Transactions Letters

A Rate-Distortion Optimal Hybrid Scalable/Multiple-Description Video Codec

Lisimachos P. Kondi

***Abstract*—We propose a novel hybrid scalable/multiple description codec which produces a base layer and two multiple description enhancement layers. The base layer is required for decoding. If one or two of the multiple description enhancement layers are also received, the signal-to-noise ratio of the received video sequence is improved. There is no hierarchy in the multiple description enhancement layers. Thus, in contrast with classic scalable coding (SC), if either enhancement layer is lost, the other enhancement layer is still useful to the decoder. This functionality comes at a minimal expense in compression efficiency compared with SC. The layers are constructed using a rate-distortion optimal partitioning of the discrete cosine transform coefficients. Experimental results are presented and conclusions are drawn.**

***Index Terms*—Multiple description coding (MDC), rate-distortion optimization, scalable coding (SC).**

I. Introduction

THERE are two main paradigms for layered video coding: scalable coding (SC) [1]–[7], and multiple description coding (MDC) [8]–[19]. A scalable codec produces a bitstream that can be partitioned into layers that form a hierarchy. One of the layers is called the base layer and is required for video reconstruction. The other layers, called enhancement layers, can be decoded along with the base layer and produce a video sequence of improved quality. However, in order for an enhancement layer to be useful in decoding, the base layer and all hierarchically higher enhancement layers also need to be available to the decoder. Thus, if a scalable encoder produces one base layer and two enhancement layers and the decoder receives only the base layer and the second enhancement layer, the latter cannot be used in the decoding and the decoder is only able to reconstruct a video sequence of base layer quality.

In contrast with scalable codecs, multiple description codecs produce layers that do not form a hierarchy. Thus, any of the layers can be decoded independently and produce a video sequence of a certain quality. Furthermore, the more layers are available to the decoder, the better the reconstructed video quality. As there is no hierarchy, any layer that is received by the decoder is used in the reconstruction. In order for this to be possible, the multiple description layers need to share some information; thus the layers are correlated. This correlation causes a decrease in coding efficiency.

Manuscript received November 26, 2003; revised July 21, 2004. This paper was recommended by Associate Editor M. Strintzis.

The author is with the Department of Electrical Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260 USA (e-mail: lkondi@eng.buffalo.edu).

Digital Object Identifier 10.1109/TCSVT.2005.848344

We propose a novel hybrid scalable/multiple-description codec (HSMDC) that combines the advantages of the SC and MDC paradigms. The HSMDC codec produces a bitstream that consists of a base layer and several multiple description enhancement layers. The base layer is required for video reconstruction. If one or more multiple-description enhancement layers are received in addition to the base layer, they can be used in the decoding and improve video quality. Thus, there is no hierarchy in the enhancement layers and any received enhancement layer is useful in decoding as long as the base layer has been successfully received. The compression efficiency of HSMDC is better than that of MDC, since only the enhancement layers need to be correlated. Furthermore, as we show in the experimental results, the compression efficiency of HSMDC is close to that of SC, since the correlation between enhancement layers does not need to be as high as in the case of MDC. This is due to the fact that the most important information is transmitted with the base layer, which needs not be correlated with any other layer. Thus, HSMDC relaxes the hierarchy of SC by only requiring the base layer to be successfully received and provides nonhierarchical enhancement layers at the expense of a small reduction in compression efficiency.

The proposed HSMDC scheme is aimed to be used in video transmission over lossy channels in conjunction with unequal error protection (UEP). UEP can be achieved using appropriate channel coding techniques. Furthermore, the future Internet will offer different priority classes for different packets. In both SC and the proposed HSMDC, the base layer needs to be better protected than the enhancement layers. In a three-layer SC scheme (one base and two enhancement layers), if only the base layer and the second enhancement layer are received (the first enhancement layer is lost), the second enhancement layer is useless to the receiver, which will only be able to decode the video at a base layer quality. In the proposed HSMDC scheme with one base and two multiple description enhancement layers, either enhancement layer will be useful in decoding as long as the base layer is successfully received and give a video quality that is better than the base layer video quality that the SC scheme would give if the first enhancement layer was lost. This is exactly the advantage of the proposed HSMDC scheme over SC. MDC does not require UEP, however, since the descriptions need to be correlated, its compression efficiency is much lower than both SC and HSMDC. MDC schemes have the edge in situations when UEP cannot be achieved.

In related work, combinations of SC and MDC have been proposed in [20], [21], but these approaches are very different from ours. In [20], the base layer is encoded using MDC (instead of the enhancement layer in our case) and three possible single description enhancement layers are encoded, depending on whether either one or both base layer descriptions are received, or expected to be received by the decoder. The transmitter can choose to transmit one or both base layer descriptions along with the appropriate enhancement layer, according to the expected channel conditions. In [21], the authors start with a multiple description code and convert its descriptions to make them scalable.

