H.264/AVC-based multiple description video coding using dynamic slice groups[☆]

Che-Chun Su^b, Homer H. Chen^{a,b,c,*}, Jason J. Yao^{a,b}, Polly Huang^{a,b,c}

^a Department of Electrical Engineering, Taipei 10617, Taiwan, ROC

^b Graduate Institute of Communication Engineering, National Taiwan University, Taipei 10617, Taiwan, ROC

^c Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei 10617, Taiwan, ROC

ARTICLE INFO

Article history:

Received 24 March 2008

Received in revised form

15 June 2008

Accepted 28 July 2008

Keywords:

H.264/AVC

Multiple description coding

Slice group

ABSTRACT

In this paper, an H.264/AVC-based multiple description video coding scheme is proposed. It utilizes the advanced video coding tools and features provided in H.264/AVC to introduce redundancy into descriptions. Two independently decodable descriptions are generated, each consisting of two slice groups. One of them, called main slice group (MSG), is encoded normally as main information. The other one, called side slice group (SSG), is encoded with fewer bits as redundancy information by using larger quantization step sizes. Spatial and temporal correlations between neighboring macroblocks in video frames are exploited to achieve efficient redundancy coding. Experimental results show that the proposed MDC scheme is superior to previous slice group based multiple description coding (MDC) schemes in terms of the rate-distortion (R-D) performance.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Although more and more multimedia applications such as IPTV and peer-to-peer content distribution have emerged as a result of the rapid growth of the Internet and wireless networks, robust video transmission [1,2] remains a challenging issue as the bandwidth is never enough to feed the increasing multimedia traffic due to, for example, the demand for high-quality video.

Multiple description coding (MDC) is an effective means developed to deal with data transmission over error-prone networks [3]. It encodes one signal into multiple bit-streams. Each bit-stream is regarded as one

description, and each description is independently decodable. If one description is received, a baseline signal can be reconstructed. With more descriptions received, the quality of the reconstructed signal is improved. Through this mechanism, MDC reduces the adverse effect of packet losses by transmitting different descriptions along different paths. In addition, a variety of error-concealment techniques can be developed to recover the lost information. The benefits of MDC come at the cost of added redundancy into descriptions. Therefore, one major objective of designing MDC schemes is to minimize the redundancy, while meeting the end-to-end rate-distortion (R-D) requirement in an error-prone network.

To apply MDC to video transmission, the principles of video coding algorithms need to be considered. Motion-compensated temporal prediction is nearly universal in today's successful hybrid video coding systems. If there is a mismatch between the motion-compensated states of the encoder and decoder, the error will accumulate and propagate until the next non-predicted frame. Thus, in designing a multiple description video coder, a key challenge is how to deal with the mismatch between the

[☆] Manuscript received March 24, 2008. This work was supported in part by grants from the Intel Corporation, ITRI, III, and the National Science Council of Taiwan under Contract 95E1053, NSC 94-2219-E-002-012, NSC 94-2219-E-002-016, NSC 94-2752-E-002-006-PAE.

* Corresponding author at: Department of Electrical Engineering, 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan, ROC.
Tel.: +886 2 33663549; fax: +886 2 23683824.

E-mail address: homer@cc.ee.ntu.edu.tw (H.H. Chen).

reference frame buffers in the encoder and decoder when only one description is received at the decoder. One way to avoid such a mismatch is to have independent prediction loops, each consisting of reference frames reconstructed from a single description. Otherwise, the mismatch signal is coded as the redundancy into the descriptions. The performance of a multiple description video coder depends greatly on how effective the mechanisms are to reduce the reference frame mismatch between the encoder and decoder.

Many multiple description video coding schemes have been proposed recently, and all are built on top of the block-based motion-compensated prediction framework. In [4], a multiple description transform coding (MDTC) video coder is presented. The prediction error of the central motion-compensated loop is coded by the pairwise correlating transform (PCT) [5,6] to produce two descriptions. The mismatch between the motion-compensated prediction in the central and side encoder is coded as redundancy, which is controlled by the PCT parameter and the quantization step size. In [11], a poly-phase down-sampling (PD) [10] technique is used for generation of descriptions. By down-sampling the input signal before the temporal prediction loop, a flexible number of descriptions can be generated. In [7], a multiple description motion compensation (MDMC) video coder is proposed. It performs motion compensation by predicting the current frame from two previously coded frames. Two descriptions are generated, containing the even and odd coded frames, respectively. When only one description is received, e.g. the one containing even frames, the decoder has prediction only from the reconstructed even frames. The mismatch signal is coded explicitly to avoid error propagation, and the total redundancy is controlled by both the predictor coefficients and the quantization step size. In [12], the multiple description motion coding (MDMC) algorithm is proposed to enhance the robustness of the motion vector field against transmission errors. First, the motion vectors are estimated by minimizing a Lagrangian cost function that takes into account the possible scenarios of received descriptions at the decoder. Then, the motion vectors and the motion-compensated prediction error are split into two descriptions following a quincunx sub-sampling lattices scheme. However, all the schemes described above are not designed for any specific video coding standard, and they are not easy to be implemented for practical applications. To solve this problem, a multiple-state MDC scheme through pre-processing is proposed in [8,9]. The input video sequence is first divided into two subsequences of frames, even and odd. Each subsequence is independently encoded as one description, and different error-concealment methods can be used to recover the lost frames. One drawback of this multiple-state scheme is that the video quality drops due to the limited reference frames in each state.

