proposed methods, including SIA, frame-based GIA and GOF-based GIA, are compared in Table V with the adaptive quantization method in [2] as the reference. Here, S, M, N, and $N'_{\rm GOF}$ stand for the maximum FS, the total number of candidate QP levels, the total number of frames and the total number of GOFs in the test sequence, respectively. Clearly, the complexity of SIA is very high. The complexity of the suboptimal heuristic, frame-based GIA, is significantly reduced. The complexity of our proposed fast approach, GOF-based GIA, can be lower than that of the reference, if frames with constant temporal and spatial characteristics are grouped efficiently so that $SN'_{\rm GOF} < N$ (e.g., the "Suzie" sequence illustrated earlier). The overall R-D performance of GOF-based GIA is typically enhanced in comparison with that of the reference.

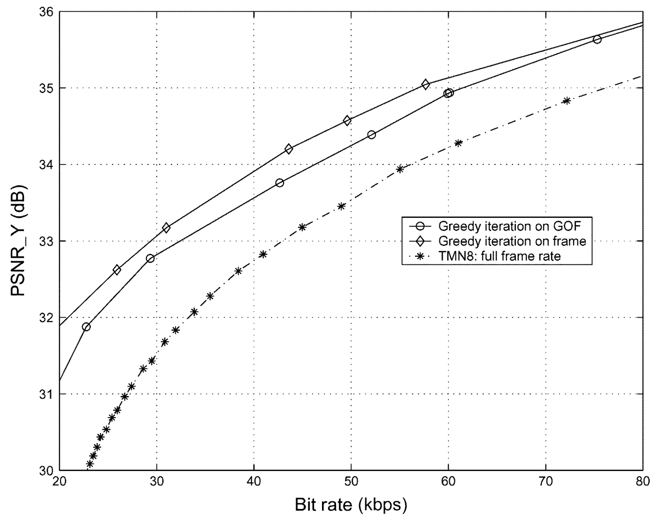


Fig. 17. R-D performance comparison between GOF-based GIA, frame-based GIA and TMN8 with fixed FS/QP for frames 1–120 of the "Suzie" sequence.

TABLE IV
TARGET AND ACTUAL BIT RATES, AVERAGE PSNR VALUES, AND PSNR VARIANCES OF GOF-BASED GIA, FRAME-BASED GIA AND TMN WITH FIXED FS/QP COMBINATIONS FOR THE "SUZIE" SEQUENCE

Budget = 40kbits/s	Rate (kbits/s)	Average PSNR (dB)	PSNR Variance
GIA (GOF)	40.66	33.66	0.43
GIA (frame)	39.27	33.68	0.38
FS=0, QP=15	38.39	32.61	0.22
FS=1, QP=11	38.69	33.41	0.85
FS=2, QP=9	40.10	33.18	3.72
FS=3, QP=8	39.37	32.80	7.90

TABLE V
COMPLEXITY COMPARISON OF PROPOSED METHODS

SIA	GIA (frame)	GIA (GOF)	Reference
$(O(M^N \sum_{i=0}^{N-2} C_{N-2}^i/M^i), O(SMN))$	$O(SMN)$	$O(SMN'_{GOF})$	$O(MN)$

The actual executable time depends on the hardware environment. When implemented in Pentium-II 600 MHz PC without code/assembly optimization, it demands more than 12 hours for bit allocation for ten frames of the "Suzie" QCIF with SIA, a couple of hours for 100 frames with frame-based GIA, and less than 20 min for 100 frames with GOF-based GIA. The above numbers however only provide a rough idea of relative execution speeds. The actual coding speed will be greatly enhanced with code/assembly optimization.

VI. CONCLUSION

Joint temporal–spatial bit allocation problem was addressed in this work and several heuristic solutions were proposed to solve the proposed problem with consideration of both frame prediction and frame interpolation dependency. Since classic optimization methods such as dynamic programming can only solve independent R-D optimization problems, exhaustive search is generally required to provide the optimal solution when the dependency among frames can not be ignored.

However, exhaustive search is infeasible due to its extremely high complexity. Thus, the SIA was proposed first to give the near-optimal R-D performance benchmark. To further reduce the complexity, the GIA was proposed as a suboptimal heuristic, which can achieve an R-D performance very close to that of SIA with a significantly reduced complexity. By adaptively grouping frames according to MAD differential and applying greedy iteration to GOF, the bit allocation process can be further greatly expedited with little sacrifice of quality. In all proposed techniques, both FS and QPs are adaptively and jointly determined so that the tradeoff between temporal and spatial qualities can be properly achieved. It was shown by experimental results that the proposed methods outperform the spatial adaptive quantization approach in [2] by taking advantage of adaptive FS. When compared with H.263+/TMN8 with fixed FS/QP, our proposed methods enhance the overall video quality at the full frame rate with higher average PSNR and lower PSNR variance values.

REFERENCES

[1] M. Gallant, G. Cote, B. Erol, and F. Kossentini, UBCs H.263+ public domain software, version 3.2.0, in Official ITU-T Study Group 16 Video Experts Group Reference Codec, 1998.

