H.264/AVC-BASED MULTIPLE DESCRIPTION CODING SCHEME

Che-Chun Su, Jason J. Yao, and Homer H. Chen

Graduate Institute of Communication Engineering, National Taiwan University

ABSTRACT

A new multiple description coding (MDC) scheme is proposed in this paper. It utilizes the advanced video coding tools and features provided in H.264/AVC to introduce redundancy into descriptions. The proposed MDC scheme produces two descriptions, each consisting of two slice groups. One of them, called main slice group (MSG), is encoded in the normal way as main information. The other one, called side slice group (SSG), is encoded with fewer bits as redundancy 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 achieves better rate-distortion (R-D) performance than previous slice-group based MDC schemes.

***Index Terms*—** Multiple description coding (MDC), H.264/AVC, slice group

1. INTRODUCTION

With the rapid growth of the Internet and wireless networks, video transmission over lossy environments has become an important issue [1]. Multiple description coding (MDC) is an effective means to combat burst packet losses in error-prone networks [2]. MDC 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 can be increased. 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 applied to recover the lost information. The benefits of MDC come at the cost of added redundancy into descriptions. Therefore, one major objective of the MDC scheme design is to minimize the redundancy while meeting the end-to-end rate-distortion requirement in an error-prone network.

Extensive work on MDC has been conducted [3]-[7]. A transform-based MDC scheme is presented in [3], where the pairwise correlating transform (PCT) is used to transform a pair of Karhunen-Loeve coefficients into another pair of correlated components. This scheme is applied to DCT coefficients for motion-compensated video coding in [4]. A multiple-state MDC scheme through pre-processing is proposed in [5]. The video sequence is 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. However, 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, to introduce redundancy into the descriptions [6], [7]. 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. In the proposed scheme, the amount of redundancy can be controlled flexibly, and the reconstructed video quality is improved.

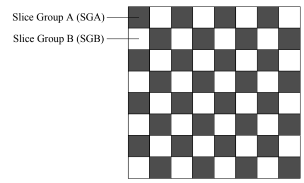


Fig. 1. The dispersed macroblock to slice group map.

The rest of this paper is organized as follows. In Section 2, the proposed MDC scheme is described. Section 3 gives the experimental results and the analysis, followed by the conclusions in Section 4.

2. PROPOSED MDC SCHEME

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. Since multiple slice groups, also known as flexible macroblock ordering (FMO), make it possible to encode macroblocks in flexible orders, effective error-concealment methods can be developed to recover the lost macroblocks. Type 1, called dispersed macroblock to slice group map, is very effective for error resilience [8], 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 macroblock 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 dynamic slice group interchanger, which determines a slice group is encoded as MSG or SSG. Next, the MSG is encoded normally, including intra and/or inter prediction with rate-distortion

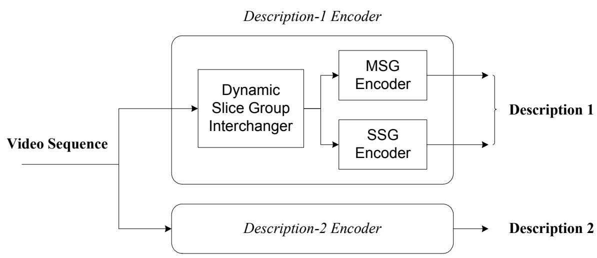


Fig. 2. Framework of the proposed MDC scheme.

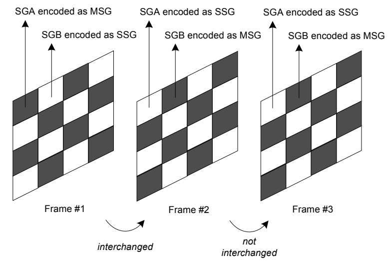
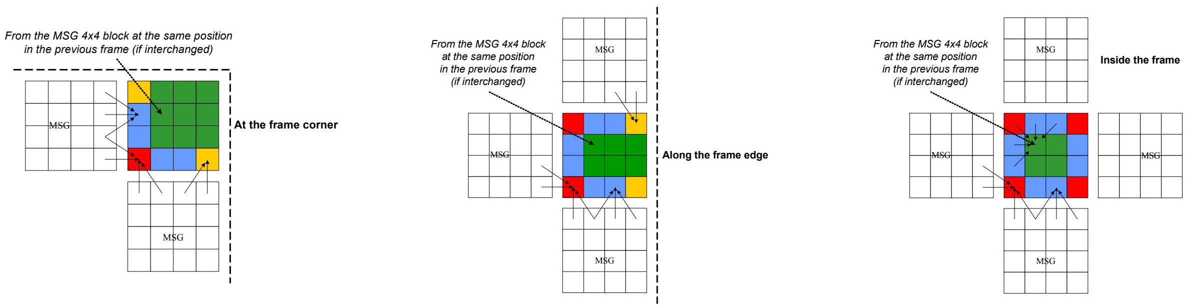


Fig. 3. The result of the dynamic slice group interchanger.