Scalability is supported by most of the current motion- compensated discrete cosine transform-based (MC-DCT) video compression standards such as MPEG-2, MPEG-4, and H.263. Version 2 of the H.263 standard (also known as H.263+) [1], [22] supports SNR scalability as well as spatial and temporal scalability. SNR scalability implies that the enhancement in quality translates in an increase in the SNR of the reconstructed video sequence, while spatial and temporal scalability imply that the spatial and temporal resolutions, respectively, are increased.

The MPEG-4 standard also supports fine granularity scalability (FGS) [7]. In FGS, the video sequence is encoded into a base layer and an enhancement layer. For the enhancement layer, the difference between the original picture and the base layer reconstructed picture is encoded using bit-plane coding of the DCT coefficients. Thus, the enhancement layer bitstream can be truncated at any point while still being able to be decoded (yielding lower video quality).

It has been shown in [23] that, for transmission over error-prone channels, it is advantageous to use scalability and apply stronger error protection to the base layer than to the enhancement layers (UEP). Thus, we can expect a basic reconstructed quality with high probability even during adverse channel conditions. Had we not used scalability but instead protected the whole bitstream equally, there would be a much higher probability of catastrophic errors that would result in a reconstructed video sequence of poor quality.

Some of the early theoretical work on MDC appears in [8], [12], [15]. The most recent multiple description image and video coding techniques include pairwise correlating transforms [11], [13], [16], [17], SC in conjunction with UEP [18], [19], correlating filter banks [10] and coefficient splitting in the DCT domain [9]. Video redundancy coding (VRC), which is supported by the H.263 standard [22], can also be seen as a MDC technique.

The proposed hybrid scalable/multiple-description algorithm operates in two steps. It is first assumed that the DCT of the displaced frame difference (DFD) (or the intensity for intrablocks) is taken and quantized. Then, during the first step, the base layer is constructed for each block by subtracting a suitable value from each quantized coefficient (see Fig. 1). The subtracted values then become the enhancement layer. The determination of the subtracted values is optimal in the operational rate-distortion sense. In the second step of the algorithm, the enhancement layer obtained in the first step is converted into two multiple description enhancement layers by selecting a threshold for each block and duplicating into both descriptions the quantized coefficients with values equal to or greater than the threshold while alternating between descriptions the other coefficients, in a fashion similar to the algorithm in [9]. Again, the determination of the threshold is optimal in the operational rate-distortion sense. Some preliminary results on HSMDC have appeared in [24].

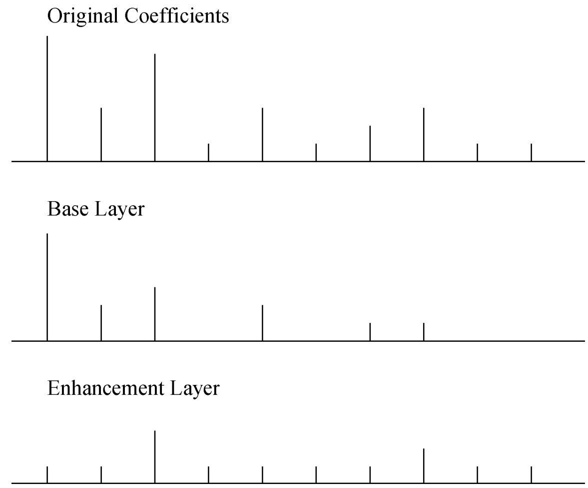

Fig. 1. Proposed partitioning of DCT coefficients for SNR scalability.

The rest of the paper is organized as follows. In Section II, the algorithm for the determination of the base layer is presented, while in Section III, the determination of the multiple description enhancement layers is discussed. In Section IV, experimental results are presented. Finally, in Section V, conclusions are drawn.

II. Step One: Determination of the Base Layer

A. Basic Codec Structure

The proposed codec is based on the architecture of the H.263 video compression standard. However, any Motion-Compensated DCT-based codec can be used as a basis. In H.263 with QCIF-sized frames, one group-of-blocks (GOB) consists of one line of 16×16 macroblocks (11 macroblocks). Each macroblock consists of four luminance and two chrominance 8×8 blocks. The DCT transform of the DFD (or the intensity for intrablocks) of a block is taken and quantized. Then, a triplet (LEVEL, RUN, LAST) is transmitted using suitable variable length code (VLC) tables, where LEVEL is the quantization level of the coefficient, RUN is the number of zero-valued coefficients that precede it and LAST specifies whether the current coefficient is the last in the block. An extra bit is appended to the VLC to denote the sign of LEVEL. Therefore, in the following discussion, LEVEL will refer to the absolute value of the quantization index.