In this paper, an H.264/AVC-based MDC scheme is proposed. It adopts H.264/AVC as the base for video codec and utilizes its advanced video coding tools, including slice groups, variable block-size motion compensation, and multiple reference frames [13–16], to counteract packet losses and enhance error-concealment. One of the

design goals is to use the tools provided in the standard as much as possible, because we want it to be standard compliant. Slice groups are used to generate independently decodable descriptions [17–19]. The proposed MDC scheme aims at introducing redundancy into descriptions in an efficient way and providing error-concealment for reliable video transmission.

The rest of this paper is organized as follows. Section 2 describes the framework of the proposed MDC scheme and the details of the redundancy coding algorithms. Section 3 shows the experimental results, followed by a conclusion in Section 4.

2. Proposed MDC scheme

The slice group is a new coding tool provided in H.264/AVC, in which a coded frame consists of one or more slice groups, and each slice group contains one or more slices. In H.264/AVC, there are seven types of macroblock (MB) to slice group maps that define which slice group an MB belongs to. Type 1, called dispersed MB to slice group map, is very effective for error resilience [20], and is adopted in the proposed MDC scheme. Fig. 1 shows the dispersed slice group map with two slice groups, SGA and SGB, each containing one independently decodable slice.

2.1. Framework

Fig. 2 shows the framework of the proposed MDC scheme, which employs the dispersed slice group map to produce two independently decodable descriptions. In each description, a coded frame consists of two slice groups, SGA and SGB, arranged according to the dispersed MB to slice group map, as shown in Fig. 1. One of the two slice groups is encoded normally, called main slice group (MSG). The other slice group, called side slice group (SSG), is encoded with fewer bits than the MSG by using larger quantization step sizes. The MSG is encoded prior to the SSG, and the redundancy is introduced into the SSG. For each description, the input video sequence is first processed by the slice group interchanger, which decides whether a slice group is encoded as MSG or SSG. Next, the MSG is encoded normally, including intra- and/or inter-prediction with R-D optimized mode decision. Finally, the

Fig. 1. The dispersed macroblock to slice group map.

Fig. 2. Framework of the proposed MDC scheme.

SSG is encoded with the aid of the motion information from the MSG. Since the two descriptions are symmetric, only the design of description-1 encoder is discussed in the following sections.

2.2. Dynamic slice group interchanger

In the proposed MDC scheme, the encoding patterns of SGA and SGB can be interchanged frame by frame. For example, if the SGA in the previous frame is encoded as MSG, and the encoding pattern is interchanged, the SGA will be encoded as SSG in the current frame. Thus, for every SSG MB, the corresponding MB at the same position in the previous frame is encoded as MSG, and vice versa. In addition, the neighboring macroblocks (MBs) of an SSG MB are encoded as MSG. This temporal and spatial relationship is illustrated in Fig. 3. Because the MSG MBs are encoded normally with motion-compensated prediction and R-D optimized mode decision, the motion information can be used to help encode the SSG MBs and introduce the redundancy. Moreover, since the quantization step size of MSG is smaller than that of SSG, the MSG MBs have better quality and can lead to more accurate predictions for SSG MBs, resulting in small residuals. However, the coding efficiency of MSG MBs may drop due to the coarse SSG MBs with larger quantization step sizes. To solve this problem, a dynamic slice group interchanger is proposed to conditionally interchange the slice group map of the current frame with that of the previous frame, and the condition is described as follows:

$$\text{Motion_Key} = \# \text{of} \left\{ \begin{array}{l} \text{MV} | \text{MV} \in \text{motion vector in the previous frame \& } \\ (|MV_x| \leq 1 \& |MV_y| \leq 1) \end{array} \right\} \quad (1)$$

$$\text{Bit_Key} = \frac{\text{total bits of SSG in the previous frame}}{\text{total bits of MSG in the previous frame}} \quad (2)$$

$$\text{Exception_Flag} = \left\{ \begin{array}{l} (\text{Motion_Key} \geq \text{Motion_Thresh}) \& \\ (\text{Bit_Key} \leq \text{Bit_Thresh}) \end{array} \right\} ? 1 : 0 \quad (3)$$

Fig. 3. The spatial and temporal relationship between MSG and SSG macroblocks.