(a) (b) (c)

Fig. 4. For STPMV, three types of macroblocks are defined in SSG: (a) corner MB, (b) edge MB, and (c) central MB.

optimized mode decision. Finally, the 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 macroblock, the corresponding macroblock at the same position in the previous frame is encoded as MSG, and vice versa. Because the MSG macroblocks are encoded normally with motion-compensated prediction and rate-distortion optimized mode decision, the optimal motion information can be used to help encode the SSG macroblocks and introduce the redundancy. Moreover, since the quantization step size of MSG is smaller than that of SSG, the MSG macroblocks with better quality can form good prediction for SSG macroblocks, resulting in small residuals. However, the coding efficiency of MSG macroblocks may drop due to the coarse SSG macroblocks coded 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. Fig. 3 demonstrates the result of the dynamic slice group interchanger.

2.3. Side Slice Group Encoder

Because the SSG is coded as redundancy by using larger quantization step sizes, the bits of header and motion information can take up a large proportion of the total bits. Thus, the SSG is encoded with the aid of the motion information from the MSG to reduce the overhead. The encoding of SSG macroblocks comprises three steps. The first step performs the inter prediction to obtain the motion vector, not by doing motion estimation, but by predicting the motion vector from the neighboring MSG macroblocks and the corresponding MSG macroblock in the previous frame. Then, the reference frame is determined by the histogram collected from the reference frames of neighboring MSG macroblocks. 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 rate-distortion (R-D) cost.

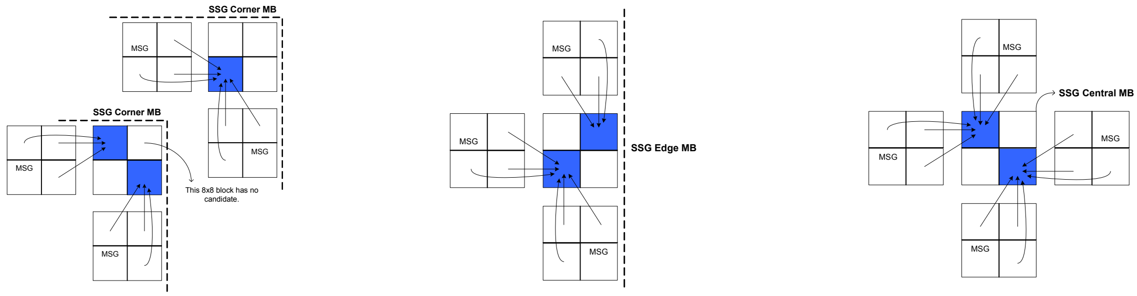
2.3.1. Spatial-Temporal Prediction of Motion Vector

In the first step, a spatial-temporal prediction of motion vector (STPMV) technique is adopted. In SSG, each macroblock is divided into sixteen 4x4 blocks, and the motion vector of each 4x4 block is predicted from the spatial 4x4 blocks of the neighboring macroblocks in MSG, and/or the temporal 4x4 block of the corresponding macroblock at the same position in the previous frame. In addition, motion-compensated prediction can achieve small distortion if blocks with small sizes are used in motion estimation. Thus, the block size 4x4, which is the smallest macroblock partition in H.264/AVC, is chosen to encode SSG.

For STPMV, three different types of macroblocks (MBs) are defined in SSG: the corner MB, the edge MB, and the central MB. Fig. 4 illustrates these three types of SSG macroblocks with their candidates for motion vector prediction. For all types of MBs, the predicted motion vector is set to the median of their candidates.