B. Rate-Distortion Problem Formulation

In forming an SNR scalable bitstream, the following problem is formulated and solved. Let X be the set of original (unquantized) DCT coefficients in a frame and C the set of quantization levels that results from the quantization of X with quantization parameter (QP).

Our goal is, given a set of DCT coefficients X with corresponding quantization levels C and "dequantized" values $\hat{X}$, to find a set of quantization levels $\tilde{C}$ by subtracting a certain value

l_i from each coefficient quantization level C_i, so that a bit constraint is satisfied. The value l_i can be different for each coefficient quantization level C_i. The set of "dequantized" values that corresponds to $\tilde{C}$ is $\tilde{X}$. We will call $\tilde{X}$ a *trimmed* version of $\hat{X}$. The set of quantization levels $\tilde{C}$ is transmitted as the base layer (along with motion vectors and overhead information). Then, given a bit budget for the base layer, our problem is to find $\tilde{C}$ as the solution to the constrained problem

$$\min_{\tilde{C}} \left[D(X, \tilde{X})|C\right] \text{ subject to } R(\tilde{C}) \leq R_{\text{budget}} \quad (1)$$

where $D(.,.)$ and $R(.)$ are the distortion and rate functions, respectively, and R_{budget} is the available bit budget for the base layer.

The problem of (1) can be solved using Lagrangian relaxation. The problem now becomes the minimization of the Lagrangian cost

$$J_1(\lambda) = D(X, \tilde{X}|C) + \lambda R(\tilde{C}) \quad (2)$$

and the specification of the Lagrange multiplier λ so that the budget constraint is satisfied.

Without lack of generality, in our implementation of the algorithm, we determine a bit budget for the base layer for a GOB. This is done because an outside rate control mechanism updates the QP at the beginning of each GOB and thus determines the total available bit budget for the GOB (for all scalable layers). The bit budget for the base layer is a fixed percentage of the total available bit budget for the GOB. This percentage is determined by the target bit rates for each scalable layer. $J_1(\lambda)$ can be expressed as the sum of individual Lagrangian costs (one for each block)

$$J_1(\lambda) = \sum_i J_{1,\text{block},i}(\lambda), \quad (3)$$

where

$$J_{1,\text{block},i}(\lambda) = D_{\text{block},i}(X, \tilde{X}|C) + \lambda R_{\text{block},i}(\tilde{C}) \quad (4)$$

is the Lagrangian cost for block i. Clearly, in order for (2) and (3) to be equivalent, the same λ should be used for all $J_{1,\text{block},i}(\lambda)$. Since the encoding of the DCT coefficients is done independently for each block (except for the dc coefficient of intrablocks which is differentially encoded and transmitted with the base layer anyway), the minimization of $J_1(\lambda)$ can be performed through the independent minimization of each of the $J_{1,block,i}(\lambda)$, using the same λ [25], [26] for each block. The λ for which the bit budget is met is found iteratively. A large λ results in a point in the rate-distortion curve with low rate and high distortion. Conversely, a small λ results in a point with high rate and low distortion. Therefore, a simple method, such as bisection, can be used to find the desired λ iteratively. More sophisticated algorithms, such as, the fitting of a Bezier curve [26], can also be used.

The problem now reduces to finding the set of quantization levels $\tilde{C}$ and corresponding trimmed DCT coefficients $\tilde{X}$ for every block that would minimize the Lagrangian cost of the block (4) for a given λ. The admissible candidate set $\tilde{C}$ is constructed as follows. Each nonzero coefficient in the block with quantization level C_i is either dropped completely or a value $l_i < C_i$ is subtracted from it. Although there is a finite number of admissible sets $\tilde{C}$, the minimization of the Lagrangian cost in (4) using exhaustive search is computationally prohibitive. The problem has however a structure which can be exploited using dynamic programming (DP) for its solution. The details of the DP algorithm can be found in [6], but are also summarized here. Similar DP algorithms have been proposed in [27], [28] for different applications (quality improvement for nonscalable JPEG or MPEG codecs).

C. DP Solution

As mentioned in the previous section, the 2-D DCT coefficients are ordered in one dimension using the zig-zag scan and encoded using VLCs that correspond to the triplets (LEVEL, RUN, LAST). Let us assume for a moment that the coefficients are coded using pairs (LEVEL, RUN), i.e., the same VLC is used whether the coefficient is the last nonzero coefficient in the block or not. We will explain the modifications to the algorithm for (LEVEL, RUN, LAST) later. Then, suppose that we consider the problem of minimizing the Lagrangian cost given that coefficient k is the last nonzero coefficient in the block to be coded and coefficients $k+1$ to 63 are all thresholded to zero. Assuming that we have the solution to this problem, it can be used to solve the problem when coefficient k' is the last nonzero coefficient, where $k' > k$.

