A Mixed Layer Multiple Description Video Coding Scheme

Mohammad Kazemi, Khosrow Haj Sadeghi, and Shervin Shirmohammadi, *Senior Member, IEEE*

***Abstract*—Multiple description coding (MDC) is a technique where multiple streams from a source video are generated, each individually decodable and mutually refinable. MDC is a promising solution to overcome packet loss in video transmission over noisy channels, particularly for real-time applications in which retransmission of lost information is not practical. A problem with conventional MDC is that the achieved side distortion quality is considerably lower than single description coding (SDC) quality except at high redundancies which in turn leads to central quality degradation. In this paper, a new mixed layer MDC scheme is presented with no degradation in central quality, and providing better side quality (approximately as much as that of SDC) compared to conventional methods. Also, this property directly leads to higher average quality when delivering the video in lossy networks. For each discrete cosine transform coefficient, we generate two coefficients: base coefficient (BC) and enhancement coefficient which are combined together. When all descriptions are available, they are decomposed and decoded to achieve high quality video. When one description is not available, we use estimation to extract as much of the BC as possible from the received description. Simulation results show that the proposed scheme leads to an improved redundancy-rate-distortion performance compared to conventional methods. The algorithm is implemented in JM16.0 and its performance for two-description and four-description coding is verified by experiments.**

***Index Terms*—Multiple description coding, redundancy-rate-distortion, video coding, video packet loss mitigation, video transmission.**

I. Introduction

TRANSMITTING video in error prone environments, such as the Internet, is a challenging problem that has attracted considerable research in recent years. Multiple description coding (MDC) is one promising solution for such applications where multiple descriptions of the video signal are generated and transmitted over different channels [1]. Each description can be decoded independently with acceptable quality, while higher quality is achieved with higher number of received descriptions. Hence, unless all channels fail simultaneously, which is much less probable than single channel failure, video is delivered to the receiver. MDC also mitigates the need for retransmission—a particularly important feature for real-time applications such as video conferencing where retransmission is not practical. But MDC is not the only solution for multi-stream delivery. Another solution is scalable video coding (SVC), where the video is encoded into a base layer (BL) and one or more enhancement layers (EL). An EL is superimposed on the video decoded from the BL and all lower ELs. Therefore, for each layer to be usable all lower layers must be received at the decoder, otherwise the layer is discarded. SVC by nature is able to support video in multiple bit rates and fits well for video delivery in heterogeneous networks; however, it has no resilience against network loss. The strengths and shortcomings of MDC versus SVC have been well studied [2]–[5] and are beyond the scope of this paper. Hybrid MDC/SVC schemes are also possible, in which SVC layers are coded using MDC, or in another approach, each MDC description is coded again using SVC, but such solutions are also beyond the scope of our work. In this paper, we concentrate on MDC only and propose a mixed layer MDC scheme that improves the quality of the video compared to existing MDC solutions.

In MDC, when all of the descriptions are available, they are decoded by the central decoder and the distortion achieved is called central distortion. Otherwise, the available descriptions are decoded by the side decoder, and the resulting distortion is called side distortion.

The ability to independently decode individual descriptions is achieved at the cost of redundancy and, subsequently, a higher bit rate compared to single description coding (SDC). Hence, there are three main challenges associated with MDC. One challenge is to minimize this redundancy for a given desired side and central distortion. The other challenge is the lower quality of the video when one or more of the descriptions are lost. While MDC allows the decoder to display the video extracted from the available descriptions (even from just one description), the quality of the video in such cases is not very high. To achieve higher quality, more redundancy is needed, but this redundancy has no contribution to the central quality and deteriorates the rate/central-distortion behavior of the encoder. The third and final challenge is the drift problem which occurs due to predictive coding of the video signals. When a frame is correctly received but its reference has not

Manuscript received August 11, 2010; revised November 15, 2010, February 22, 2011, and May 18, 2011; accepted June 6, 2011. Date of publication June 13, 2011; date of current version February 8, 2012. This work was supported in part by the Iran Telecommunication Research Center. This paper was recommended by Associate Editor R. Hamzaoui.

M. Kazemi and K. H. Sadeghi are with the Department of Electrical Engineering, Sharif University of Technology, Tehran 11365, Iran (e-mail: m_kazemi@ee.sharif.edu; ksadeghi@ee.sharif.edu).

S. Shirmohammadi is with the Distributed and Collaborative Virtual Environment Research Laboratory, School of Information Technology and Engineering, University of Ottawa, Ottawa, ON K1N 6N5, Canada, and also with the Multimedia Processing Laboratory, School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran 14395, Iran (e-mail: sshirmohammadi@ut.ac.ir; shervin@discover.uottawa.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2011.2159431

been reconstructed exactly, the frame is reconstructed with an error. This new erroneous frame is then used as the reference of the next frames and hence the error propagates until the next Intra frame arrives. Therefore, a low side quality, even for one frame affects the whole group of pictures (GOP) unless an appropriate drift technique is used. In this paper, we address the above three challenges and propose a novel MDC-based scheme called mixed layer MDC (MLMDC) that provides a central distortion comparable to conventional schemes while achieving a side quality approximately as much as that of SDC, in particular for high activity videos. Similar to BL and EL layers in coarse granular scalable (CGS) coding [6], we utilize a base coefficient (BC) and an enhancement coefficient (EC) which are combined and inserted into the descriptions so that they are decomposed similar to decoding BL and EL in CGS at the central decoder. However, MLMDC does not have the hierarchical restriction of CGS and the source can be reconstructed from any subset of descriptions. The tradeoff to achieve this is a higher bit rate, which can be justified due to higher video quality and loss resiliency provided by MLMDC.

The rest of this paper is organized as follows. Section II provides an overview of the related work. In Section III, the framework of MLMDC including the encoder and the central and side decoders are presented. The side and central distortions and the redundancy rate are studied in Section IV, while in Section V a comparative study of our method with three other MDC schemes is presented. The experimental results are given in Section VI, and finally this paper is concluded in Section VII.

II. Related Work

MDC approaches can be categorized based on the domain in which partitioning of the video is done. In spatial domain MDC, the video source is partitioned in the spatial domain before the encoding process which is common for video and image MDC [7]–[9]. But subsampling in the spatial domain does not allow controlling the amount of redundancy. This problem can be solved by zero padding in the discrete cosine transform (DCT) domain where the amount of redundancy is determined by the number of padded zeros [8]. The optimum direction of zero padding is discussed in [9]. In temporal domain MDC, the descriptions are generated by subsampling the video source in time. For example, one can alternate the frames between descriptions such that even frames are in the first description and odd frames in the second description [10]–[12]. In both spatial and temporal domain approaches, the resolution of the descriptions is smaller than the original video. In the side decoder, due to the spatial-temporal correlation within each frame (spatial) and between neighboring frames (temporal), the lost data can be estimated from the received data using error concealment techniques. In frequency domain MDC, the process of MDC generation is carried out in the DCT domain, such as DCT coefficient partitioning [13], [14], multiple description scalar quantizer (MDSQ) [15], [16], [18], quantization based MDC [17], [29], [41], and multiple description transform coding (MDTC) [19]–[23]. In [13], the coefficients larger than a certain threshold are replicated in all descriptions and the smaller ones are divided between descriptions. In [14], the coefficients are split so that both descriptions are completely balanced in terms of both side quality and bit rate. In MDSQ, each source is assigned two indices, one for each description, such that with a given rate and side distortion, the maximum central quality is achieved. The methods which work with quantization decision levels can be placed in the category of MDSQ [18]. In quantization based MDC, the redundant data is the coarse quantized (lower-rate) representation of the primary data. The lower rate part might be GOP [17], a frame [41], or a slice of macroblocks (MBs) [29]. In [17], a GOP is encoded multiple times but in different rates, one for each description and the lower rate descriptions are discarded in the side and central decoders. The generated descriptions are unbalanced for each GOP and switching between descriptions is carried out when receiving an Intra frame. In this approach, there is no drift in the decoder but the video quality may be low for one or multiple GOPs. In [41], each description contains the frames that alternatively are high quality (fine quantization) and low quality (coarse quantization); therefore, the one-description video is not subjectively satisfactory. In [29], each frame is composed of high rate and low rate slices and, in contrast to [17] and [41], the descriptions are almost balanced with respect to rate and distortion. In MDTC, for both image and video, some of the DCT coefficients are paired and combined by a correlating transform and the transformed coefficients are sent to the decoder. A missed coefficient can be estimated from the received ones due to the correlation introduced by the transform.

There are also MDC schemes in which the partitioning is performed in multiple domains, for example spatial-temporal [24], and spatial-frequency [25]. In another group, namely forward error correction MDC, video partitioning is done on the compressed signal at the final stage [26], [27]. Finally, there are other MDC techniques which cannot be categorized in the above groups such as partitioning motion vectors [28], using matching pursuit [30], or using lapped orthogonal transform [31]. A more detailed review of MDC can be found in [32] and [1], the former being particularly for video while the latter is more general.

Our proposed MDC technique differs from the above since it combines the superimposing property generally seen in SVC with estimation techniques. As such, our technique achieves better central and side rate-distortion performance, making it suitable for channels with both high and low packet loss rates, as will be shown in Section VI. Next, we will present the details of our proposed MDC technique.