Motion_key is the total number of motion vectors in the previous frame, in which the magnitudes of the x-component and the y-component are both smaller than or equal to one. Bit_key is the bit ratio of SSG to MSG in the previous frame. Motion_Thresh and Bit_Thresh are the thresholds for Motion_key and Bit_key, respectively. Finally, the value of Exception_Flag determines if the slice group map of the current frame should be interchanged with that of the previous frame: one means no interchange, and zero means there is interchange. From (3), the slice group map is not interchanged if the following two conditions are both satisfied: (Motion_Key ≥ Motion_Thresh) and (Bit_Key ≤ Bit_Thresh). The first condition means that the motion of the previous frame is low or the scene is still. It implies that the current frame probably has low motion, and the motion vectors are small. The second condition means that the quality of SSG MBs is

Fig. 4. The result of the dynamic slice group interchanger.

much worse than that of the MSG MBs. If both conditions are satisfied and the slice group map is interchanged, the MSG MBs will have the coarse SSG MBs as prediction, resulting in large residual and diverse motion vectors. This significantly decreases the coding efficiency of MSG MBs. Thus, the interchange is turned off to raise the coding efficiency when both conditions are satisfied. Fig. 4 demonstrates the result of the dynamic slice group interchanger.

2.3. SSG encoder

The encoding of SSG MBs comprises three steps. The first step performs the inter prediction, not by doing motion estimation, but by predicting the motion vector from the neighboring MSG MBs and the corresponding MSG MB in the previous frame. Then, the reference frame is determined by the histogram collected from the reference frames of neighboring MSG MBs. If the SSG is in an intra-frame, only the normal intra-prediction is performed. The final step is to determine the best mode according to the R-D cost.

2.3.1. Spatial-temporal prediction of motion vector (STPMV)

In the first step, a STPMV technique is adopted. In SSG, each MB is divided into sixteen 4×4 blocks, and the motion vector of each 4×4 block is predicted from the spatial 4×4 blocks of the neighboring MSG MBs, and/or the temporal 4×4 block at the same position of the MSG MB in the previous frame. For simplicity, we use S- 4×4 block and T- 4×4 block to denote the spatial and temporal 4×4 MSG blocks, respectively. If the slice group map of the current frame is interchanged from one of the previous frame, the T- 4×4 block is used in the prediction; otherwise only the S- 4×4 blocks are used. In addition, motion-compensated prediction can achieve small distortion if blocks with small sizes are used in motion estimation. Thus, the block size 4×4 , which is the smallest MB partition in H.264/AVC, is chosen to encode the SSG MBs.

For STPMV, three different types of SSG MBs are defined according to their positions in the coded frame: the corner MB, the edge MB, and the central MB, as

illustrated in Fig. 5. For each SSG MB type, a different STPMV method is applied.

Corner and edge MBs, as shown in Fig. 5(a) and (b), have two different kinds of corner 4×4 blocks. One of them has two neighboring MSG MBs, while the other kind of corner 4×4 block has only one. The motion vector of the corner 4×4 block that has only one neighboring MSG MB is predicted from two S- 4×4 blocks and one T- 4×4 block. For the corner 4×4 block that has two neighboring MSG MBs, its motion vector is predicted from four S- 4×4 blocks and one T- 4×4 block. The motion vector of the 4×4 block along the frame edge is predicted from three S- 4×4 blocks and one T- 4×4 block. Finally, the motion vectors of the remaining 4×4 blocks are predicted only from their T- 4×4 blocks in the previous frame.

For central MBs shown in Fig. 5(c), the motion vectors of the boundary 4×4 blocks are predicted in the same way as the corner and edge MBs. The motion vectors of the four interior 4×4 blocks are predicted from five neighboring 4×4 blocks in the same SSG MB and one T- 4×4 block.

Finally, for all types of MBs mentioned above, the predicted motion vector is set to the median of their candidates.