Let us utilize the incremental Lagrangian cost $\Delta J_{j,k}^{l_k}$ as the difference in the cost incurred by including coefficient k trimmed by l_k in the base layer when the previous nonzero coefficient is j. It is defined by

$$\Delta J_{j,k}^{l_k} = -E_k^{l_k} + \lambda R_{j,k}^{l_k} \quad \text{for} \quad j < k, \quad (5)$$

where $E_k^{l_k}$ represents the difference in distortion incurred by including coefficient k and is defined by $E_k^{l_k} = X_k^2 - (X_k - \tilde{X}_k^{l_k})^2$, X_k is the original kth unquantized coefficient and $\tilde{X}_k^{l_k}$ is the "dequantized" coefficient which corresponds to quantization level $\tilde{C}_k^{l_k} = C_k - l_k$, with C_k the original quantization level, and $R_{j,k}^{l_k}$ is the rate (in bits) that would be required to encode quantization level $\tilde{C}_k^{l_k}$ given that the previous nonzero coefficient was coefficient j.

If we drop all ac coefficients of an intrablock, the rate will be zero and the distortion will be equal to

$$J_0^* = E_{\text{intra}} = \sum_{i=1}^{63} X_i^2 \quad (6)$$

since the DCT transform is unitary and we can therefore calculate the mean squared error in either the spatial or the frequency domain. For interblocks, we allow for the possibility of dropping all coefficients, including the dc. Then, we define

$$J_{-1}^* = E_{\text{inter}} = \sum_{i=0}^{63} X_i^2. \quad (7)$$

As mentioned earlier, we need to take into account the fact that different VLCs are used depending on whether the coefficient to be encoded is the last one in the block or not. Therefore, we define a second incremental cost $\Delta J_{j,k,\text{last}}^{l'_k} = -E_j^{l'_k} +$

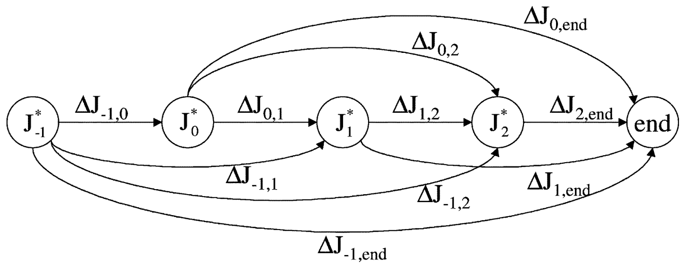

Fig. 2. Directed acyclic graph representation of the optimal DCT coefficient partitioning problem.

$\lambda R^{l'_k}_{j,k,\mathrm{last}}$, where $R^{l_k}_{j,k,\mathrm{last}}$ is the number of bits that are required to encode quantization level $\tilde{C}^{l_k}_k$ given that j was the previous nonzero coefficient and coefficient k is the last one to be encoded in the block. We also keep the minimum Lagrangian costs $J'_{k,last}$ for each coefficient k given that it is the last coefficient to be coded in the block.

For $k=1$, we can either keep coefficients 0 and 1 or just coefficient 1. Again, we need to determine the value to be subtracted from C_1. Now, the minimum cost will be $J'_1 = \min_{i,l_1} J^*_i + \Delta J^{l_1}_{i,1}$, for $i=-1,0$. We also need to calculate $J^*_{1,\mathrm{last}}$ in a similar manner.

For a general k, the minimum costs are found as

$$J^*_k = \min_{i,l_k} \left\{ J^*_i + \Delta J^{l_k}_{i,k} \right\}, \quad \text{for} \quad i=-1,\ldots,k-1 \tag{8}$$

and

$$J^*_{k,\mathrm{last}} = \min_{i,l'_k} \left\{ J^*_i + \Delta J^{l'_k}_{i,k,\mathrm{last}} \right\}, \quad \text{for} \quad i=-1,\ldots,k-1. \tag{9}$$

The algorithm calculates J^*_k and $J^*_{k,last}$ for all $k=0,\ldots,63$ and also stores the last nonzero coefficients (predecessors) i and the subtracted values l_k and l'_k which minimize (8) and (9), respectively. The k which results in the minimum $J^*_{k,\mathrm{last}}$ will be denoted as k^*. Clearly, $J^*_{k^*,\mathrm{last}}$ is equal to the minimum Lagrangian cost J_{block} for the whole block. Therefore, we know that coefficient k^* will be included in the base layer and we look up the value to be subtracted from it. Then, we look up the optimal predecessor i that resulted in $J^*_{k^*,\mathrm{last}}$. Let us denote this coefficient as k_1. Then, k_1 will be included in the base layer and the value to be subtracted from it is looked up. Then we look up the predecessor that resulted in $J^*_{k_1}$ and continue recursively in the same fashion until we arrive at the imaginary coefficient -1. "Pruning" of nonoptimal predecessors i in (8) and (9) can also be performed as explained in [6].