2.3.2. Multiple Reference Frames Selection

After the motion vector is obtained, the reference frame of the SSG macroblock needs to be decided. Based on the same idea in STPMV, each SSG macroblock is divided into sixteen 4x4 blocks. H.264/AVC requires that the four 4x4 blocks in an 8x8 block use the same reference frame. Thus, in the SSG macroblock, only one reference frame is determined for each 8x8 block that consists of four 4x4 blocks. The reference frame of each 8x8 SSG block is selected from the reference frames of neighboring 8x8 MSG blocks. As in STPMV, three different types of macroblocks (MBs) are defined, and their candidates for reference frame selection are chosen as shown in Fig. 5. The reference frame is determined by the histogram of all candidates as follows:



(a) (b) (c)

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

$$Val = \max[Hist(i)] \quad (1)$$

$$Key = \underset{i}{Arg}\left\{\max\left[Hist(i)\right]\right\} \quad (2)$$

$i \in$ possible reference frames

$$ref = \left(Val \geq Thresh\right) ? Key : 0 \quad (3)$$

$Hist(i)$ 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 $i \in$ other possible reference frames. The threshold, $Thresh$, is set to one half of the total number of reference frame candidates. If the maximum value is larger than or equal to the threshold, it means the current 8x8 SSG block tends to have the same reference frame as its neighboring 8x8 MSG blocks; otherwise, the neighboring 8x8 blocks can not provide useful information about the reference frame of the current SSG 8x8 block, and it is set to the previous reconstructed frame.

2.3.3. Improved Mode Decision

By using STPMV with reference frame selection to encode the SSG macroblock, the bits of the header and motion data can be saved. However, the motion vector determined by STPMV and the selected reference frame may produce large residual, resulting in non-optimal rate-distortion (R-D) cost. In order to obtain the best coding efficiency, an improved mode decision is proposed. The best mode is chosen among the proposed mode employing STPMV with reference frame selection and all the modes provided in H.264/AVC. First, the normal rate-distortion 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 macroblock is inter-coded by the predicted motion vector in STPMV and the selected reference frame. 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 proposed mode is compared with the one of the normal mode. The mode with the lower R-D cost is chosen as the best mode to encode the SSG macroblock.

2.3.4. 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. If one description is lost, and some packets are lost in the other description, the lost macroblocks are recovered by using the successfully reconstructed ones. If the lost packets belong to an SSG macroblock, its motion vector is predicted from neighboring MSG macroblocks. The lost reference frame index is recovered by using the selection method presented in Section 2.3.2. However, if the lost packets belong to an MSG macroblock, the reconstructed macroblock 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 macroblock loss. If the lost macroblocks in one description are reconstructed successfully in the other description, the reconstructed macroblocks are directly copied to compensate the drift error. If the same macroblocks are lost in both descriptions, the error-concealment techniques described in [9] can be applied.

3. EXPERIMENTAL RESULTS

This section presents the performance of the proposed MDC scheme. Two different kinds of experiments are performed to evaluate its coding efficiency: the rate-distortion (R-D) performance of the complete reconstruction and the R-D performance of the single-channel reconstruction. In the experiment of the complete reconstruction, the MDC scheme is assumed error-free, and both descriptions are received successfully. All the MSGs are extracted from both reconstructed sequences to produce the best-quality output video. In the experiment of the single-channel reconstruction, it is assumed that one description is entirely lost during transmission. The goal is to evaluate the redundancy coding efficiency of an MDC scheme.

The proposed MDC scheme is compared with the previous slice-group based MDC (SG-MDC) scheme [6]. In [6], the slice group pattern is fixed, and the SSG macroblocks are coded by using spatial prediction of motion vectors without reference frame selection. The proposed scheme is implemented on JM 10.1, the reference software of H.264/AVC. The rate-distortion optimization is turned on. The GOP structure is IPPP… without B-frame, and the number of reference frame is set to 10. The test sequences are of CIF size with frame rate 30. Fig. 6 shows the experimental results of *Stefan* sequence. The R-D curves are obtained by changing the quantization parameter (QP) of SSG (QPS) while keeping the quantization parameter (QP) of MSG (QPM) at a fixed value. Due to the page limit, we only present the results of *Stefan* sequence; however, the results of other test sequences are similar. The results of the single-channel reconstruction show that the proposed MDC scheme achieves higher PSNR over almost the entire bit-rate range, and the maximum improvement approaches 3 dB. The performance gain is attributed to the efficient redundancy coding of the proposed MDC scheme. Under the interchanged slice group pattern, the corresponding macroblock at the same position in the previous frame of each SSG macroblock is an MSG macroblock. This allows the proposed MDC scheme to add the motion vector of this MSG macroblock to the pool of STPMV and increase the accuracy of the predicted motion vector, resulting in smaller residuals and better PSNR. In addition, the reference frame selection saves the bits of reference frame indices. Finally, the improved mode decision achieves optimal SSG macroblock encoding by minimizing the rate-distortion cost. All these novel designs 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 [6] 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 PSNR 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.