2.3.2. Reference frame selection

After predicting the motion vector of SSG MBs, the reference frame needs to be decided. H.264/AVC supports multiple reference frames, and the maximum number of reference frames depends on the profile and level. Based on the same idea in STPMV, each SSG MB is divided into sixteen 4×4 blocks to achieve efficient motion-compensated prediction. However, H.264/AVC requires that the four 4×4 blocks in an 8×8 block use the same reference frame. Thus, in the SSG MB, only one reference frame is determined for each 8×8 block that consists of four 4×4 blocks.

In the proposed method of reference frame selection, the reference frame of each 8×8 SSG block is selected from the reference frames of neighboring 8×8 MSG blocks. As in STPMV, three different types of MBs are defined (Fig. 6). A different selection method is applied to the 8×8 block at different positions in the SSG MB. The selection comprises two steps. First, the candidates of the

Fig. 5. For STPMV, three types of macroblocks are defined in SSG: (a) corner MB, (b) edge MB, and (c) central MB.

reference frame are chosen from the reference frames of the neighboring 8×8 MSG blocks. Then, the reference frame is determined by the histogram of all candidates.

For corner MBs shown in Fig. 6(a), the reference frames of the two 8×8 blocks that have only one neighboring MSG MB are selected from the reference frames of three 8×8 MSG blocks. The reference frame of the 8×8 block which have two neighboring MSG MBs is selected from the reference frames of six 8×8 MSG blocks. For the 8×8 block, which has no neighboring MSG MB, its reference frame is directly set to the previous reconstructed frame. For edge and central MBs shown in Fig. 6(b) and (c), the reference frames of all 8×8 blocks are selected in the same way as the corner MBs described above.

After its candidates are chosen, the reference frame is determined as follows:

$$\text{Val} = \max[\text{Hist}(i)] \quad (4)$$

$$\text{Key} = \underset{i}{\text{Arg}}\{\max[\text{Hist}(i)]\} \quad (5)$$

$i \in \text{possible reference frames}$

$$\text{ref} = (\text{Val} \geq \text{Thresh})? \text{Key} : 0 \quad (6)$$

Hist(i) is the histogram computed from the reference frame candidates. For example, if there are six candidates, four of them are 1 and two of them are 3. Then, Hist(1) = 4, Hist(3) = 2, and Hist(i) = 0, for $i \in \text{other}$

Fig. 6. For reference frame selection, three types of macroblocks are defined in SSG: (a) corner MB, (b) edge MB, and (c) central MB. Candidates are chosen for 8×8 blocks at different positions.

possible reference frames. The threshold, *Thresh*, is set to one half of the total number of reference frame candidates. If the maximum value of the histogram is larger than or equal to the threshold, it means the current 8×8 block in SSG tends to have the same reference frame as its neighboring 8×8 blocks. On the contrary, if the maximum histogram value is smaller than the threshold, it means the neighboring 8×8 blocks in MSG cannot provide useful information about the reference frame of the current 8×8 block in SSG. Thus, the reference frame

of the current 8×8 block is set to the previous reconstructed frame.

Simulations are performed to examine the effectiveness of the proposed method of reference frame selection. Four CIF sequences are tested: *Foreman*, *Mobile Calendar*, *Stefan*, and *Table Tennis*. The platform is the reference software JM10.1 of H.264/AVC [21]. The dispersed slice group with two slice groups is adopted. Each slice group has one slice. Both slice groups, SGA and SGB, undergo the normal encoding process. The reference frame of each

Table 1
Simulation results of reference frame selection

Sequence	Frame number	Average error per 8×8 block				Average error per macroblock	Match percentage (%)
		1	2	3	4		
Foreman	40	0.912	0.993	0.975	0.962	0.961	55.6
	59	1.737	1.512	1.787	1.643	1.673	41.5
Mobile Calendar	40	1.643	1.825	1.681	1.850	1.752	47.5
	59	1.462	1.537	1.506	1.656	1.541	48.6
Stefan	40	1.693	1.443	1.706	1.543	1.596	47.3
	59	0.331	0.350	0.381	0.394	0.364	77.5
Table Tennis	40	0.287	0.337	0.325	0.381	0.332	78.9
	59	0.325	0.262	0.318	0.293	0.301	79.1

8×8 block in each slice group is recorded. The proposed method is performed on each 8×8 block in SGB, and the selected reference frames are compared with the reference frames determined by the normal encoding process. Table 1 shows the simulation results.