It is interesting to point out that the proposed algorithm is equivalent to finding the shortest path in an directed acyclic graph (DAG). Fig. 2 shows a DAG for the case of just three DCT coefficients (instead of 64). The vertices of the DAG correspond to the Lagrangian costs J^*_k while the edges correspond to the differential costs $\Delta J_{i,k}$. For simplicity, in this graph we assume that coefficients can either be included or dropped from the base layer, i.e., no "trimming" is involved. The first vertex takes the value $J^*_{-1} = E$, where E is equal to E_{intra} or E_{inter}, depending on the type of the macroblock. The last vertex designated as "end" is needed to show that the last coefficient for the block has been encoded. Clearly, $\Delta J_{i,\mathrm{end}} = 0$ for all i. The solution of finding the shortest path of the DAG is exactly the algorithm we described.

III. Step Two: Determination of the Multiple Description Enhancement Layers

We have thus far presented an optimal algorithm for partitioning a set of quantized DCT coefficients into two layers. We are now ready to split the second layer into two multiple description enhancement layers, by also adding an appropriate amount of redundancy.

Let us assume that we have already partitioned the DCT coefficients into two layers and the set of coefficient quantization levels for the second layer is C_{enh}. We now want to partition C_{enh} into two sets of coefficients, namely C_2 and C_3. The coefficients of the base layer C_1 have already been selected during the partitioning of the coefficients into two layers. Let X_{1+2} be the "dequantized" DCT coefficients when layer 1 (base layer) and layer 2 (first multiple description enhancement layer) are utilized. Similarly, let X_{1+3} be the "dequantized" DCT coefficients when layer 1 (base layer) and layer 3 (second multiple description enhancement layer) are utilized. In order to construct X_{1+2} and X_{1+3}, the decoder adds up the corresponding quantization levels before "dequantization", as in the case of the scalable video codec in [6], [29]. The two sets of coefficients C_2 and C_3 are constructed from C_{enh} by determining a threshold T for each 8×8 image block and duplicating on both C_2 and C_3 those coefficient quantization levels that are equal to or greater than the threshold, while alternating between layers the remaining coefficients. Thus, the first coefficient that is below the threshold will go to C_2, the second one to C_3 and so on. Clearly, the smaller the threshold T, the greater the redundancy introduced. Let T_{GOB} be the set of thresholds for a whole GOB. Given the algorithm of the creation of the multiple description enhancement layers, it is expected that both multiple description enhancement layers will have similar rate and distortion

$$[D(X, X_{1+2})|C_{\mathrm{enh}}] \simeq [D(X, X_{1+3})|C_{\mathrm{enh}}] \tag{10}$$

$$R(C_2) \simeq R(C_3). \tag{11}$$

Thus, we choose to define the problem of determining the multiple-description enhancement layers as follows: Determine the set of thresholds T_{GOB} such that

$$\min_{T_{\mathrm{GOB}}} [D(X, X_{1+2})|C_{\mathrm{enh}}] \text{ subject to } R(C_2) + R(C_3) \leq R_{\mathrm{budget,enh}}. \tag{12}$$

where $R_{\mathrm{budget,enh}} = R(C_{\mathrm{enh}}) \cdot \alpha$, where $1 \leq \alpha \leq 2$ determines the added redundancy. $\alpha = 1$ corresponds to $R(C_2) + R(C_3) = R(C_{\mathrm{enh}})$ (no added redundancy) while $\alpha = 2$ corresponds to $C_2 = C_3 = C_{\mathrm{enh}}$.