III. MLMDC Method

A. Encoder

Our proposed MDC technique is inspired from [19]–[21], but instead of combining different-frequency DCT coefficients, we combine base layer coefficients (BC) with enhancement layer coefficients (EC). Fig. 1 shows the block diagram of the encoder, the central decoder, and the side decoder of MLMDC. In Fig. 1(a), x_i and $\hat{x}_i$ are DCT coefficients at the ith position before and after quantization by Q, respectively. Subtracting

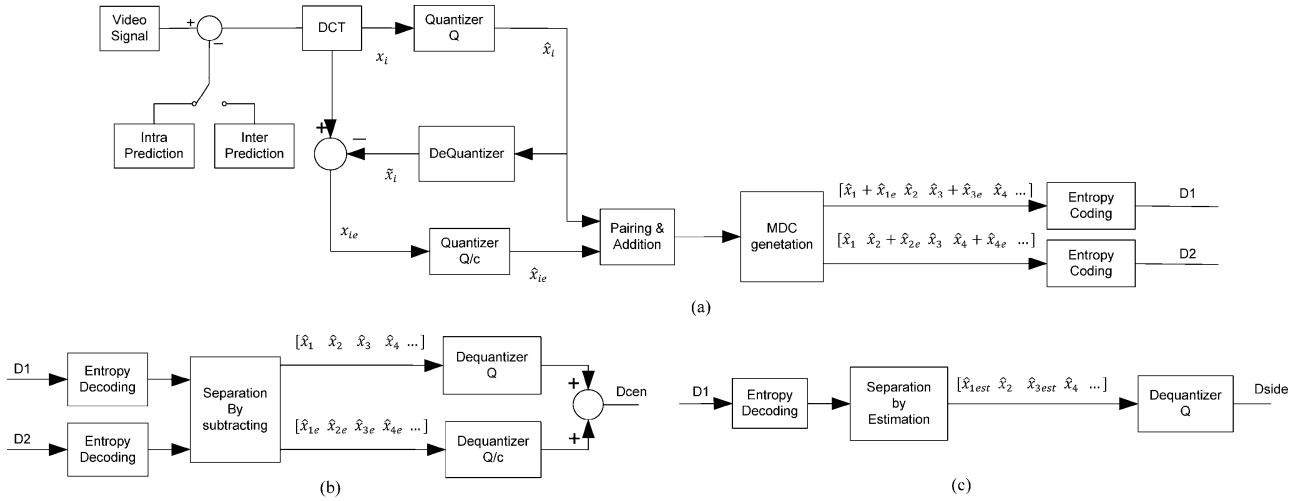


Fig. 1. Block diagram of MLMDC. (a) Encoder. (b) Central decoder. (c) Side decoder.

$\hat{x}_i$ from x_i results in a coarse quantization error which clearly cannot be quantized again by Q, and so a smaller quantization step size is required. In MLMDC, the second quantization is carried out with $Q/c, c > 1$. The lager the value of c, the lower the residual error between the original coefficients and the coefficients reconstructed from both descriptions. Hence, a higher central quality is achieved, but at the cost of a higher estimation error and higher bit rate. As shown in Fig. 1(a), the combined and not-combined coefficients are alternated between descriptions so that the redundancy due to combination and the number of coefficients that must be estimated are divided between descriptions. Similarly, in four-description coding, each description contains one quarter of the total combined coefficients. Generally, for n-description coding, in each description $1/n$ of the coefficients are combined which are the sources of redundancy and estimation error. The rest $(n-1)/n$ coefficients remain as base coefficients. This means that redundancy and estimation error in each description are diminished (with $1/n$) by the number of descriptions in n-description coding. Therefore, the efficiency of MLMDC will increase with increasing number of descriptions.

Assume a signal, composed of two coefficients $\{x_1, x_2\}$, is to be encoded using MDC. The coefficients $\{x_1, x_2\}$ are first quantized with a quantization step size of Q to form the BCs $\{\hat{x}_1, \hat{x}_2\}$. Then, the corresponding quantization errors, $\{x_{1e}, x_{2e}\}$, are quantized again to form ECs, $\{\hat{x}_{1e}, \hat{x}_{2e}\}$, which are then added to BCs to eventually form the CCs, $\{\hat{z}_1, \hat{z}_2\}$. Using the quantization rule of H.264/AVC, the BC, EC, and CC coefficients are obtained as given in

$$\begin{aligned} &\text{BC}: \hat{x}_i = sign(x_i)\left\lfloor \frac{|x_i|}{Q} + f \right\rfloor, \tilde{x}_i = Q\hat{x}_i \\ &\qquad x_{ie} = x_i - \tilde{x}_i \\ &\text{EC}: \hat{x}_{ie} = sign(x_{ie})\left\lfloor \frac{|x_{ie}|}{Q/c} + f \right\rfloor \\ &\text{CC}: \hat{z}_{ij} = \hat{x}_i + \hat{x}_{je} \quad i, j = 1, 2 \end{aligned} \tag{1}$$

where $\lfloor\ \rfloor$ denotes rounding to the nearest integer toward minus infinity, $\tilde{x}_i$ is the reconstructed coefficient, and Q and f are the quantization step size and the rounding offset, respectively.

There are two possibilities to construct the descriptions, such that each description contains the combined and not-combined coefficients. Equations (2) and (3) show both descriptions (denoted as D_1 and D_2) for these two possibilities, referred to as *CaseI* and *CaseII* hereafter

$$CaseI: \begin{cases} D_1 : \{(\hat{x}_1 + \hat{x}_{1e}), (\hat{x}_2)\} \\ D_2 : \{(\hat{x}_1), (\hat{x}_2 + \hat{x}_{2e})\} \end{cases} \tag{2}$$

$$CaseII: \begin{cases} D_1 : \{(\hat{x}_1 + \hat{x}_{2e}), (\hat{x}_2)\} \\ D_2 : \{(\hat{x}_1), (\hat{x}_2 + \hat{x}_{1e})\}. \end{cases} \tag{3}$$

Fig. 1(a) shows *CaseI* MLMDC in which, unlike *CaseII*, BC and EC are obtained from the same DCT coefficient and are hence correlated. This is the key difference between *CaseI* and *CaseII* and has an important role in rate-distortion performance, as will be shown later.

B. Central Decoder

In the central decoder, both descriptions are available and hence BC and EC can be separated by subtracting the uncombined coefficient $(\hat{x}_i)$ from the combined coefficient $(\hat{z}_{ij})$. As shown in Fig. 1(b), after dequantization of BC and EC, they are added to reconstruct the DCT coefficient. Therefore, the central distortion can be computed as follows:

$$\begin{aligned} &\hat{x}_{je} = \hat{z}_{ij} - \hat{x}_i \\ &x_{\text{cen}_j} = Q\hat{x}_j + Q\hat{x}_{je}/c \\ &D_{\text{cen}_j} = E\left[\left(x_j - x_{\text{cen}_j}\right)^2\right] \end{aligned} \tag{4}$$

where x_{cen} and D_{cen} are the reconstructed coefficient and distortion, respectively, achieved by the central decoder. Note that in terms of central distortion, *CaseI* and *CaseII* produce identical results.

C. Side Decoder

In the side decoder, we must estimate the base coefficient from $\hat{z}_{ij}$. As an example, consider *CaseI* when only D_1 is received, $\hat{x}_1$ is not available and we extract it from the $\hat{x}_1 + \hat{x}_{1e}$ coefficient by estimation as shown in Fig. 1(c). Independent

of which case is used, we define $\hat{z}$ to be the sum of $\hat{x}_b$ (BC) and $\hat{x}_e$ (EC) as in

$$\hat{z} = \hat{x}_b + \hat{x}_e. \tag{5}$$

From *estimation theory*, the optimal estimator of $\hat{x}_b$ from $\hat{z}$ is the conditional expected value of $\hat{x}_b$ when $\hat{z}$ is available [33]

$$\hat{x}_{est} = E\{\hat{x}_b|\hat{z}\}. \tag{6}$$

If the distributions of $\hat{x}_b$ and $\hat{x}_e$ are known, we can use (7)–(10), (11) for estimating $\hat{x}_b$

$$\hat{x}_{est} = \sum_{\hat{x}_b} \hat{x}_b P_{\hat{x}_b|\hat{z}}(\hat{x}_b|\hat{z}) \tag{7}$$

$$P_{\hat{x}_b|\hat{z}}(\hat{x}_b|\hat{z}) = \frac{P_{\hat{x}_b,\hat{z}}(\hat{x}_b,\hat{z})}{P_{\hat{z}}(\hat{z})} = \frac{P_{\hat{x}_b}(\hat{x}_b)P_{\hat{z}|\hat{x}_b}(\hat{z}|\hat{x}_b)}{P_{\hat{z}}(\hat{z})} \tag{8}$$

$$P_{\hat{z}|\hat{x}_b}(\hat{z}|\hat{x}_b) = P_{\hat{x}_e|\hat{x}_b}(\hat{z}-\hat{x}_b) \tag{9}$$

$$\begin{aligned} p_{\hat{z}}(\hat{z}) &= \textstyle\sum_{\hat{x}_b} P_{\hat{x}_b,\hat{z}}(\hat{x}_b,\hat{z}) = \sum_{\hat{x}_b} P_{\hat{x}_b}(\hat{x}_b)P_{\hat{z}|\hat{x}_b}(\hat{z}|\hat{x}_b) \\ &= \textstyle\sum_{\hat{x}_b} P_{\hat{x}_b}(\hat{x}_b)P_{\hat{x}_e|\hat{x}_b}(\hat{z}-\hat{x}_b) \end{aligned} \tag{10}$$

and consequently

$$\hat{x}_{est} = \frac{\sum_{\hat{x}_b} \hat{x}_b P_{\hat{x}_b}(\hat{x}_b)P_{\hat{x}_e|\hat{x}_b}(\hat{z}-\hat{x}_b)}{\sum_{\hat{x}_b} P_{\hat{x}_b}(\hat{x}_b)P_{\hat{x}_e|\hat{x}_b}(\hat{z}-\hat{x}_b)}. \tag{11}$$

It can be seen that the above estimation equations are dependent on the individual as well as the joint probability functions of BC and EC. BC and EC are dependent in *CaseI* and independent in *CaseII*, meaning that their joint probability functions and, consequently, their corresponding side distortions may be different.