In the simulation, the GOP size is set to 30, and the number of reference frames is set to 10. Table 1 shows the results of frames 40 and 59, where 10 reference frames can be used to perform the motion-compensated prediction. In addition, other frames have the same results as the frames 40 and 59. The average error is the difference between the reference frame selected by the proposed method and the one determined by the normal encoding process. The match percentage, which represents the percentage that the selected reference frame is identical to the one selected by the normal encoding process, indicates how accurate the proposed selection method is. The 8×8 block numbers stand for the raster-scanned order of the 8×8 blocks in the MB. For *Mobile Calendar*, the background is very complicated, and the scene moves with a horizontal velocity. This makes the reference frames of neighboring 8×8 blocks uncorrelated. Although the average error is larger than 1, the match percentage approaches 50%, which means the proposed method can select the accurate reference frame for half the 8×8 blocks in SSG. For *Stefan*, there is an irregular camera motion in the horizontal direction, and the upper background is quite complicated. This leads to different simulation results for different frames. For *Table Tennis*, the average error is quite low and the match percentage reaches 80%. In summary, the match percentages in the simulation results show that the proposed method of reference frame selection works well for different sequences.

2.3.3. Improved mode decision

In the third step of the SSG MB encoding, the best mode is determined. By using STPMV with reference frame selection to encode the SSG MB, the bits of the header and motion data can be saved, because of the fixed 4×4 MB partition with the predicted motion vector and

Fig. 7. The improved mode decision flow.

the selected reference frame. However, the motion vector predicted by STPMV and the selected reference frame may produce large residual, resulting in non-optimal R-D cost. In order to obtain the best coding efficiency, an improved mode decision flow is proposed, as shown in Fig. 7. The best mode is chosen among the STPMV mode and all the modes provided in H.264/AVC.

First, the normal R-D optimized mode decision is performed. This normal mode is determined by the minimum R-D cost computed according to the Lagrangian cost function. Then, the SSG MB is inter-coded by the predicted motion vector in STPMV and the selected reference frame, resulting in the R-D cost of the STPMV mode. This R-D cost is computed from the distortion and the bits needed for coding the quantized transform coefficients. Finally, the R-D cost of the STPMV mode is compared with the R-D cost of the normal mode.

The mode with lower R-D cost is chosen as the best mode to encode the SSG MB.

2.4. Complete block diagram

Fig. 8 shows the complete block diagram of the proposed H.264/AVC-based MDC scheme, which has two encoding loops, one for each description. Since the two descriptions are symmetric, only the structure of the description-1 encoder is detailed, and the same operations are applied to the description-2 encoder.

First, the encoding pattern of MSG and SSG is determined in the dynamic slice group interchanger. The

resulting slice group map is fed into the slice group interchanger in the description-2 encoder to determine the MSG and SSG symmetrically. Then, the MSG is encoded normally, and the motion data (Motion Data) are fed into the SSG encoder. Taking advantage of this motion data, the SSG is inter-coded by STPMV with reference frame selection. The normal encoding process is also performed for the SSG. Finally, the improved mode decision determines the best mode according to the R-D cost. The output description contains the residual (Residual), header (Header), and motion data (Motion Data) of both MSG and SSG. In addition, because the SSG can be encoded by STPMV with reference frame selection or the mode defined in H.264/AVC, a macroblock coding map

Fig. 8. The complete block diagram of the proposed MDC scheme.

(MBC-map) needs to be encoded. The MBC-map uses one bit for each SSG MB to indicate if it is to be encoded by STPMV with reference frame selection or not. Thus, for a coded frame, the total number of bits of the MBC-map is one half of the number of MBs. Generally, this overhead is smaller than 1% of the bit-rate consumed. We define a new type of NAL unit packet through the `nal_unit_type` parameter, which is a 5-bit number, to indicate the MBC-map in the H.264/AVC stream. Specifically, we use

Table 2

Test conditions

Platform	JM 10.1
Sequence length	150
Frame rate	30
Format	CIF 4:2:0
Motion vector resolution	1/4-pel
Motion estimation search range	± 16
Number of reference frames	10
Rate-distortion optimization	On
GOP structure	IPPP...
B frame	No

the non-specified value 24 of `nal_unit_type` for the new type.

In the proposed MDC scheme, there are two (MSG and SSG) encoders. While maintaining the common reconstructed frame buffer, they possess independent transform-quantization processes. Each of them has their own quantization parameter (QP), QP of MSG (QPM), and QP of SSG (QPS). Thus, the QPM controls the quality of the reconstructed video when both descriptions are received successfully, and the amount of redundancy is adjusted by the QPS to control the quality if only one description is received.

2.5. Error-resilient mechanism

There are different kinds of errors for video transmission over error-prone packet-networks: single packet lost, burst packet loss, and channel failure. To combat these network errors, the proposed scheme provides an error-resilient mechanism. Video sequences are encoded into two independently decodable descriptions, which are transmitted along two different paths in the network.

