A MULTIPLE DESCRIPTION VIDEO CODING MOTIVATED BY HUMAN VISUAL PERCEPTION

Saeed Moradi[*]

Department of Electrical and Computer Engineering, University of Toronto
saeed@comm.utoronto.ca

Saeed Gazor

Department of Electrical and Computer Engineering, Queen's University
s.gazor@queensu.ca

Tamás Linder

Department of Mathematics and Statistics, Queen's University
linder@mast.queensu.ca

ABSTRACT

We propose a simple multiple description (MD) video coding technique motivated by human visual perception. This method employs two simple parameters to characterize the smoothness and edge features of DCT blocks and measure their perceptual tolerance against visual distortion. We duplicate the key components of compressed video and split the remaining according to the calculated perceptual tolerance parameter. Simulation results reveal that our simple MD video coding method achieves superior performance compared to other similar techniques which lack to consider the perceptual distortion in the design problem.

Index Terms— **Multiple description, video coding, human visual perception**

I. INTRODUCTION

Multiple description coding (MDC) is a promising approach to enhance the error resilience for transmission over a diversity communication system with two channels. The objective is to send source information over each channel such that good performance is obtained when both channels work and the degradation is small when either channel fails (see, e.g., [1] for a comprehensive review). A general multiple description problem consists of encoding the source packet into two descriptions with rates R_1 and R_2, respectively. The reconstruction of the received description induces the side distortions D_1 and D_2, and the central distortion D_0, respectively. In many applications, a balanced design is considered, where $R_1 = R_2$ and $D_1 = D_2$ [2], [3].

Current MD video coding techniques widely use peak signal to noise ratio (PSNR) to measure the quality of reconstructed video. Since the perceptual distortion is ignored in calculation of PSNR, the distortion calculated by PSNR may not match the subjective perception of a human viewer [4], [5]. However, visual distortion and the quality of the video perceived by the human visual system (HVS) is a very crucial factor in assessment of quality of reconstructed video, and the existing MD video coding algorithms lack to address this issue. Recently, the problem of constructing visually optimized balanced multiple descriptions of images was considered in [6]. The basic idea in [6] is to use the DCQ algorithm of [5] to compute the appropriate quantization step size for different coefficients and achieve better visually reconstruction of the image. This idea can be extended to MD coding of video, however, this would require changing the standard core of the video encoder. This introduces extra overhead in terms of the side information, and increases the complexity. To address this problem, we propose a multiple description video coding technique motivated by human visual perception in order to generate two correlated streams. The implementation of the proposed MD video coding preserves the core of the standard video coder and does not significantly increase the coding complexity. We first calculate two simple parameters to characterize the smoothness and edge features of each block of an MPEG video frame. We employ these two parameters as a measure of the perceptual tolerance of DCT blocks against visual distortion. We then duplicate the key information such as motion vectors and some low-frequency DCT coefficients, and split the remaining DCT coefficients of prediction errors according to the calculated perceptual tolerance parameter. We compare our proposed MD video coder to the simple rate-splitting method introduced in [7] in terms of the single-channel reconstruction performance.

The organization of this paper is as follows. Section II explains the general scheme of a rate-splitting MD video coder and determination of visual distortion. Section III presents the proposed method. Simulation results are provided in Section IV. Finally, Section V concludes the paper.

II. RATE SPLITTING AND VISUAL DISTORTION

In general, the objective of a general MD video coding with rate splitting is to encode a video sequence into two video substreams with equal bit rates [7]. The splitting unit operates on the quantized DCT coefficients. Two descriptions are created by duplicating the header information and motion vectors. Then, the splitting algorithm determines the appropriate redundancy associated to each block. As a result, some of the DCT coefficients are duplicated in both descriptions and the rest are split between the two descriptions. The splitting is done so that when a quantized coefficient value is sent to one of the descriptions, a corresponding zero is sent to the other description.