IV. Rate and Distortions in MLMDC

A. Background

MLMDC is a DCT-domain MDC and is thus performed on DCT coefficients. To better understand the rate-distortion in MLMDC, we present here a short review of some important concepts of DCT coefficients applicable to MLMDC.

Three models for distribution of DCT coefficients have been introduced, namely, generalized Gaussian [34], Laplacian [35], and Cauchy [36], which progressively provide better fitness. However, Laplacian is more popular and widely used in rate control applications. Furthermore, the exponential form of Laplacian is more suitable for our analytical derivations; hence, we apply it as follows:

$$\begin{aligned} f_b(x_b) &= \tfrac{\lambda}{2}\mathrm{e}^{-\lambda|x_b|} \\ \sigma^2 &= \tfrac{2}{\lambda^2} \end{aligned} \tag{12}$$

where λ and σ are the Laplace distribution parameter and standard deviation of the signal x_b, respectively. Each frequency component of the DCT block has its particular λ. The low frequency coefficients have larger variances and hence a smaller λ. This is known as the compacting property of the DCT transform which concentrates the energy of the signal mostly in low frequency coefficients.

Sullivan has shown that for Laplacian sources, the quantization rule of (1) is nearly optimal [38], [39]. In the normal quantization mode of H.264/AVC, the rounding offset, f, is set to 1/3 for intra frames and 1/6 for inter frames due to different distributions of coefficients in intra and inter frames. In adaptive rounding offset mode, f is adapted so that the nonzero reconstructed coefficient is placed at the centroid of the corresponding quantization interval, as described in (13). This option is also available in JM and achieves about 1 dB improvement in rate-distortion performance [40]. For Laplacian distribution, the adapted f is the same and less than 1/2 for all quantization levels, although it varies with λ and Q. The value of f which achieves minimum distortion is 1/2, and choosing $f < 1/2$ causes an increase in distortion. However, the resulting bit rate reduction due to the introduced dead zone is such that rate-distortion is improved. In all of our simulations in this paper, in accordance with the adaptive rounding offset technique of JM, f is determined so that (13) is satisfied

$$\begin{aligned} Q\hat{x}_b &= E[x|_{\hat{x}_b=N}] \\ N &\neq 0. \end{aligned} \tag{13}$$

The quantization distortion for Laplacian sources has been obtained already in [37] as

$$D_q = \frac{\lambda Q e^{f\lambda Q}(2+\lambda Q-2f\lambda Q)+2-2e^{\lambda Q}}{\lambda^2(1-e^{\lambda Q})}. \tag{14}$$

It can be proved that for a small enough λQ, D_q approaches D_{qu} given in

$$D_{qu} = \left(f^2 - f + \frac{1}{3}\right)Q^2 \tag{15}$$

which is the quantization noise of a uniform signal quantized with Q.

B. Side Distortion

In addition to quantization distortion, the side decoder also deals with estimation distortion. The estimation distortion is computed using the "reconstructed" coefficient as the reference of distortion

$$D_{est} = E[(\tilde{x} - Q\hat{x}_{est})^2]. \tag{16}$$

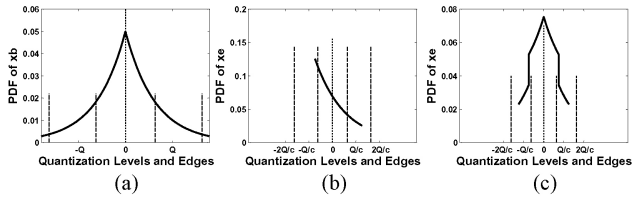
Side distortion, which considers both quantization and estimation distortions, will be computed using the "original" coefficient as the reference of distortion

$$D_{side} = E[(x - Q\hat{x}_{est})^2]. \tag{17}$$

For calculating the estimation and the side distortions, we first need to know $\hat{x}_{est}$. In the following parts, the optimal estimator is obtained for both *CaseI* and *CaseII* and then the performance of MLMDC with respect to the aforementioned distortions is studied.

If $\hat{x}_b$ is the quantized form of x_b its probability distribution function can be found as follows:

$$\begin{aligned} p_b(0) &= p_b(\hat{x}_b = 0) = 2\textstyle\int_0^{(1-f)Q} f_b(x_b)dx_b = 1- \\ &\quad e^{-(1-f)\lambda Q} = P_{b0} \\ p_b(n) &= p_b(\hat{x}_b = n) = \textstyle\int_{(n-f)Q}^{(n+1-f)Q} f_b(x_b)dx_b = \\ &\quad \tfrac{(1-e^{-\lambda Q})e^{\lambda Qf}}{2}e^{-\lambda Q|n|} = P_{bN}e^{-\lambda Q|n|}. \end{aligned} \tag{18}$$



(a) (b) (c)

Fig. 2. Quantization of a Laplacian source in MLMDC when $\lambda = 1/10$, $Q = 16$, and $c = 2$ for generation of: (a) BC, (b) EC *CaseI* and $\hat{x}_b = 1$, and (c) EC *CaseII*.

The distribution of x_e in *CaseI* is obtained directly by the distribution of x_b divided by the probability of the corresponding $\hat{x}_b$ as follows:

$$f^{I}_{(x_e|\hat{x}_b)}(x_e|\hat{x}_b) = \begin{cases} \text{if } \hat{x}_b > 0 & \begin{cases} (\lambda e^{-\lambda x_e})/P_{bN} & -fQ < x_e < (1-f)Q \\ 0 & \text{otherwise} \end{cases} \\ \text{if } \hat{x}_b < 0 & \begin{cases} (\lambda e^{\lambda x_e})/P_{bN} & -(1-f)Q < x_e < fQ \\ 0 & \text{otherwise} \end{cases} \\ \text{if } \hat{x}_b = 0 & \begin{cases} (\lambda e^{-\lambda|x_e|})/P_{bo} & -(1-f)Q < x_e < (1-f)Q \\ 0 & \text{otherwise} \end{cases} \end{cases} \tag{19}$$

where P_{bN} and P_{b0} were defined in (18). The distribution of x_e is determined by the sign of $\hat{x}_b$ if $\hat{x}_b \neq 0$, and the sign of x_b if $\hat{x}_b = 0$, as expressed in (19).

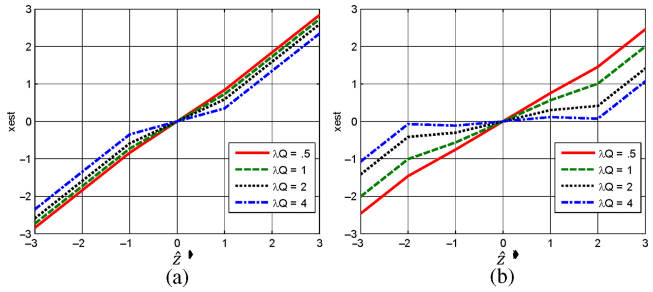
In *CaseII*, the coefficients are independent and hence to find the distribution of x_e, we can use that of *CaseI* without the mentioned divisions and then adding them together for all possible $\hat{x}_b$ values. The closed form of that with a good approximation is given by

$$f^{II}_{(x_e|\hat{x}_b)}(x_e|\hat{x}_b) = f^{II}_{x_e}(x_e) = \begin{cases} \frac{\lambda}{2}e^{-\lambda|x_e|} + \frac{\lambda}{e^{\lambda Q}-1} & |x_e| \le fQ \\ \frac{\lambda}{2}\frac{e^{\lambda Q}}{e^{\lambda Q}-1}e^{-\lambda|x_e|} & fQ < |x_e| \le (1-f)Q. \end{cases} \tag{20}$$

Derivation details of (20) can be found in the Appendix. Here, it is assumed that both BC and EC have the same Laplace parameter in *CaseII*. The case of BC and EC with unequal Laplace parameters will be discussed later in Section V.

Signals x_b and x_e and their corresponding quantized coefficients, $\hat{x}_b$ and $\hat{x}_e$, are shown in Fig. 2 for *CaseI* and *CaseII*. In *CaseI,* when $\hat{x}_b$ is positive [Fig. 2(b)], the interval at which the distribution of x_e is nonzero is not symmetric around zero and it has a shift toward plus infinity. This happens since the optimum f obtained by (13) is less than 1/2 and the interval $[-fQ \quad (1-f)Q]$ described in (19) is not symmetric.