As in Step 1, the constrained optimization problem of (12) can be converted to an unconstrained optimization problem through the use of Lagrangian multipliers. Thus, we have the Lagrangian cost

$$J_2(\mu) = [D(X, X_{1+2})|C_{\mathrm{enh}}] + \mu R(C_2). \tag{13}$$

TABLE I
PSNR COMPARISON BETWEEN HSMDC AND SC FOR THE "FOREMAN" SEQUENCE AT 32-48-64 kb/s

	$\alpha = 1.0$	$\alpha = 1.4$	SC
Layer 1	29.30	29.24	29.35
Layers 1+2	30.21	30.78	30.66
Layers 1+3	30.47	30.56	N/A
Layers 1+2+3	31.68	31.31	31.73

TABLE II
PSNR COMPARISON BETWEEN HSMDC AND SC FOR THE "FOREMAN" SEQUENCE AT 64-96-128 kb/s

	$\alpha = 1.0$	$\alpha = 1.4$	SC
Layer 1	31.88	31.69	31.70
Layers 1+2	32.47	32.63	32.54
Layers 1+3	32.56	32.45	N/A
Layers 1+2+3	33.27	32.92	33.29

TABLE III
PSNR COMPARISON BETWEEN HSMDC AND SC FOR THE "FOREMAN" SEQUENCE AT 96-144-192 kb/s

	$\alpha = 1.0$	$\alpha = 1.4$	SC
Layer 1	32.97	33.11	33.01
Layers 1+2	33.78	34.19	34.11
Layers 1+3	33.83	34.00	N/A
Layers 1+2+3	34.87	34.60	34.89

TABLE IV
PSNR COMPARISON BETWEEN HSMDC AND SC FOR THE "FOREMAN" SEQUENCE AT 128-192-256 kb/s

	$\alpha = 1.0$	$\alpha = 1.4$	SC
Layer 1	34.24	34.33	34.30
Layers 1+2	35.15	35.52	35.44
Layers 1+3	35.16	35.31	N/A
Layers 1+2+3	36.31	36.00	36.30

TABLE V
PSNR COMPARISON BETWEEN HSMDC AND SC FOR THE "AKIYO" SEQUENCE AT 32-48-64 kb/s

	$\alpha = 1.0$	$\alpha = 1.4$	SC
Layer 1	39.12	39.21	39.20
Layers 1+2	39.65	39.82	39.81
Layers 1+3	39.66	39.74	N/A
Layers 1+2+3	40.27	40.08	40.29

TABLE VI
PSNR COMPARISON BETWEEN HSMDC AND SC FOR THE "AKIYO" SEQUENCE AT 64-96-128 kb/s

	$\alpha = 1.0$	$\alpha = 1.4$	SC
Layer 1	40.59	40.68	40.61
Layers 1+2	40.91	41.09	40.98
Layers 1+3	40.92	41.03	N/A
Layers 1+2+3	41.28	41.26	41.25

As before, $J_2(\mu)$ can be broken into a sum of individual Lagrangian costs, one for each block i

$$J_{2,\text{block},i}(\mu) = [D_{\text{block},i}(X, X_{1+2})|C_{\text{enh}}] + \mu R_{\text{block},i}(C_2). \quad (14)$$

Then, the appropriate μ that will meet the target bit rate can be determined using a method like the bisection method, as in Step 1. A similar technique to the one we describe here has been proposed in [9] for MDC.

The HSMDC decoder works as follows. If only the base layer is available, it is decoded by itself, just like nonscalable H.263. If the base plus one of the two multiple description enhancement layers are available, the decoder adds up the corresponding quantization indexes before inverse quantization. If the base plus both multiple description enhancement layers are available, the decoder first processes the multiple description enhancement layer in order to reconstruct the original enhancement layer produced in Step 1. The decoder compares the values of the coefficients of the two layers. If the value of a particular coefficient is the same in both layers, this means that this coefficient was duplicated on both layers in Step 2. If the value of a particular coefficient is zero in one layer and nonzero in the other, this means that the cofficient was alternated in Step 2. This way, the decoder is able to recover the enhancement layer produced in Step 1. Then, it adds that layer to the base layer and performs "inverse quantization."

IV. EXPERIMENTAL RESULTS

We next present experimental results using the proposed HSMDC. In order to gauge the performance penalty incurred in utilizing multiple description enhancement layers instead of hierarchical enhancement layers, we compare the proposed HSMDC codec with the three-layer scalable codec in [6], [29]. The "Foreman," "Akiyo," and "Container" sequences were used in the experiments. Four different sets of target bit rates were used. For both codecs, an external rate control was used to maintain the total rate for all layers at 64, 128, 192, and 256 kb/s. For both encoders, the base layer rate was set to be equal to 50% of the total rate (32, 64, 96, and 128 kb/s, respectively). In all cases, the two enhancement layers had equal bit rates. Thus, the bit rates for layer 1, layers 1 and 2 and layers 1, 2, and 3 for the four cases considered were 32-48-64, 64-96-128, 96-144-192, and 128-192-256 kb/s. Clearly, in all cases, the total target bit rate for layers 1 and 3 is equal to the total target bit rate for layers 1 and 2. However, only the proposed HSMDC codec can decode layers 1 and 3. In the SC codec, layer 3 would be useless and it would only be able to decode the video at base layer quality. Two different values for α were used: $\alpha = 1$ and $\alpha = 1.4$.