At the receiver side, if both descriptions arrive at the central decoder, then the coefficients from the two streams are simply merged into a single stream that can be decoded by the MPEG standard decoder. However, if only one of the descriptions arrives at a side decoder, the received stream can still be simply decoded with the MPEG decoder. Suppose the bit rates used to send descriptions, in bits per source sample, are $R_1 = R_2$, and the total rate is $R = R_1 + R_2 = 2R_1$. One way to measure the efficiency of an MD video coder is by using the redundancy-rate distortion (RRD) curve [8]. We define the distortion of the best single description coder to be D_0 when R^* bits are used. Then,

[*]The author performed the work while at Queen's University.

redundancy is defined to be $\rho = \frac{R-R^*}{R^*}$, where R is the rate when the MD coder has central distortion D_0.

In order to numerically quantify the perceptual distortion, we use a simple method to evaluate tolerance of DCT coefficients against visual distortion. In general, side descriptions of an MD video coding system provide coarse reconstruction of the original data at the side decoders, while the central decoder uses each description to refine another one and achieve better performance over side decoders. As a result, video frames may be corrupted by degradations such as noise or blocking artifacts at the side decoders. These sources of degradation arise during generating two coarse descriptions of the original packet, and may have a perceptible effect on visual quality of the side reconstructed. Therefore, there is always a natural trade off between MD coding efficiency and video quality at the side decoders. In general higher redundancy results in better side video quality. Thus, we need to consider HVS characteristics and visual perception in efficiently designing the MD video coding algorithm such that the degradation in visual quality of reconstructed side video streams be as little as possible.

To address this issue, we employ a simple smoothness and edge parameters reported by Dittman et al. in [9], which was originally introduced for the purpose of video watermarking. We herein use these parameters to determine the amount of redundancy of each block of DCT coefficients of prediction errors. Although sophisticated techniques are developed to characterize the human visual perception, the calculation of smoothness and edge parameters of blocks is kept quite simple under this method. This is due to the fact that this calculation is to be done in the MPEG-stream domain and in real-time. We introduce the *Smooth* parameter defined simply as the number of DCT-coefficients which are not zero after quantization [9]. Thus, high value of Smooth indicates many frequency components in the block, and therefore a large perceptual tolerance against additional distortions introduced through splitting the DCT coefficients.

However, blocks with edge characteristics often have a lot of frequency components too. Thus, a second parameter *Edge* is also introduced to reduce the artifacts. The parameter Edge is calculated as the sum of the absolute values of the DCT coefficients 1, 2, 8, 9, 10, 16, and 17, as marked in the MPEG quantization matrix [9], which represent the lower DCT frequencies. High values in these components indicate that the block is likely to have edge characteristics which need to be preserved. The perceptual tolerance parameter *level* is then defined as [9]

$$L = \max \left\{ \min \left\{ -10S + 0.27E + 50, 100 \right\}, 0 \right\}, \qquad (1)$$

where L, S, and E denote, respectively, the parameters Level, Smooth, and Edge. The min and max operators keep the value of L between 0 and 100.

The constant values in equation (1) are evaluated through experiments [9]. The splitting rate factor is then determined based on the computed Level parameter. Generally, the larger values of Level indicate less perceptual tolerance against change of high frequency coefficients of the corresponding DCT block. We investigated the effect of Level parameter on the rate-splitting strategy, and observed that the blocks with more frequency components show higher perceptual tolerance against change of coefficients.

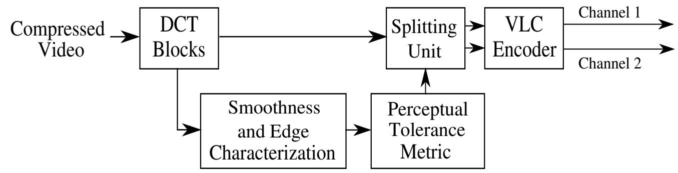


Fig. 1. Block diagram of the proposed MD video coding scheme.

III. PROPOSED RATE SPLITTING METHOD

In this section we present an MD video coding technique based on the method introduced in Section II. This technique applies the perceptual tolerance parameter Level introduced in Section II in order to determine the visually acceptable redundancy which can be introduced in a specific DCT block. Our proposed technique is very simple and allows only simple alternation and duplication of the non-zero DCT coefficients produced by a traditional one-layer encoder.