Fig. 9. The R-D performance of *Stefan* sequence: (a) single-channel reconstruction and (b) complete reconstruction.

If one of the two transmission paths fails, the decoder is able to maintain an acceptable reconstructed quality by decoding the description that is successfully received. If packets are lost, the error-concealment tools provided in H.264/AVC [22,23] are adopted to reconstruct the lost MBs. Details are described as follows.

If one description is lost, and the other description is received without error, the decoder can reconstruct the video with acceptable quality by decoding the received description.

If one description is lost, and some packets are lost in the other description, the lost MBs are recovered by using the successfully reconstructed MBs. If the lost packets belong to an SSG MB, its motion vector is predicted from the neighboring MSG MBs. The lost reference frame index is recovered by using the selection method presented in Section 2.3.2. Then, the lost SSG MB is reconstructed. However, if the lost packets belong to an MSG MB, no neighboring MBs can help recover the lost MB. The reconstructed MB at the same position in the previous frame is copied.

If both descriptions are received, but some packets are lost, there are two different types of MB loss. If the lost

MBs in one description are reconstructed successfully in the other description, the reconstructed MBs are directly copied to compensate the drift error. If the same MBs are lost in both descriptions, the error-concealment techniques described in [22,23] can be applied.

3. Experimental results

3.1. Scenario

In this section, the performance of the proposed MDC scheme is presented. Two different kinds of experiments are performed to evaluate its coding efficiency: the R-D performance of the complete reconstruction and the R-D performance of the single-channel reconstruction. In both scenarios, two independently decodable descriptions are generated. In the experiment of the complete reconstruction, the MDC scheme is assumed error free. Both descriptions are received successfully, and the decoder reconstructs the signal with the best quality. In the experiment of the single-channel reconstruction, it is assumed that only one of the two descriptions is

Fig. 10. The R-D performance of Mobile Calendar sequence: (a) single-channel reconstruction and (b) complete reconstruction.

successfully received and the other one is entirely lost during transmission. If only one description is received successfully, the multiple description decoder checks which description is available, and the corresponding decoder reconstructs the signal with an acceptable quality. The goal of the simulation of single-channel reconstruction is to examine the efficiency of redundancy coding. If an MDC scheme encodes the redundancy more efficiently, it can use the bits that are saved to encode the baseline signal under the same bandwidth constraint. A quality baseline is important to MDC, because the baseline data will be used to conceal errors when random packet loss occurs. Therefore, more efficient redundancy coding means better ability of error recovery.

For the complete reconstruction, the proposed MDC scheme decodes the two successfully received descriptions in the description-1 and -2 decoders, respectively. Each frame of the reconstructed sequence comprises a MSG and a SSG. All the MSGs are extracted from both reconstructed sequences, and they are rearranged to produce the best-quality output video.

In these two experiments, the proposed MDC scheme is compared with the previous two slice group based MDC

(SG-MDC) schemes [17,18]. Both of them were implemented, and the experiments of single-channel reconstruction were performed. The three-loop scheme [18] has better R-D performance than the one with two independent description encoders [17]. Thus, the SG-MDC scheme [18] is compared against the proposed MDC scheme. In [18], the slice group pattern is fixed, and the SSG MBs are coded by using spatial prediction of motion vectors without reference frame selection.

The proposed MDC scheme is implemented on JM 10.1 [21]. The R-D optimization (RDO) is turned on. The GOP structure is IPPP... without B-frame, and the number of reference frames is set to 10. The test sequences are of CIF size with frame rate 30. The test conditions are summarized in Table 2.

3.2. R-D performance

In both experiments, different bit-rate values on the R-D curves are obtained by changing the QP of SSG (QPS), while keeping the QP of MSG (QPM) at a fixed value. Since the MSGs are used to reconstruct the video when two

Fig. 11. The R-D performance of Bus sequence: (a) single-channel reconstruction and (b) complete reconstruction.

descriptions are received successfully, keeping the QPM at a fixed value maintains the quality of the complete reconstruction. In addition, because the SSG is coded as redundancy in each description, the bit-rate, which depends on the amount of redundancy, can be controlled by changing the QPS. For the R-D performance of the

complete reconstruction, we compute the PSNR of the reconstructed sequence that consists of the MSGs from both descriptions, and the bit-rate is the total bits in both descriptions. For the single-channel reconstruction, each PSNR value on the R-D curve is obtained by computing the difference between the original video sequence and the

Fig. 12. The redundancy vs. bit-rate curves of three sequences: (a) *Stefan*, (b) *Mobile Calendar*, and (c) *Bus*.

decoded video sequence corresponding to the successfully received description, and the bit-rate is also obtained from the successfully received description.