Tables I–IV show results for the "Foreman" sequence, while Tables V–VIII and Tables IV–XII show results for the "Akiyo" and "Container" sequences, respectively. Since the most important video information goes to the base layer, the multiple de-

TABLE VII
PSNR COMPARISON BETWEEN HSMDC AND SC FOR THE "AKIYO" SEQUENCE AT 96-144-192 kb/s

	$\alpha = 1.0$	$\alpha = 1.4$	SC
Layer 1	41.62	41.71	41.63
Layers 1+2	42.02	42.17	42.16
Layers 1+3	42.02	42.01	N/A
Layers 1+2+3	42.46	42.27	42.44

TABLE VIII
PSNR COMPARISON BETWEEN HSMDC AND SC FOR THE "AKIYO" SEQUENCE AT 128-192-256 kb/s

	$\alpha = 1.0$	$\alpha = 1.4$	SC
Layer 1	42.70	42.75	42.71
Layers 1+2	43.11	43.26	43.11
Layers 1+3	43.12	43.19	N/A
Layers 1+2+3	43.58	43.47	43.54

TABLE IX
PSNR COMPARISON BETWEEN HSMDC AND SC FOR THE "CONTAINER" SEQUENCE AT 32-48-64 kb/s

	$\alpha = 1.0$	$\alpha = 1.4$	SC
Layer 1	34.82	34.89	35.01
Layers 1+2	35.19	35.32	35.41
Layers 1+3	35.20	35.25	N/A
Layers 1+2+3	35.62	35.47	35.76

TABLE X
PSNR COMPARISON BETWEEN HSMDC AND SC FOR THE "CONTAINER" SEQUENCE AT 64-96-128 kb/s

	$\alpha = 1.0$	$\alpha = 1.4$	SC
Layer 1	36.26	36.26	36.37
Layers 1+2	36.57	36.68	36.70
Layers 1+3	36.58	36.61	N/A
Layers 1+2+3	36.91	36.83	37.01

scription enhancement layers need not be as correlated as in traditional MDC. An increase in α translates in an increase of the PSNR of decoding layers $1 + 2$ and layers $1 + 3$ (since the redundancy is increased) and in a decrease of the PSNR of decoding layers $1 + 2 + 3$. It should be emphasized that the external rate control keeps the total rate for all layers the same for all choices of α. Results for $\alpha = 1.0$ are provided to point out that, in contrast with MDC, the proposed HSMDC algorithm even works with no added redundancy, at the expense, of course, of a clearly reduced PSNR for layers $1 + 2$ and $1 + 3$. As can be seen from the results for $\alpha = 1.4$, the PSNR of layers $1 + 2$ and $1 + 3$ is comparable to and sometimes better than that of the corresponding SC case. The tradeoff is a lower SNR when layers $1+2+3$ are decoded. The appropriate value of α for given channel conditions needs to be considered in a

TABLE XI
PSNR COMPARISON BETWEEN HSMDC AND SC FOR THE "CONTAINER" SEQUENCE AT 96-144-192 kb/s

	$\alpha = 1.0$	$\alpha = 1.4$	SC
Layer 1	37.34	37.39	37.38
Layers 1+2	37.76	37.91	37.83
Layers 1+3	37.78	37.83	N/A
Layers 1+2+3	38.24	38.11	38.31

TABLE XII
PSNR COMPARISON BETWEEN HSMDC AND SC FOR THE "CONTAINER" SEQUENCE AT 128-192-256 kb/s

	$\alpha = 1.0$	$\alpha = 1.4$	SC
Layer 1	38.41	38.43	38.39
Layers 1+2	38.77	38.89	38.88
Layers 1+3	38.78	38.81	N/A
Layers 1+2+3	39.17	39.06	39.18

joint source-channel coding framework and should depend on the probabilities of losing each layer for given channel conditions.

V. CONCLUSION

We have presented a hybrid scalable multiple description video codec. We have shown that the functionality of having multiple description enhancement layers comes at the expense of a slight decrease of the PSNR when using all layers in the decoding. We are currently working on the application of the proposed codec on wireless video transmission in a joint source-channel coding framework.

REFERENCES

[1] *Video Coding for Low Bitrate Communications*, Jan. 1998.

[2] *Generic Coding of Moving Pictures and Associated Audio*, ISO/IEC IS-13818, Nov. 1994.

[3] R. Koenen, "MPEG-4: multimedia for our time," *IEEE Spectrum*, vol. 36, no. 2, pp. 26–33, Feb. 1999.