Fig. 7 shows the relationship of the bit-rate and the redundancy in the single 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. Note that different MDC schemes may have different amount of redundancy at the same bit-rate, resulting in the intersection of the two curves.

Note that the simulation results are run on a Pentimu-4 PC with 1.25GB 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 time is about 0.16 (fps) for different test sequences.

4. CONCLUSIONS

In this paper, an H.264/AVC-based multiple description coding scheme has been presented. Two independently decodable descriptions are produced, and each description consists of two slice groups, the main slice group (MSG) and the side slice group (SSG). The MSG is encoded in the normal way to carry the basic information, and the SSG is encoded as redundancy by using larger quantization step sizes. By exploiting the motion information in neighboring MSG macroblocks, the proposed coding algorithms achieve efficient SSG encoding. Experimental results show that the proposed MDC scheme is superior to previous schemes by 3dB in PSNR.

5. ACKNOWLEDGMENT

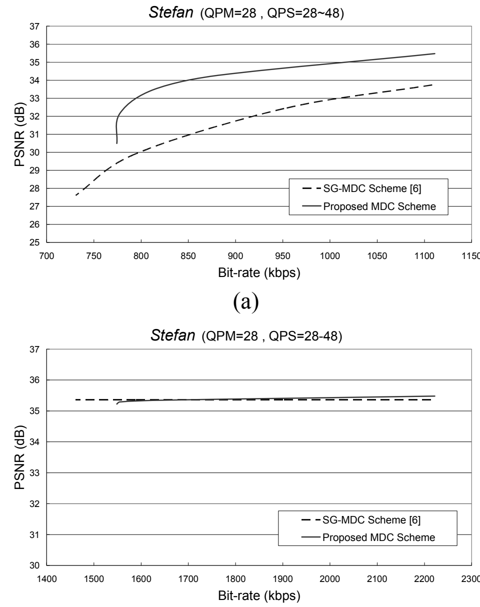
The authors thank Mr. Dong Wang for providing the software for generating the results presented in Section 3.

6. REFERENCES

[1] Y. Wang and Q.-F. Zhu, “Error control and concealment for video communication: a review,” *Proceedings of the IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.

[2] Y. Wang, A. R. Reibman, and S. Lin, “Multiple description coding for video delivery,” *Proceedings of the IEEE*, vol. 93, no.1, pp. 57–70, Jan. 2005.

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



(a)

(b)

Fig. 6. The R-D performance of *Stefan* sequence. (a) Single-channel reconstruction and (b) complete reconstruction.

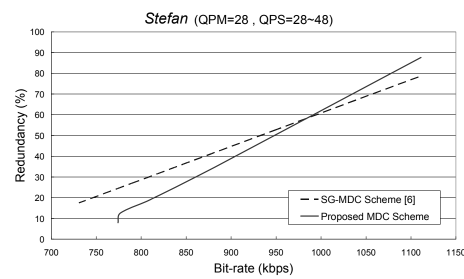


Fig. 7. The redundancy vs. bit-rate curve of *Stefan* sequence.

[4] A. Reibman, H. Jafarkhani, Y. Wang, M. Orchard, and R. Puri, “Multiple-description video coding using motion-compensated temporal prediction,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 193–204, Mar. 2002.

[5] J. G. Apostolopoulos, “Reliable video communication over lossy packet networks using multiple state encoding and path diversity,” in *Proc. VCIP*, Jan. 2001, pp. 392–409.

[6] D. Wang, N. Canagarajah, and D. Bull, “Slice group based multiple description video coding with three motion compensation loops,” in *Proc. IEEE Int. Symposium on Circuits and Systems*, Kobe, Japan, May 2005, pp. 960–963.

[7] C.-C. Su, J. J. Yao, and H. H. Chen, “Multiple description video coding based on slice group interchange,” *Picture Coding Symposium*, Beijing, China, Apr. 2006.

[8] W. Hantanong and S. Aramvith, “Analysis of macroblock-to-slice group mapping for H.264 video transmission over packet-based wireless fading channel,” in *Proc. IEEE Midwest Symposium on Circuits and Systems*, Aug. 2005, pp. 1541–1544.

[9] Ye-Kui Wang, M. Hannuksela, and V. Varsa, “The error concealment feature in the H.26L test model,” in *Proc. ICIP*, New York, Sept. 2002, pp. II-729–II-732.