Figure 1 shows the block diagram of our proposed MD video coding scheme. In general, motion vectors constitute a crucial part of the video stream since they are used by the decoder to reconstruct the temporally predicted frames. Even though the motion vectors constitute very small portion of video stream, based on motion activity of the temporally predicted frame, the loss of motion vectors results in significant distortion. Inspired by this, we duplicate motion vectors and the side header information needed by the decoder in order to identify the received information and determine the mode of operation in each side description. Then, for each block in the frame, we compute the number of non-zero low-frequency DCT coefficients which needs to be duplicated in order to maintain a minimum visual quality. For each DCT block, the Level parameter is computed according to the algorithm described in Section II. Then, the number of non-zero DCT coefficients that need to be duplicated is

$$M = N \left(1 - (1-\rho)\left(1 - \left[\frac{L}{100}\right]\right)\right), \qquad (2)$$

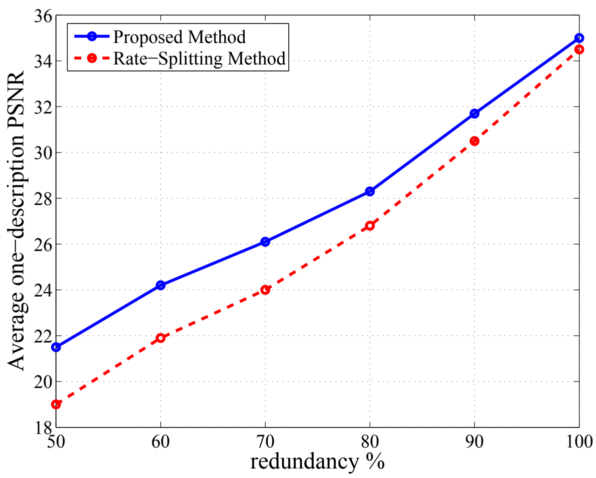
where N indicates the total number of non-zero DCT coefficients in the block, $0 \leqslant \rho \leqslant 1$ is the average target redundancy, $[x]$ returns the nearest integer to x, and L is the Level parameter. If we fix the Level value and increase the redundancy ρ, then M increases. On the other hand, if we fix ρ and decrease the Level value, then M decreases. Consequently, we duplicate the first M non-zero low-frequency coefficients into each description, and the remaining $N-M$ coefficients will be distributed in an alternation fashion between the descriptions according to their magnitude. It should be noted that we always duplicate the DC coefficient of each DCT block.

At the decoder, if both descriptions are received, it is a simple matter to merge the coefficients from the two bitstreams either into a single block of coefficients, or into a single bitstream that can be decoded by a standard MPEG decoder. In either case, the exact single-description video can be produced. If only one description is received, that bitstream can simply be decoded using the same method as a standard MPEG decoder.

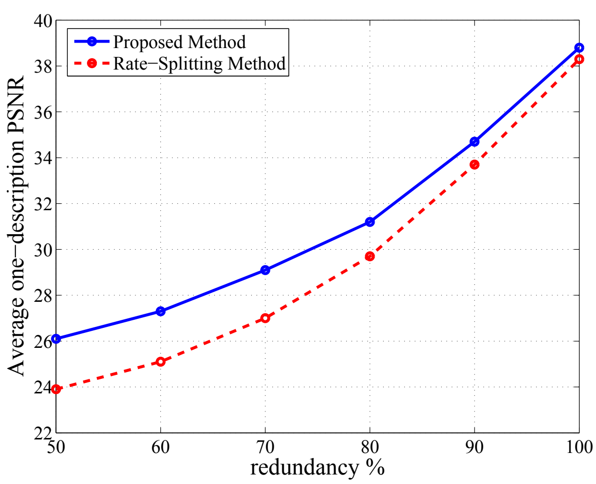
Our proposed method is summarized in the following steps:

1) Extract each DCT block of prediction errors from MPEG video stream.