The number of enhancement levels is determined by the value of c. A larger value of c clearly leads to more enhancement levels. For example, for the case demonstrated in Fig. 2, we have three possible levels for EC : $0, +1, -1$, and there is no data for the higher quantization levels. The probability of



(a) (b)

Fig. 3. Estimation curves for *CaseI* . (a) c = 1.5. (b) c = 3.5.

the levels in Fig. 2, p_e, can be computed as follows:

$$p_{e|\hat{x}_b}(1) = p_{e|\hat{x}_b}(\hat{x}_e = 1) = \int_{(1-f)Q/c}^{(1-f)Q} f_{x_e|\hat{x}_b}(x_e|\hat{x}_b)\,dx_e$$

$$p_{e|\hat{x}_b}(-1) = p_{e|\hat{x}_b}(\hat{x}_e = -1) = \int_{A}^{-(1-f)Q/c} f_{x_e|\hat{x}_b}(x_e|\hat{x}_b)\,dx_e$$

where A is $-fQ$ for *CaseI* and $-(1-f)Q$ for *CaseII*

$$p_{e|\hat{x}_b}(2) = p_{e|\hat{x}_b}(\hat{x}_e = 2) = 0, \quad p_{e|\hat{x}_b}(-2) = p_{e|\hat{x}_b}(\hat{x}_e = -2) = 0$$

$$p_{e|\hat{x}_b}(0) = p_{e|\hat{x}_b}(\hat{x}_e = 0) = 1 - (p_{e|\hat{x}_b}(1) + p_{e|\hat{x}_b}(-1) + p_{e|\hat{x}_b}(2) + p_{e|\hat{x}_b}(-2)). \tag{21}$$

For *CaseI* and *CaseII*, the corresponding EC distribution function from (19) and (20) must be used. One can easily prove that for *CaseI*

$$\begin{aligned} &\forall\, m\,,\ n\ \in \mathbb{Z} \ge 0 \\ &p_{e|m}(1) = p_{e|n}(1) = p_{e|-m}(-1) = p_{e|-n}(-1) = P_{e1p} \\ &p_{e|m}(2) = p_{e|n}(2) = p_{e|-m}(-2) = p_{e|-n}(-2) = P_{e2p} \\ &p_{e|m}(-1) = p_{e|n}(-1) = p_{e|-m}(1) = p_{e|-n}(1) = P_{e1n} \\ &p_{e|m}(-2) = p_{e|n}(-2) = p_{e|-m}(2) = p_{e|-n}(2) = P_{e2n} \\ &p_{e|m}(0) = p_{e|n}(0) = 1 - \left(P_{e1p} + P_{e2p} + P_{e1n} + P_{e2n}\right) = P_{e0}. \end{aligned} \tag{22}$$

For *CaseII*, clearly the statements of (22) are true and furthermore

$$\begin{aligned} P_{e1p} &= P_{e1n} \\ P_{e2p} &= P_{e2n}. \end{aligned} \tag{23}$$

Using the distributions given in (18) and (21), the closed form of (11) is obtained as in

$$\hat{x}_{est} = \begin{cases} \hat{z} - sign(\hat{z})\,Z_0 & |\hat{z}| \ge 2 \\ N_0\hat{z} & |\hat{z}| < 2 \end{cases} \tag{24}$$

where

$$Z_0 = \frac{2P_{e2p}e^{2\lambda Q} + P_{e1p}e^{\lambda Q} - P_{e1n}e^{-\lambda Q} - 2P_{e2n}e^{-2\lambda Q}}{P_{e2p}e^{2\lambda Q} + P_{e1p}e^{\lambda Q} + P_{e0} + P_{e1n}e^{-\lambda Q} + P_{e2n}e^{-2\lambda Q}}$$

$$N_0 = \frac{3P_{e2n}e^{-2\lambda Q} + P_{e0} + 2P_{e1n}e^{-\lambda Q} - P_{e2n}}{P_{e2n}e^{-2\lambda Q} + P_{e0} + P_{e1n}e^{-\lambda Q} + P_{e2n} + P_{b0}P_{e1p}e^{\lambda Q}/P_{bN}}. \tag{25}$$

It can be shown that Z_0 and N_0 are functions of λQ. The estimation curves for some typical values of λQ and $c = 1.5$ and $c = 3.5$ are depicted in Figs. 3 and 4 for *CaseI* and *CaseII*, respectively.

These figures show that *CaseI* and *CaseII* have approximately the same estimators. This happens because $\hat{x}_b$ related terms are dominant and common in both cases. In addition,

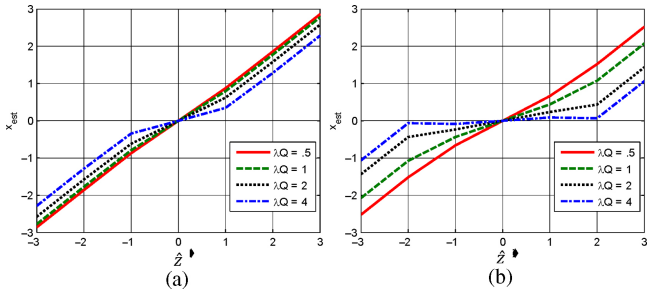


Fig. 4. Estimation curves for *CaseII*. (a) c = 1.5. (b) c = 3.5.

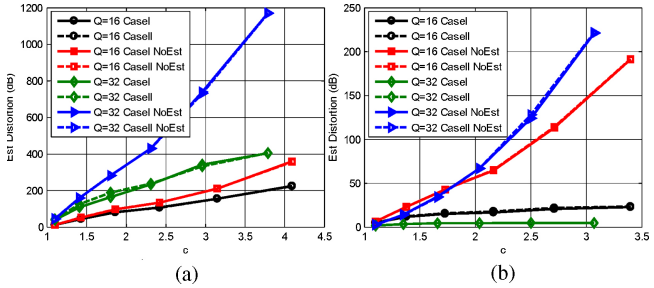


Fig. 5. Estimation distortion for *CaseI* and *CaseII*, Q = 16, 32. (a) $\lambda = 1/20$. (b) $\lambda = 1/5$.

some estimator lines behave approximately similar to the identity function; i.e., the input and output of the estimators are approximately the same. An identical estimator means that the standard decoder does not need to be modified for side decoding; an important feature for applications that are already deployed and difficult to change. Equation (24) declares that the value of Z_0 determines the difference of the estimator and the identity function. Using a numerical approach, we find that as long as c satisfies (26), Z_0 in (24) is less than 0.2. This means that using the estimation function for the combined coefficients larger than 2, the estimated coefficient will be the combined coefficient minus 0.2; i.e., the change in the combined coefficient is less than 10%; hence estimating can be skipped for that case

$$c \leq \left(\frac{6.425}{\lambda Q - 0.4274} \right)^{0.1131}. \tag{26}$$

Now, $\hat{x}_{est}$ for some typical values of λ and Q is computed and D_{est} and D_{side} are obtained from (16) and (17) and are depicted in Figs. 5 and 6 for both cases. Both cases are similar with respect to estimation distortion but very different in side distortion; this difference originates from the correlation between sources in *CaseI*. To see how the correlation affects the distortions, the estimation and side distortion resulted by an identity function estimator are obtained in (27) and (28), respectively

$$\begin{aligned} D_{est} &= E\left[(\tilde{x}_b - Q\hat{x}_{est})^2\right] = E\left[(\tilde{x}_b - Q\hat{z})^2\right] = \\ & E\left[(\tilde{x}_b - Q(\hat{x}_b + \hat{x}_e))^2\right] = Q^2 E\left[\hat{x}_e^2\right]. \end{aligned} \tag{27}$$

The $E\left[\hat{x}_e^2\right]$ term is common for both *CaseII* and *CaseI*, hence we have the same estimation distortion for both cases

$$\begin{aligned} & D_{side} = E\left[(x_b - Q\hat{x}_{est})^2\right] = E\left[(x_b - Q\hat{z})^2\right] = \\ & E\left[(x_b - Q(\hat{x}_b + \hat{x}_e))^2\right] = E\left[x_b^2\right] + Q^2 E\left[\hat{x}_b^2\right] + \\ & Q^2 E\left[\hat{x}_e^2\right] - 2QE\left[x_b\hat{x}_b\right] + 2Q^2 E\left[\hat{x}_b\hat{x}_e\right] - \\ & 2QE\left[x_b\hat{x}_e\right] = A_c + A_d. \end{aligned} \tag{28}$$

The first four terms are common for both cases and are denoted as A_c; the remaining two terms are denoted as A_d. A_d is zero for *CaseII* since the variables are independent. To obtain A_d for *CaseI*, we approximate x_b with the coefficients reconstructed by the central decoder; so we have

$$\begin{aligned} E\left[x_b\hat{x}_e\right] &\approx E\left[(Q\hat{x}_b + Q\hat{x}_e/c)\hat{x}_e\right] \\ &= QE\left[\hat{x}_b\hat{x}_e\right] + QE\left[\hat{x}_e^2\right]/c \end{aligned} \tag{29}$$

and therefore

$$\begin{aligned} A_d \approx 2Q^2 E\left[\hat{x}_b\hat{x}_e\right] - 2Q\left(QE\left[\hat{x}_b\hat{x}_e\right] + QE\left[\hat{x}_e^2\right]/c\right) = \\ -2Q^2 E\left[\hat{x}_e^2\right]/c. \end{aligned} \tag{30}$$

While A_d is zero for *CaseII*, it is explicitly negative for *CaseI*. This is the reason behind a smaller side distortion in *CaseI* compared to *CaseII*.

The distortion achieved by not using the estimation functions is also shown in Figs. 5 and 6. For small values of c, as we had predicted, there is no need for estimation. As Fig. 5 shows, the estimation distortion for $\lambda = 1/5$ and $Q = 32$ is less than when $Q = 16$. For each λ, there exists a particular value of Q after which the estimation error is decreased with Q. Increasing Q leads to a higher percentage of coefficients which can be estimated without error; this happens when both $\hat{x}_b$ and $\hat{x}_e$ are zero. One can easily show that this situation will be more probable for larger Q. Therefore, it is reasonable to have a smaller estimation distortion for large enough values of Q.