Figs. 9–11 show the experimental results for *Stefan*, *Mobile Calendar*, and *Bus*, respectively.

The R-D performance of the single-channel reconstruction shows that the proposed MDC scheme achieves higher PSNR over almost the entire bit-rate range. For *Stefan* and *Bus*, 3 dB PSNR gains are achieved. For *Mobile Calendar*, the maximum improvement approaches 6 dB.

Fig. 13. The R-D performance of single-channel reconstruction varying the QPM: (a) *Stefan*, (b) *Mobile Calendar*, and (c) *Bus*.

As the bit-rate drops and the amount of redundancy decreases, the PSNR difference increases for all test sequences. The performance gain is attributed to the efficient redundancy coding of the proposed MDC scheme. Under the interchanged slice group pattern, the corresponding MB at the same position in the previous frame of each SSG MB is an MSG MB. This allows the proposed MDC scheme to add the motion vector of this MSG MB to the pool of STPMV and increase the accuracy of the predicted motion vector, resulting in smaller residuals and better PSNR. In addition, because the reference frame selection can give a correct reference frame from the neighboring MSG MBs, the reference frame indices of SSG MBs need not be coded, and the bits of motion data are saved. Finally, the improved mode decision achieves optimal SSG MB encoding by minimizing the R-D cost. All these novel designs of coding algorithms contribute to the superior R-D performance of the proposed MDC scheme.

The R-D performance of the complete reconstruction shows that the PSNR of the proposed scheme slightly drops as the total bit-rate decreases. Because the bit-rate decreases with larger QPSs, the quality of SSG also drops. The MSG, which is predicted from SSG, needs more bits to encode the residual, because of the relatively coarse quality of SSG, resulting in the PSNR drop of the complete reconstruction and the turning points in single-channel RD curves. The SG-MDC scheme [18] keeps the reconstructed PSNR at a fixed value due to its three-loop structure. However, the drop of the proposed scheme is smaller than 0.3 dB, and the difference between the two schemes is negligible. Thus, the proposed MDC scheme achieves superior single-channel performance, while providing comparable quality of the complete reconstruction.

It can be seen from Figs. 9 to 11(a) that the PSNR difference between H.264/AVC and the proposed scheme is negligible over most of the bit-rate range. However, it is inevitable that the performance of MDC suffers at low bit-rates due to the reduced redundancy.

Fig. 12 shows the relationship of the bit-rate and the redundancy in one description. Since the bit-rate is in proportion to the amount of redundancy, the proposed MDC scheme can control the total bit-rate under different channel bandwidths by adjusting the QPS. Fig. 12 also indicates that the redundancy corresponding to the bit-rates at which the turning points occur is approximately 10%, which is very small, for all three test sequences. Recall that the turning points shown in Fig. 11 all occur at lower bit-rates. Therefore, the effective bit-rate range of the proposed MDC scheme is broad enough for practical applications.

Fig. 13 shows the experimental results of the single-channel reconstruction obtained by varying the QPM. From the results, it can be seen that the proposed scheme has competitive performance with the single description of H.264/AVC over the entire bit-rate range. It can also be seen that by increasing the QP of MSG, the proposed scheme performs equally well at low bit-rates.

Note that the simulation results are run on a Pentium-4 PC with 1.25 GB RAM. The computational power

required for encoding one description by our system, which is implemented on JM10.1, is almost the same as for encoding the single description by the original JM10.1. The average encoding rate is about 0.16 (fps) for different test sequences.

4. Conclusion

A new H.264/AVC-based multiple description coding scheme has been presented in this paper. It adopts the advanced video coding tools and features provided in H.264/AVC. Slice groups are used to produce independently decodable descriptions. The correlations between neighboring MBs of different slice groups are exploited to introduce redundancy into descriptions. Experimental results show that the proposed MDC scheme is superior to previous SG-MDC schemes in terms of the R-D performance. More efficient redundancy coding is achieved. With the aid of well-designed error-concealment methods, the proposed MDC scheme provides a practical solution for video transmission over error-prone packet networks.

Acknowledgment

The authors thank Mr. Dong Wang for providing the software for generating the results presented in Section 3.