2) Compute the Smooth parameter S as the number of the non-zero quantized DCT coefficients. Compute the Edge parameter E as the sum of the absolute values of the DCT coefficients 1, 2, 8, 9, 10, 16, and 17, as marked in MPEG quantization matrix.
3) Calculate the Level parameter $L = \max[\min(-10S + 0.27E + 50, 100), 0]$.
4) Compute the number of non-zero DCT coefficients that need to be duplicated as $M = 64 \left(1 - (1-\rho)(1 - \left[\frac{L}{100}\right])\right)$, where the number of coefficients of each DCT block is assumed to be 64, ρ is the target redundancy, and $0 \leqslant \rho \leqslant 1$.
5) Duplicate the first M non-zero DCT coefficients in two descriptions, and split the remaining $(64 - M)$ DCT coefficients between descriptions.



(a)

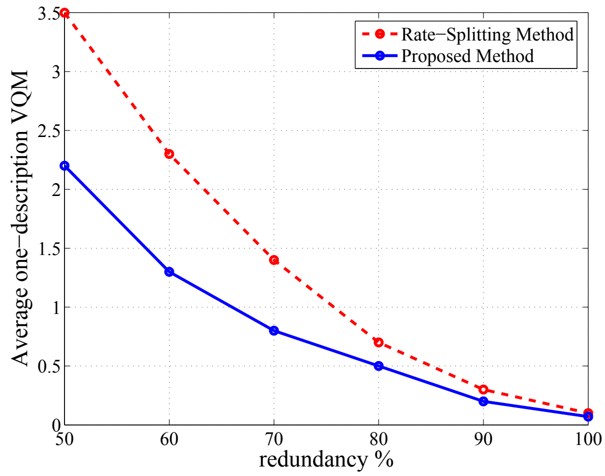


(b)

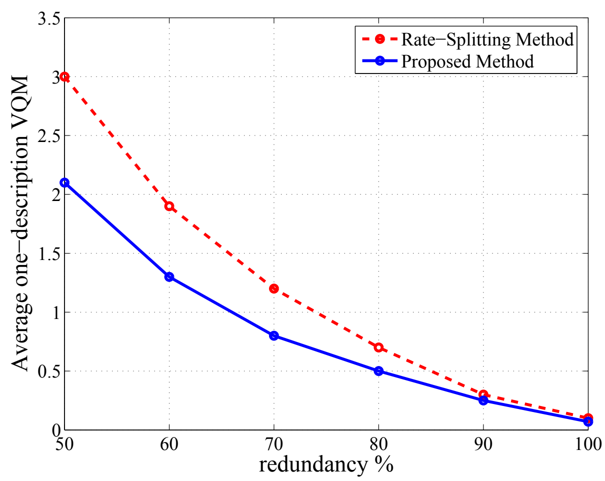
Fig. 2. PSNR performance comparison of proposed method and simple rate-splitting method for (a) Foreman MPEG video, and (b) Suzie MPEG video.

IV. SIMULATION RESULTS

We compare our proposed MD video coder to the simple rate-splitting method introduced in [7] in terms of the single-channel reconstruction performance. In both cases, the coders are built using an MPEG-1 coder. This comparison is made by assuming an entire description is lost; each coder produces identical video when both descriptions are received.



(a)



(b)

Fig. 3. VQM performance comparison of proposed method and simple rate-splitting method for (a) Foreman MPEG video, and (b) Suzie MPEG video. Note that lower VQM values correspond to higher perceptual video quality.

We consider two distortion measures to evaluate and compare the performance of our method. The first distortion measure is the average luminance PSNR across time. We have also evaluated the performance with VQM metric. VQM is developed by ITS (The Institute for Telecommunication Science) to provide an objective measurement for perceived video quality [10]. It measures the perceptual effects of video impairments including blurring, jerky and unnatural motion, global noise, block and color distortions, and combines them into a single metric. The testing results [10] show that VQM has a high correlation with subjective video quality assessment and has been adopted by ANSI (American National Standards Institute) as an objective video quality standard. In general, the lower VQM value indicates higher perceptual video quality.