Surprisingly, the side distortion is decreasing with c for a particular range of c values for *CaseI* and $\lambda = 1/5$. In other words, contrary to what was expected, the combination has reduced side distortion in this region. In *CaseI*, as shown in Fig. 2, most of the times BC and EC have the same sign and therefore the expected absolute value of CC is larger than BC; i.e., the combination causes the quantization decision levels to move toward zero. In other words, the combination of BC and EC for small values of c is equivalent to a simple quantization (no combination) but with the quantization decision levels shifted toward zero or equivalently larger rounding offset (closer to 1/2). We already stated that $f = 1/2$ achieves less distortion compared to the rate/distortion optimized value obtained using (13), which is smaller than 1/2 particularly for signals with a large λ.

C. Central Distortion

If both descriptions (BC and CC) are available, EC will be obtained easily by subtracting BC from CC. The central distortion is equal to the quantization noise of x_e when quantized by Q/c. It can be easily shown that for integer values of c, the central distortion is equivalent to quantization distortion of (14) when Q/c is used as quantization step size. However, simulations show that for non-integer values, this

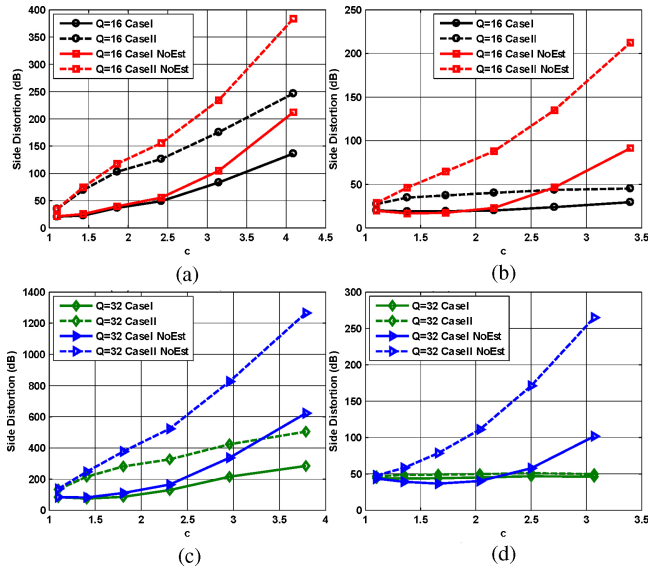


Fig. 6. Side distortion for *CaseI and CaseII*. (a) $\lambda = 1/20$ Q = 16. (b) $\lambda = 1/5$ Q = 16. (c) $\lambda = 1/20$ Q = 32. (d) $\lambda = 1/5$ Q = 32.

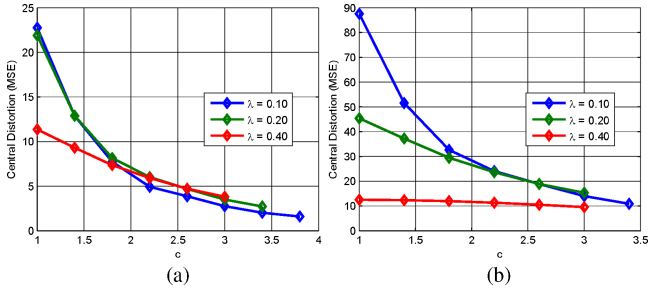


Fig. 7. Central distortion versus c for some typical values of λ. (a) $Q = 16$. (b) $Q = 32$.

equality is still true but approximately

$$D_{cen} \cong D_q|_{Q=\frac{Q}{c}}. \tag{31}$$

The central distortion versus c has been depicted in Fig. 7 for typical values of λ and Q.

For $Q = 16$ and $\lambda = 1/10, 1/5$, D_q is obtained from (15) and is independent of λ. One can see that the central distortion by a given c is degraded when λ or Q increases; in other words, the performance of MLMDC vanishes for large enough λQ. Hence, this scheme is not very effective for low content (small variance) sources with a high degree of compression.

D. Enhancement Rate

The other effect of the combination, in addition to estimating distortion in the side decoder, is the excess rate needed to transmit the combined coefficients. In *CaseII*, $\hat{z}$ is composed of two independent coefficients; hence, it has a larger variance compared to the uncombined coefficient ($\hat{x}_b$). As a result, the rate needed to transmit $\hat{z}$ is clearly more than when $\hat{x}_b$ is transmitted. In *CaseI*, the variance of $\hat{z}$ also depends on the correlation between BC and EC; however as shown later, this correlation does not significantly change the variance of $\hat{z}$ and the rate in *CaseI* is increased as well.

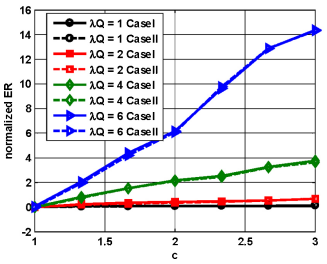


Fig. 8. ER_n versus c for some typical values of λQ.

The excess rate due to combination is denoted as enhancement rate (ER) and the rate needed for BC transmission is denoted as base rate (BR). For bit rate estimation, we calculate the entropy rate for each DCT coefficient. Therefore ER is computed as follows:

$$ER = \sum_{\hat{z}} P_{\hat{z}}(\hat{z}) \log_2\left(\frac{1}{P_{\hat{z}}(\hat{z})}\right) - \sum_{\hat{x}_b} P_{\hat{x}_b}(\hat{x}_b) \log_2\left(\frac{1}{P_{\hat{x}_b}(\hat{x}_b)}\right). \tag{32}$$

The entropy rate is optimal, but the context adaptive algorithms used in H.264/AVC, in particular CABAC, are very efficient and encode the quantized coefficients near the entropy rate. To study the effect of combination on the transmission rate, the absolute value of ER is not suitable. Instead, the normalized value, ER_n, is defined as the ratio of ER and BR

$$ER_n = \frac{ER}{BR}. \tag{33}$$

ER_n is a function of λQ. Fig. 8 shows the variation of this variable with c, for some typical values of λQ. ER and BR are the average of 40 independent simulations performed with 2000 samples. It can be seen that, first, both *CaseI* and *CaseII* have similar ER_n behaviors and, second, the signals with smaller λQ have a lower ER_n. In other words, for small values of λ or Q, combination with EC does not incur a considerable excess rate. But for large λQ, BR is small and the excess rate due to combination (ER) will be larger; i.e., both the nominator and the denominator in (33) vary in the direction of a larger ER_n.

For larger λQ, in addition to larger ER_n, we saw that the central distortion performance is not as good as smaller λQ. This means that the improvement in the central distortion for small variance sources is not very much and it may not be able to compensate the redundancy rate resulting from combination.

V. Performance Comparison

So far we have studied the performance of MLMDC with respect to distortion and rate, and found that central distortion will be reduced at the cost of excess side distortion and excess rate. Actually, this is a common property for all MDC schemes where redundancy, side distortion, and central distortion are compromising variables, and the improvement in central distortion must compensate the side distortion and redundancy rate. To judge the performance of our method, the redundancy-rate-distortion (RRD) criterion is used [20], which is defined as the side distortion achieved at a given

redundancy and a fixed central distortion. Here, redundancy is the bit rate needed to send both descriptions in excess of SDC. In a fixed central distortion, an amount of redundancy results in a particular side distortion, and the algorithm with the smaller side distortion is therefore more satisfactory.

The performances of three MDC scenarios namely Splitting, Replication, and MDTC [21] are compared with MLMDC. In Splitting, coefficients are divided between descriptions: odd coefficients in the first description and even coefficients in the second description. The absent coefficients in each description are filled with zeros. Zero filling is beneficial for both rate and distortion. All DCT quantized coefficients are most of the time zero than any other value (this is the reason for using the famous zero run-length coding scheme); therefore, zero filling increases the probability of zeros and decreases the entropy, hence reducing bit rate. Splitting has the best rate/central-distortion performance but the lowest side quality. In Replication, the coefficients are replicated in both descriptions, giving the best side decoding and the worst central decoding performance. Splitting and Replication are extremes of MDC strategies. In MDTC, instead of splitting coefficients themselves, they are first combined by a correlating transform and then the transformed coefficients are split. At the decoder, when one description is not received correctly, the missed coefficient can be estimated from the received one, due to the imposed correlation. We consider two sources, x_1 and x_2, both with and without the same Laplace parameters, to discuss the rate-distortion and MDC functionality of the MLMDC and to compare it with the other MDC scenarios.