References

- [1] Y. Wang, Q.-F. Zhu, Error control and concealment for video communication: a review, *Proc. IEEE* 86 (5) (May 1998) 974–997.
- [2] Y. Wang, A.R. Reibman, S. Lin, Multiple description coding for video delivery, *Proc. IEEE* 93 (1) (January 2005) 57–70.
- [3] V.K. Goyal, Multiple description coding: compression meets the network, *IEEE Signal Process. Mag.* 18 (5) (September 2001) 74–93.
- [4] A. Reibman, H. Jafarkhani, Y. Wang, M. Orchard, R. Puri, Multiple-description video coding using motion-compensated temporal prediction, *IEEE Trans. Circuits Syst. Video Technol.* 12 (March 2002) 193–204.
- [5] M. Orchard, Y. Wang, V. Vaishampayan, A. Reibman, Redundancy rate-distortion analysis of multiple description coding using pairwise correlating transforms, in: *Proceedings of the IEEE International Conference Image Processing*, Santa Barbara, CA, October 1997, pp. 608–611.
- [6] Y. Wang, M. Orchard, V. Vaishampayan, A. Reibman, Multiple description coding using pairwise correlating transforms, *IEEE Trans. Image Process.* 10 (March 2001) 351–366.
- [7] Y. Wang, S. Lin, Error-resilient video coding using multiple description motion compensation, *IEEE Trans. Circuits Syst. Video Technol.* 12 (6) (January 2002) 438–452.
- [8] J. G. Apostolopoulos, Error-resilient video compression via multiple state streams, in: *Proceedings of the VLBV*, Kyoto, Japan, October 1999, pp. 168–171.
- [9] J. G. Apostolopoulos, Reliable video communication over lossy packet networks using multiple state encoding and path diversity, in: *Proceedings of the VCIP*, January 2001, pp. 392–409.
- [10] W. Jiang, A. Ortega, Multiple description coding via polyphase transform and selective quantization, in: *Proceedings of the VCIP*, vol. 3653, February 1999.
- [11] N. Franchi, M. Fumagalli, R. Lancini, S. Tubaro, Multiple description video coding for scalable and robust transmission over IP, *IEEE Trans. Circuits Syst. Video Technol.* 15 (3) (March 2005) 321–334.
- [12] C.-S. Kim, S.-U. Lee, Multiple description coding of motion fields for robust video transmission, *IEEE Trans. Circuits Syst. Video Technol.* 11 (9) (September 2001) 999–1010.
- [13] Draft ITU-T Recommendation and final draft international standard of joint video specification (ITU-T Rec.H.264|ISO/IEC 14496-10 AVC), JVT-G050r1, Geneva, May 2003.

- [14] T. Wiegand, G.J. Sullivan, G. Bjontegaard, A. Luthra, Overview of the H.264/AVC video coding standard, *IEEE Trans. Circuits Syst. Video Technol.* 13 (7) (July 2003) 560–576.
- [15] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, Rate-constrained coder control and comparison of video coding standards, *IEEE Trans. Circuits Syst. Video Technol.* 13 (July 2003) 688–703.
- [16] Y. Dhondt, P. Lambert, S. Notebaert, R. V. de Walle, Flexible macroblock ordering as a content adaptation tool in H.264/AVC, in: *Proceedings of the SPIE*, 2005, pp. 44–52.
- [17] D. Wang, N. Canagarajah, D. Bull, Slice group based multiple description video coding using motion vector estimation, in: *Proceedings of the ICIP*, Singapore, October 2004, pp. 3237–3240.
- [18] D. Wang, N. Canagarajah, D. Bull, Slice group based multiple description video coding with three motion compensation loops, in: *Proceedings of the IEEE International Symposium on Circuits and Systems*, Kobe, Japan, May 2005, pp. 960–963.
- [19] C.-C. Su, J. J. Yao, H. H. Chen, Multiple description video coding based on slice group interchange, in: *Picture Coding Symposium*, Beijing, China, April 2006.
- [20] W. Hantanong, S. Aramvith, Analysis of macroblock-to-slice group mapping for H.264 video transmission over packet-based wireless fading channel, in: *Proceedings of the IEEE Midwest Symposium on Circuits and Systems*, pp. 1541–1544, August 2005.
- [21] JM, JVT of ISO/IEC MPEG and ITU-T VCEG, Joint Model Reference Software Version 10.1.
- [22] S. Kumar, L. Xu, M.K. Mandal, S. Panchanathan, Error resiliency schemes in H.264/AVC standard, *Elsevier J. Visual Commun. Image Representation* 17 (April 2006) 425–450.
- [23] Y.-K. Wang, M. Hannuksela, V. Varsa, The error concealment feature in the H.26L test model, in: *Proceedings of the ICIP*, New York, September 2002, pp. II-729–II-732.