[2] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with application to multiresolution and MPEG video coders," *IEEE Trans. Image Process.*, vol. 3, no. 9, pp. 533–545, Sep. 1994.

[3] *Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 mbit/s*, June 1996. ISO/IEC JTC1/SC29/WG11.

[4] *Generic Coding of Moving Pictures and Associated Audio Information*, Oct. 2000. ISO/IEC JTC1/SC29/WG11.

[5] *Overview of the MPEG-4 Standard*, Mar. 2002. ISO/IEC JTC1/SC29/WG11.

[6] *Video Codec for Audiovisual Services at* $\mathrm{p} \times 64$ *kbits*, Mar. 1993. ITU-T Recommendation H.261.

[7] *Video Codec for Low Bitrate Communication*, May 1996. ITU-T Recommendation H.263.

[8] *The Emerging JVT/H.26L Video Coding Standard*, Sep. 2002. ICIP Tutorial.

[9] A. M. Tekalp, *Digital Video Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[10] A. Ortega, "Optimization techniques for adaptive quantization of image and video under delay constraints," Ph.D. dissertation, Dept. Elect. Eng., Columbia Univ., New York, Jun. 1994.

[11] H. Song, J. Kim, and J. Kuo, "Real-time H.263+ frame rate control for low bit rate VBR video," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 4, 1999, pp. 307–310.

[12] E. C. Reed and J. S. Lim, "Multidimensional bit rate control for video communication," in *Proc. SPIE*, vol. XXIII, San Diego, CA, Jul. 2000, pp. 277–288.

[13] D. P. Bertsekas, *Dynamic Programming: Deterministic and Stochastic Models*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[14] T. Kuo and C.-C. J. Kuo, "Motion-compensated interpolation for low-bit-rate video quality enhancement," in *Proc. SPIE Applicat. Digital Image Process.*, vol. 3460, 1998, pp. 277–288.

[15] K. Zhang, M. Sober, and J. Kittler, "Video coding using affine motion compensated prediction," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, vol. 4, 1996, pp. 1978–1981.

Shan Liu received the B.E. degree in electronic engineering from Tsinghua University, Beijing, China and the M.S. and Ph.D. degrees from the University of Southern California, Los Angeles, both in electrical engineering.

She is currently with Sony R&D Laboratory, San Jose, CA. Before joining Sony, she had worked and interned with several Research labs including Samsung Information Systems America, IBM T. J. Watson Research Center, Rockwell Science Center, AT&T Labs. Research, as well as industrial companies such as InterVideo Inc. She has broad interests in the field of multimedia technologies, with emphasis on video compression, transmission and analysis.

C.-C. Jay Kuo (S'86–M'87–SM'92–F'99) received the B.S. degree from the National Taiwan University, Taipei, Taiwan, R.O.C., in 1980 and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1985 and 1987, respectively, all in electrical engineering.

He is with the Department of Electrical Engineering, the Signal and Image Processing Institute (SIPI), and the Integrated Media Systems Center (IMSC) at the University of Southern California (USC), Los Angeles, as Professor of Electrical Engineering and Mathematics. His research interests are in the areas of digital media processing, multimedia compression, communication and networking technologies, and embedded multimedia system design. He has guided 55 students to their Ph.D. degrees and supervised 15 postdoctoral research fellows. Currently, his research group at USC consists of around 30 Ph.D. students and five postdoctors (please visit website http://viola.usc.edu), which is one of the largest academic research groups in multimedia technologies. He is coauthor of more than 700 technical publications in international conferences and journals as well as the following seven books: *Content-based Audio Classification and Retrieval for Audiovisual Data Parsing* (with Tong Zhang, New York: Kluwer, 2001), *Semantic Video Object Segmentation for Content-based Multimedia Applications* (with Ju Guo, New York: Kluwer, 2001), *Intelligent Systems for Video Analysis and Access over the Internet* (with Wensheng Zhou, Englewood Cliffs, NJ: Prentice-Hall, 2002), *Video Content Analysis Using Multimodal Information* (with Ying Li, New York: Kluwer, 2003) *Quality of Service for Internet Multimedia* (with Jitae Shin and Daniel Lee, Englewood Cliffs, NJ: Prentice-Hall, 2003), *Radio Resource management for Multimedia QoS support in Wireless Cellular Networks* (with Huan Chen, Lei Huang, and Sunil Kumar, New York: Kluwer, 2003) and *High Fidelity Multichannel Audio Coding* (with Dai Tracy Yang and Chris Kyriakakis, Sylvania, OH: Hindawi, 2004).

Dr. Kuo is a Fellow of SPIE. He received the National Science Foundation Young Investigator Award (NYI) and Presidential Faculty Fellow (PFF) Award in 1992 and 1993, respectively. He is Editor-in-Chief for the *Journal of Visual Communication and Image Representation*, Editor for *the Journal of Information Science and Engineering*, and *the RURASIP Journal of Applied Signal Processing*. He is also on the Editorial Board of the *IEEE Signal Processing Magazine*. He served as Associate Editor for *IEEE Transactions on Image Processing* in 1995–1998, *IEEE Transactions on Circuits and Systems for Video Technology* in 1995–1997 and IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING in 2001–2003.