The descriptions are generated as in (2) and (3) for MLMDC with $c = 2$. The MDTC algorithm has been developed and optimized for Gaussian distribution; however, for comparison with MLMDC, we used Laplacian sources. The quantization described in (1) with $f = 1/2$, the optimal non-orthogonal combination matrix, and the optimum linear estimator are used to generate MDTC descriptions as follows:

$$\begin{bmatrix} \hat{z}_1 \\ \hat{z}_2 \end{bmatrix} = T \begin{bmatrix} \hat{x}_{1b} \\ \hat{x}_{2b} \end{bmatrix}$$

$$T = \begin{bmatrix} \sqrt{\dfrac{\cot(\theta)}{2}} & \sqrt{\dfrac{\tan(\theta)}{2}} \\ \sqrt{\dfrac{\tan(\theta)}{2}} & -\sqrt{\dfrac{\cot(\theta)}{2}} \end{bmatrix} \tag{34}$$

$$\text{if } \hat{z}_1 \text{is available} \Rightarrow \begin{cases} \hat{x}_{1est} = \beta_1 \hat{z}_1 & \beta_1 = \dfrac{E[\hat{x}_{1b}\hat{z}_1]}{E\left[\hat{z}_1^2\right]} \\ \hat{x}_{2est} = \beta_2 \hat{z}_1 & \beta_2 = \dfrac{E[\hat{x}_{2b}\hat{z}_1]}{E\left[\hat{z}_1^2\right]}. \end{cases}$$

A typically high redundancy ($\theta = \pi/10$ in (34)) is used for MDTC. Rate-distortion (RD) curves of the side and central decoder, as well as RRD curves for some typical values of λ, are depicted in Fig. 9 for both the same and different Laplace parameters of x_1 and x_2. Each signal has 2000 samples and the distortions and rates are averaged over 40 independent signals.

As we had seen before, the figures show that while the rate/central distortion of both *CaseI* and *CaseII* is nearly the same, the side distortion in *CaseI* is much lower. Mostly, *CaseI's* side distortion curve is closer to Replication (optimum), while its central distortion is much better than Replication. The best central distortion belongs to Splitting, and MDTC is placed at the next position in this respect.

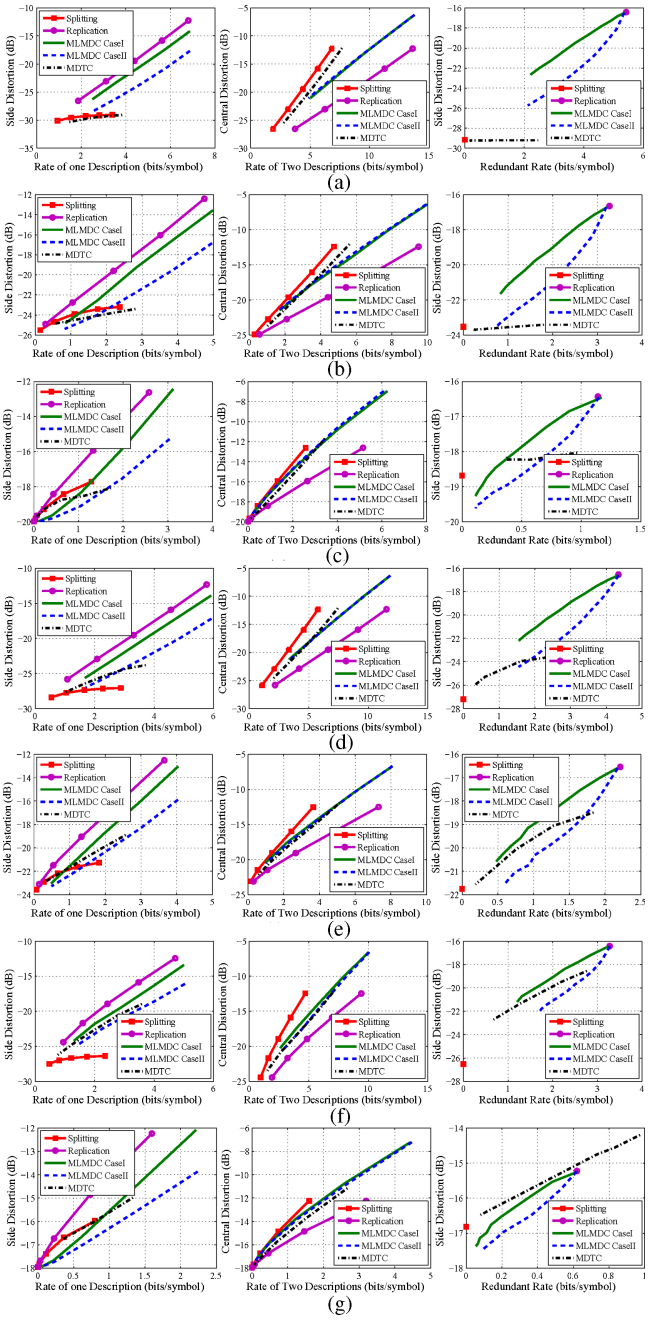


Fig. 9. RD curves for Splitting, Replication, MDTC, and MLMDC scenarios for various distribution parameters (λ_1, λ_2). Left: rate/side-distortion, middle: rate/central-distortion, right: redundancy-rate-distortion with $Q_0 = 16$. (a) $\lambda 1 = 1/20$, $\lambda 2 = 1/20$. (b) $\lambda 1 = 1/10$, $\lambda 2 = 1/10$. (c) $\lambda 1 = 1/5$, $\lambda 2 = 1/5$. (d) $\lambda 1 = 1/20$, $\lambda 2 = 1/10$. (e) $\lambda 1 = 1/10$, $\lambda 2 = 1/5$. (f) $\lambda 1 = 1/20$, $\lambda 2 = 1/5$. (g) $\lambda 1 = 1/5$, $\lambda 2 = 1/2.5$.

In [20], the redundancy rate is consumed for a lower side distortion, while in MLMDC this is used for a lower central distortion. Although both are equivalent, for comparison we must first set a common central distortion and then compare their redundancy/side-distortion curves. For example, at

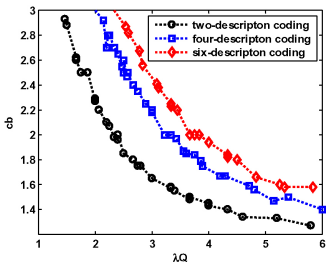


Fig. 10. MLMDC: values of c_b versus values of λQ.

$Q = Q_0$, MDTC achieves a central distortion of $D = D_0$ while the redundancy in MLMDC leads to a central distortion lower than D_0. Therefore, we must find an appropriate $Q_1 > Q_0$ so that the corresponding central distortions are equal and then measure the obtained side distortion with $Q = Q_1$. The optimal RRD curves of MDTC (using non-orthogonal combination matrix) as well as that of Splitting, Replication, and MLMDC are depicted in Fig. 9. For MDTC and MLMDC, there exist parameters that control the amount of redundancy, but for Splitting and Replication, we have zero and a fixed amount of redundancy, respectively. In other words, we have only one RRD point for Splitting and Replication, as shown in the figures. In the RRD curves of MLMDC, the higher redundancies correspond to smaller values of c, due to the fact that when c increases, the appropriate Q_1 at which the central distortion is equal to D_0 increases too, so an increase in Q_1 leads to a decrease in the redundancy rate. Therefore, the tail (right most part) of the RRD curves of MLMDC corresponds to small c parameters, near to the RRD point of Replication which is equivalent to MLMDC with $c = 1$. It can be observed that even though the redundancy of MLMDC is larger, its RRD curve is better than the others particularly for signals with small Laplace parameter. From these simulations, the following analyses are deduced.

1) The performance of *CaseI* is much better than that of *CaseII* in all situations.
2) MDTC was proposed for sources with different distribution parameters. Essentially, it combines small-variance sources with-large variance sources. For signals with the same distribution parameter, the redundancy introduced by the correlating transform has no gain in side distortion. In other words, MDTC for coefficients with the same λ is inefficient. However, for natural image/video sources, it is common that some DCT coefficients have the same distribution parameters. MLMDC outperforms MDTC except for signals with very small distribution parameters. In addition, MLMDC defines a new region in RRD which is not achievable by MDTC.
3) The performance of MLMDC is degraded for large variance signals in a constant quantization step size; first, due to the larger excess rate caused by the combination, as was shown in Fig. 8, and second, due to smaller improvement in the central distortion when EC is superimposed, as was shown in Fig. 7. In other words, the small variance signals have both inferior rate and distortion behaviors in MLMDC. The situation will be better when we have a smaller quantization parameter and hence a smaller ER_n.
4) These simulations are performed for two-description coding. If we use more descriptions, e.g., four-description coding, the excess rate and hence the side distortion per description will be reduced and MLMDC achieves higher efficiency.
5) At low redundancy or equivalently higher values of c, the RRD curve of MLMDC becomes inefficient. For example, the side distortion of MLMDC is getting closer to that of Splitting when c increases, while MLMDC consumes more redundancy rate. This means that there is a value for c, denoted as c_b, after which Splitting is more efficient than MLMDC [seen in Fig. 9(c) and (g)]. Simulations show that this bound is a function of λQ and the number of descriptions. Fig. 10 shows the values of c_b versus λQ, for two-description, four-description, and six-description coding.

A line with a good fitness can be obtained as follows:

$$c_b = \frac{6.9\left(1 + \frac{1}{3}\ln\left(\frac{n}{2}\right)\right)}{\lambda Q + 0.9} \tag{35}$$

where n is the number of descriptions. Therefore, there is a bound for the efficient performance of MLMDC; with given λ and Q. If $c \geq c_b$, the incurred side distortion and redundancy rate is quite high and MLMDC is not better than Splitting in this case.

VI. Experimental Results

We implemented MLMDC in H.264/AVC reference software, JM 16.0, although any other DCT-based codec may be used as well. In H.264/AVC, instead of DCT, a multiply-free integer DCT (IDCT) with a scaling multiplication is used. The scaling multiplication is integrated into the quantizer which leads to fewer multiplications and hence saves computational resources. However, this scaling is compensated in the quantization process. Since in MLMDC combination and estimation is performed in quantized levels, there is no need to worry about them. It is sufficient to calculate and substitute the actual DCT coefficients distribution in the formulations. That is, IDCT coefficients must first be scaled and then their distribution be measured.