[4] Y. Andreopoulos, A. Munteanu, G. Van der Auwera, P. Schelkens, and J. Cornelis, "Wavelet-based fully-scalable video coding with in-band prediction," *Proc. Benelux Signal Processing Symp.*, pp. 217–220, 2002.

[5] K. Rose and S. L. Regunathan, "Toward optimality in scalable predictive coding," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 965–976, Jul. 2001.

[6] L. P. Kondi and A. K. Katsaggelos, "An operational rate-distortion optimal single-pass SNR scalable video coder," *IEEE Trans. Image Process.*, vol. 10, no. 11, pp. 1613–1620, Nov. 2001.

[7] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 301–317, Mar. 2001.

[8] J. K. Wolf, A. Wyner, and J. Ziv, "Source coding for multiple descriptions," *Bell Syst. Tech. J.*, vol. 59, pp. 1417–1426, Oct. 1980.

[9] A. R. Reibman, H. Jafarkhani, Y. Wang, and M. Orchard, "Multiple description video using rate-distortion splitting," in *Proc. Int. Conf. Image Process.*, vol. 1, Thessaloniki, Greece, Oct. 2001, pp. 978–981.

[10] X. Yang and K. Ramchandran, "Optimal multiple description subband coding," in *Proc. Int. Conf. Image Process.*, vol. 1, Chicago, IL, Oct. 1998, pp. 654–658.

[11] V. K. Goyal and J. Kovacevic, "Generalized multiple description coding with correlating transforms," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2199–2224, Sep. 2001.

[12] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 6, pp. 851–857, Nov. 1982.

[13] Y. Wang, M. Orchard, V. Vaishampayan, and A. R. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Trans. Image Proc.*, vol. 10, pp. 351–366, Mar. 2001.

[14] V. K. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 74–93, Sep. 2001.

[15] L. Ozarow, "On a source coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, pp. 1909–1921, Dec. 1980.

[16] V. K. Goyal, J. Kovacevic, R. Arean, and M. Vetterli, "Multiple description transform coding of images," in *Proc. Int. Conf. Image Processing*, vol. 1, Chicago, IL, Oct. 1998, pp. 674–678.

[17] A. R. Reibman, H. Jafarkhani, Y. Wang, M. Orchard, and R. Puri, "Multiple description coding for video using motion compensated prediction," in *Proc. Int. Conf. Image Processing*, vol. 3, Kobe, Japan, Oct. 1999, pp. 837–841.

[18] A. E. Mohr, E. A. Riskin, and R. E. Ladner, "Unequal loss protection: graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE J. Sel. Areas Commun.*, vol. 18, pp. 819–828, Jun. 2000.

[19] R. Puri, K. Ramchandran, K. W. Lee, and V. Bharghavan, "Forward error correction (FEC) codes based multiple description coding for internet video streaming and multicast," *Signal Process. Image Commun.*, vol. 16, no. 8, pp. 745–762, May 2001.

[20] H. Wang and A. Ortega, "Robust video communication by combining scalability and multiple description coding techniques," in *Proc. Conf. Image and Video Communicatiosn and Processing*, Santa Clara, CA, Jan. 2003, pp. 111–124.

[21] P. A. Chou, H. J. Wang, and V. N. Padmanabhan, "Layered multiple description coding," presented at the Packet Video Workshop, Nantes, France, Apr. 2003.

[22] University of British Columbia, "TMN Version 3.2,", H.263+ Public domain implementation.

[23] R. Aravind, M. R. Civanlar, and A. R. Reibman, "Packet loss resilience of MPEG-2 scalable video coding algorithms," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 5, pp. 426–435, Oct. 1996.

[24] L. P. Kondi, "A rate-distortion optimal hybrid scalable/multiple-description video codec," in *Proc. Int. Conf. Acoustics, Speech and Signal Process.*, Montreal, QC, Canada, May 2004, pp. 269–272.

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

[26] G. M. Schuster and A. K. Katsaggelos, *Rate-Distortion Based Video Compression, Optimal Video Frame Compression, and Object Boundary Encoding*. Norwell, MA: Kluwer, 1997.

[27] K. Ramchandran and M. Vetterli, "Rate-distortion optimal fast thresholding with complete JPEG/MPEG decoder compatibility," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 700–704, Sep. 1994.

[28] M. Crouse and K. Ramchandran, "Joint thresholding and quantizer selection for transform image coding: entropy-constrained analysis and applications to baseline JPEG," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 285–297, Feb. 1997.

[29] L. P. Kondi and A. K. Katsaggelos, "An optimal single pass SNR scalable video coder," in *Proc. Int. Conf. Image Processing*, Kobe, Japan, 1999, pp. 276–280.