(a)

(b)

Fig. 4. Reconstruction results for Foreman MPEG video, frame no. 296, with $\rho = 0.6$. (a) Proposed method, (b) Simple rate-splitting method.

The redundancy is expressed in terms of the overhead percentage over the reference total bit rate. The test sequences are two MPEG-1 video sequences, Foreman and Suzie, with a constant bit rate of 1.5 Mb/s and a frame rate of 25 frames per second and a GOP size of 18 and 12 frames, respectively. Figures 2 and 3 compare the performance of our proposed method to that of the simple rate-splitting method for the Foreman and Suzie MPEG video sequences in terms of the PSNR and VQM metrics. In both cases, our method outperforms the simple rate-splitting method. The PSNR performance is also improved since we duplicate the DCT coefficients according to their magnitude. It is worth mentioning that since Suzie MPEG video on average has lower frequency components, the VQM of proposed method is very close to that of rate-splitting method. This suggests that our proposed method results in better performance when applied to video sequences with high-frequency components. For higher redundancy, only very high-frequency DCT coefficients, which have almost imperceptible effect on visual distortion, are split between two descriptions. Consequently, it is expected that our proposed method performs better in lower redundancy, and this is confirmed in Figure 3 as well. Figure 4 presents subjective results for our proposed method against the simple rate spitting method observed for frame no. 296 of the Foreman sequence for $\rho = 0.6$ and a target bit rate of 1.5 Mb/s for the encoder.

V. CONCLUSION

We proposed a multiple description video coding technique that uses visual distortion criteria to generate two correlated streams. The main virtue of this technique is its low complexity. Our method duplicates motion vectors and some low-frequency DCT coefficients of prediction errors in both descriptions, and then applies a simple smoothness and edge detection algorithm in order to determine the amount of redundancy in each block of DCT coefficients that can be tolerated perceptually. Simulation results reveal that this MD video coding technique has superior performance in terms of visual quality and average PSNR compared to the traditional rate splitting methods which lack to address perceptual distortion in the design problem.

VI. REFERENCES

[1] V. K. Goyal, “Multiple description coding: Compression meets the network,” *IEEE Signal Processing Mag.*, vol. 18, pp. 74-93, Sep. 2001.

[2] L. Ozarow, “On a source coding problem with two channels and three receivers,” *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909-1921, Dec. 1980.

[3] A. A. El Gamal and T. M. Cover, “Achievable rates for multiple descriptions,” *IEEE Trans. on Inform. Theory*, vol. IT-28, pp. 851-857, Nov. 1982.

[4] J. Chen and T.N. Pappas, “Perceptual metrics and perceptual coders,” *Proceeding of Human Vision and Electronic Imaging*, San Jose, CA, Jan. 2001

[5] D.M. Chandler and S.S. Hemami, “Dynamic contrast-based quantization for lossy wavelet image compression,” *IEEE Trans. Image Processing,* vol. 14, no. 4, pp. 397-410, Apr. 2005.

[6] C. Tian and S. S. Hemami, “Visually Optimized Multiple Description Image Coding,” *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. II-1–II-4, May 2006.

[7] A. Reibman et al., “Multiple description video using rate-distortion splitting,” *Proc. IEEE Int. Conf. Image Processing*, vol. 1, 2001, pp. 978-981.

[8] Y. Wang et al., “Multiple description coding using pairwise correlating transforms,” *IEEE Trans. Image Processing*, vol. 10, pp. 351-366, Oct. 2001.

[9] J. Dittmann, M. Stabenau, and R. Steinmetz, “Robust MPEG Video Watermarking Technologies,” *Proceedings of the sixth ACM international conference on Multimedia*, pp. 71-80, 1998.

[10] M. Pinson and S. Wolf, “A New Standardized Method for Objectively Measuring Video Quality,” *IEEE Trans. Broadcasting*, vol. 50, no. 3, pp. 312-322, Sept. 2004.