Due to poor side distortion behavior, *CaseII* is excluded from the experiments. The whole 300 QCIF frames of *Akiyo*, *Foreman*, and *Mobile* test video sequences, as representatives of low-rate/content, medium-rate/content, and high-rate/content sources, respectively, are used to test the performance of MLMDC. Only 4×4 DCT transform is allowed, 4×4 Intra Mode is enabled for intra frames, and RDO and rate-control are off. This simplifies the algorithm incorporation in JM without loss of generality. GOP size is set to be 10 and CABAC is used for the entropy encoder. The error resiliency tools of H.264/AVC such as B-frames, Flexible MB Ordering, intra MB-refresh, and others are disabled, in order to observe the performance of MDCs in noisy channels more clearly. For the same reason, we have not used any error concealment techniques either.

The DCT coefficients distribution is computed frame-wise and with the given parameter c, the optimal estimation parameters are calculated for each frame. In each QCIF frame, there exist 1584 samples for each DCT coefficient of a 4×4 block—sufficient for finding the distribution parameter (λ). A practical MDC approach is to replicate the larger coefficients and to split the smaller ones. There are several papers which discuss how to optimally find the threshold for distinction between the replicating and splitting coefficients for a given redundancy [13], [14]. However, instead of solving the optimization problem, we consider two cases denoted as Dup2Sp and Dup4Sp where the quantized coefficients larger than 2 and 4, respectively, are replicated and the rest are alternated among the descriptions. Splitting is a special case of the above approach in which all coefficients are split. Intuitively, this has a central quality better than Dup4Sp and a side quality inferior to Dup4Sp. Therefore, its performance against MLMDC can be predicted based on the performance of Dup4Sp and hence we have not performed a direct comparison with Splitting.

To select another existing MDC method to compare against MLMDC, methods which preserve the spatial-temporal resolution of the source were considered [17], [29], [41]. Among these, and based on the discussion in Section II, [29] provides subjectively better side quality compared to [17] and [41] and is chosen for our comparison. The redundancy controlled by the coarse quantization parameter in [29] can be optimized considering channel loss rate and drift. However, for fair comparison among all algorithms we do not perform any optimization and consider two cases for [29] denoted as Q2MDC and Q4MDC. In Q2MDC, QP of the low rate slices, QP_{LR}, is $QP_{HR} + 2$ for two-description coding, and $QP_{HR} + \{2, 4, 6\}$ for four-description coding, where QP_{HR} is the quantization parameter used for the primary (high rate) slices. Similarly, the increment of QP_{LR} in Q4MDC is 4 units. Each frame is partitioned into two(four) slices of MBs and each slice is coded by a quantization parameter determined as explained above, in two(four)-description coding case. These schemes together with Replication and MLMDC are implemented and the comparative rate-distortion curves for QP = $\{20, 24, 28, 32\}$ and multiple values of c are obtained. All MDC scenarios are applied to the luminance component and the resulting luminance texture bits are counted as bit rate in the results.

The experimental one-description performances for *Akiyo*, *Foreman*, and *Mobile* sequences are shown in Figs. 11, 12, and 13, respectively. The figures include side and central decoder RD curves for two-description and four-description coding. In these experiments, each side encoder has its own prediction loop, with the same ones used in the decoder, so reference frames are identical in the encoder and the decoder for all MDC algorithms; i.e., we do not have any reference mismatch and hence no error propagation. We aim to compare the performance of all the methods in reconstructing a frame for which only one description is available, as well as when all descriptions are received intact. The results show that MLMDC, particularly for *Mobile* and *Foreman* with large enough coefficients variance, preserves central performance while achieving one-description quality approximately

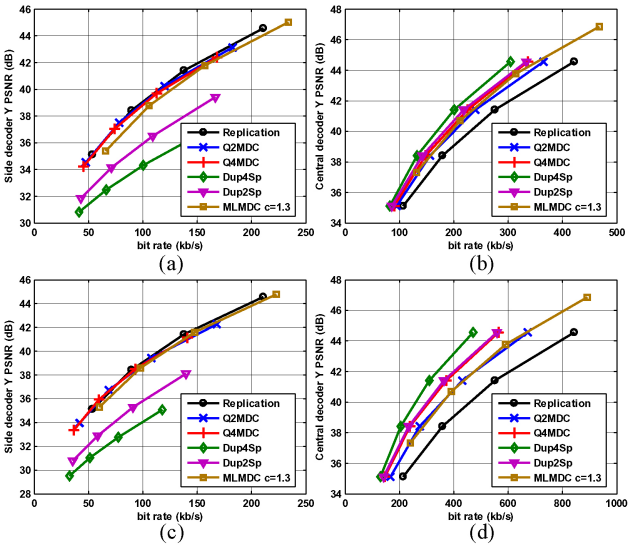


Fig. 11. Experimental rate-distortion curves using MDC scenarios for *Akiyo* sequence. (a) Two-description coding, side decoder. (b) Two-description coding, central decoder. (c) Four-description coding, side decoder. (d) Four-description coding, central decoder.

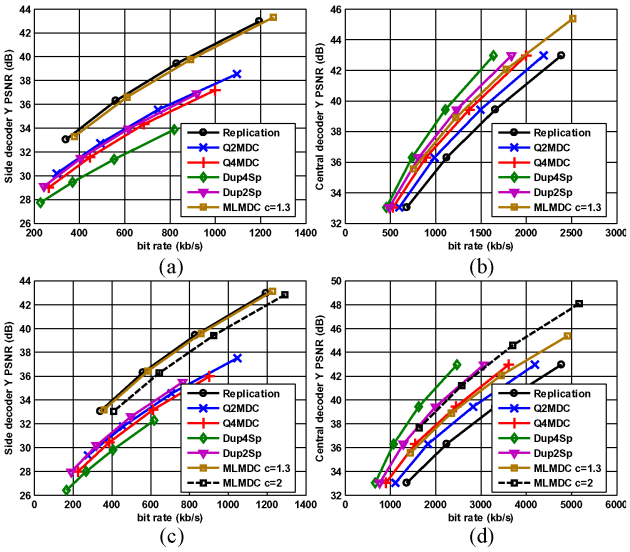


Fig. 12. Experimental rate-distortion curves using MDC scenarios for *Foreman* sequence. (a) Two-description coding, side decoder. (b) Two-description coding, central decoder. (c) Four-description coding, side decoder. (d) Four-description coding, central decoder.

as much as that of Replication. As mentioned before, for four-description coding, both side distortion degradation and rate redundancy per description will be reduced and MLMDC achieves higher efficiency.

Several points are observed from the figures, as previously anticipated and explained in this paper.

1) Except for the *Akiyo* sequence, although the central decoder RD of MLMDC is near or even lower than that of the others, its side quality performance is much better and near Replication (optimal), but with considerably higher central RD curve.
2) *Akyio* is a low motion sequence and the DCT coefficients have a small variance, which leads to a larger ER_n

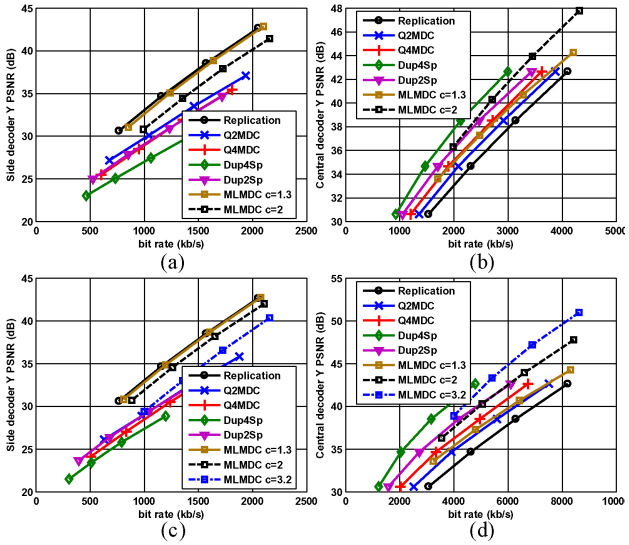


Fig. 13. Experimental rate-distortion curves using MDC scenarios for *Mobile* sequence. (a) Two-description coding, side decoder. (b) Two-description coding, central decoder. (c) Four-description coding, side decoder. (d) Four-description coding, central decoder.

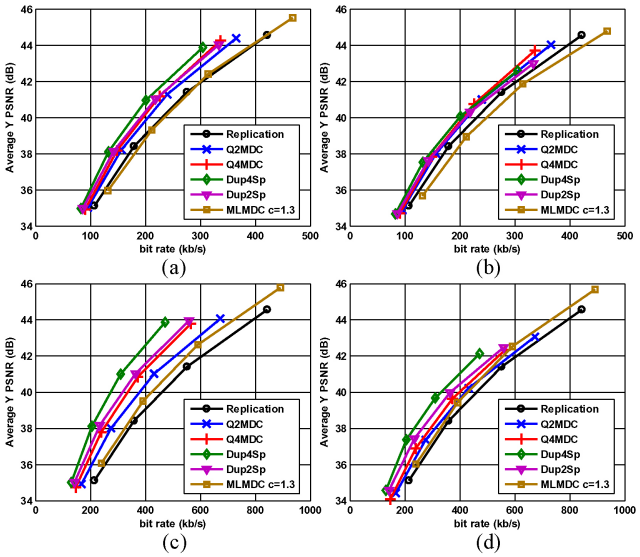


Fig. 14. Experimental rate-distortion curves using MDC scenarios for *Akiyo* sequence on channels with nonzero packet loss rate. (a) Two-description coding and PLR = 5%. (b) Two-description coding and PLR = 20%. (c) Four-description coding and PLR = 5%. (d) Four-description coding and PLR = 20%.

making MLMDC inefficient as discussed already in Section V.

3) For $c = 1.3$, the side distortion for MLMDC is slightly less than Replication. This is due to the combination which virtually moves the rounding offset closer to 1/2, the value which gives the minimum distortion.
4) For the central decoder, the quality improvement in MLMDC is slightly less at low rates compared to high rates. At low rate encoding or with a large quantization parameter, as we had predicted before (at fixed c and λ), the efficiency of MLMDC is degraded with increasing Q.

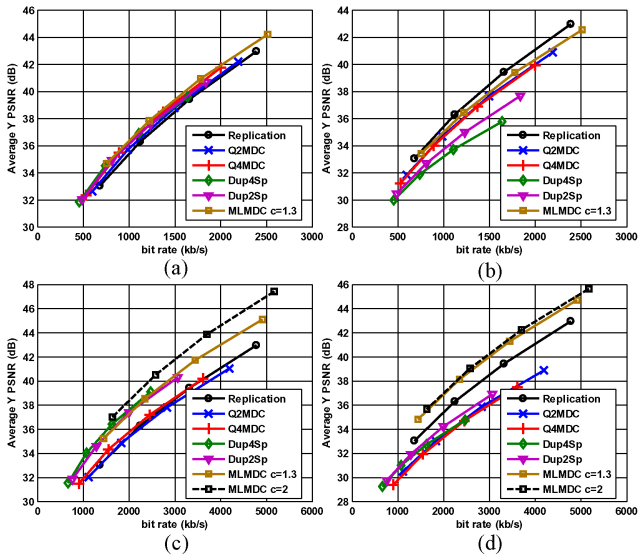


Fig. 15. Experimental rate-distortion curves using MDC scenarios for *Foreman* sequence on channels with nonzero packet loss rate. (a) Two-description coding and PLR = 5%. (b) Two-description coding and PLR = 20%. (c) Four-description coding and PLR = 5%. (d) Four-description coding and PLR = 20%.

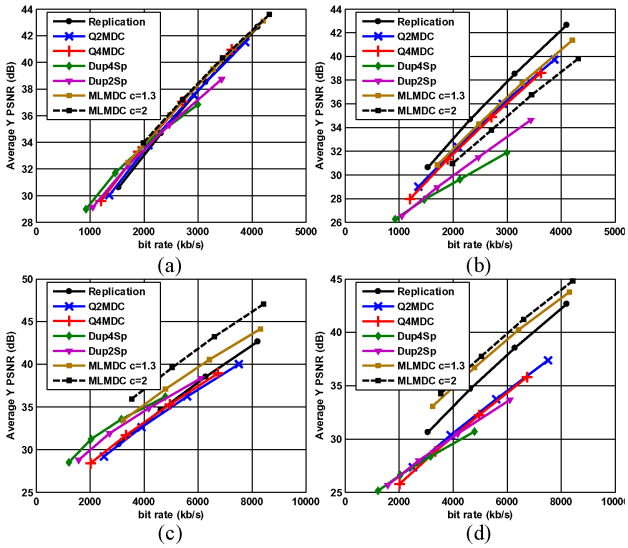


Fig. 16. Experimental rate-distortion curves using MDC scenarios for *Mobile* sequence on channels with nonzero packet loss rate. (a) Two-description coding and PLR = 5%. (b) Two-description coding and PLR = 20%. (c) Four-description coding and PLR = 5%. (d) Four-description coding and PLR = 20%.

The above results show the performance of MLMDC in a lossy scenario where one description is always received correctly while the others are completely lost. Another loss scenario is one where any channel randomly drops packets, which we evaluate next. The packets depending on the channel status might be received intact or completely lost. Each packet, in order to be independently decodable, must contain an integer number of slices (slice is the smallest independently decodable unit) and smaller than the maximum transportation packet (MTU) of the network (to avoid fragmentation). The widely-used and typical size of MTU is 1500 bytes, due to Ethernet limitations, and the fact that the great majority of

end-points are on Ethernet. For MTU size of 1500 bytes, in order to maximize channel throughput, the payload size is set to be 1460 bytes where 40 bytes are reserved for RTP/UDP/IP header information. Each slice contains a group of MBs of the frame in raster scan order and also the header information, and has a maximum size of $1460*8$ bits. In the case that the entire frame as one slice cannot fill the payload of the transport packet, slices are terminated such that an integer number of slices can be encapsulated into a 1460-byte packet and hence each packet, independent of the other packets, will be decodable. The same configuration has been used for all of the evaluated MDC methods so the comparison is fair. It is worth mentioning that in this scenario for Q2MDC and Q4MDC, where we had low-rate and high-rate slices, the quantization parameter is not changed slice-wise. QP is constant for each half (quarter) of the total MBs in a frame for two(four)-description coding just as before, but all MBs with the same QP are not coded as one slice and slices are built based on the fixed packet size scenario explained.

For this scenario in our experiment, all MDC algorithms are implemented in Class-B as defined in [32]; i.e., predictors use the same predictor that would be used by an SDC. Therefore, there will exist a reference mismatch if all of the descriptions are not available. In MLMDC, each CC is inserted into one description and the corresponding coefficient in the other descriptions is BC. Therefore, in the case of four-description coding and with at least two received descriptions, all BCs are available and the reference picture can be perfectly reconstructed, even when coding in Class-B. In the experiments, given the packet loss rate (PLR), 50 random packet loss patterns with Bernoulli distribution are generated and applied on each channel. The channels are assumed to be independent and the average video quality over 300 frames of test sequences is measured. The results are shown in Fig. 14, 15, and 16 for *Akiyo*, *Foreman*, and *Mobile* sequences, respectively.

The following can be observed.

1) For the *Foreman* and *Mobile* sequences, our proposed method is the best except for two-description coding and PLR = 20%, for which MLMDC is the second best after Replication. Note that Replication has poor performance when the loss rate is lower.
2) If a packet containing a slice in one description is lost, that slice is decoded by the side decoder and hence we have a side quality for that slice. Now, the following frames that use the blocks of this slice as the prediction reference cannot be decoded at the central quality even when using all descriptions. This is the reason for the superior performance of Replication in two-description coding and high packet loss rate.
3) The other MDCs are more efficient than MLMDC for *Akiyo* and generally low content videos, as we had discussed before.
4) Comparison of the two-description and four-description coding curves shows that with respect to rate-distortion performance, two-description coding for both values of PLR is more efficient than four-description coding. At high enough loss rates, the performance of four-description coding will be better than that of two-description coding; however, usually the channels are not that lossy. But, there are reasons other than channel loss for using four descriptions or even higher number of descriptions. For example, to cope with bandwidth heterogeneity, MDC with higher number of descriptions provides better bit rate scalability. Furthermore, MDC is usually paired with multipath transmission (MPT) where each description is sent over a separate path. Some features of MPT such as increase in total aggregated bandwidth, decrease in probability of outage and delay variability is improved with higher number of paths and hence with higher number of descriptions (as much as system complexity allows) [42]. With the higher number of paths (descriptions) the required bandwidth per channel is reduced. For example, for having an average quality of 40 dB when PLR = 20% for MLMDC with $c = 2$ and *Mobile* sequence, in the case of two-description coding, each channel needs to have at least 2150 kb/s bandwidth [Fig. 16(b)], while using four-description coding reduces this value to 1500 kb/s [Fig. 16(d)].

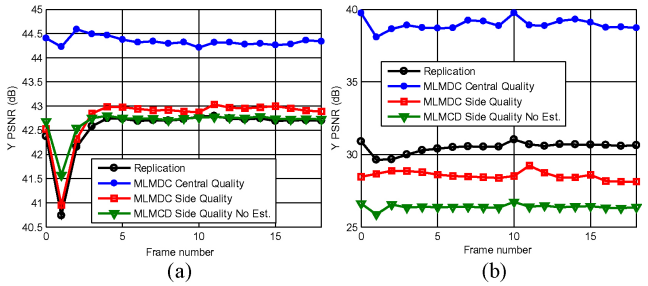


Fig. 17. Effect of estimation for two-description coding, *Mobile* sequence. (a) c = 1.3 and QP = 20. (b) c = 3.2 and QP = 32.

We showed that for small values of c such that (26) is satisfied, the estimation function is approximately the identity function and there is no need for estimation. The curves in Fig. 17 verify this claim; they are the samples of when c is respectively smaller [Fig. 17(a)] and larger [Fig. 17(b)] than the bound given in (26). This figure is a frame-wise description of the experiment in Fig. 13 with the addition of "No Estimation" case. We also see a steep fall in the PSNR of frame #1, which can be explained because the rounding offset f (determined by the adaptive algorithm of JM) has not yet converged to the optimum value for the first P frame. For $QP = 32$, the effect of this non-optimum f is less significant.

Therefore, for small values of c compared to $1/\lambda Q$, the estimator does not change the combined coefficients; meaning that a standard decoder without any modifications can be used for side decoding. However, similar to other MDC schemes, the central decoding routine must be incorporated in standard decoders.

VII. Conclusion

A new drift-free multiple description coding scheme, called MLMDC, was presented with the purpose of increasing the quality of the video compared to the other MDC schemes,

when some packets are lost. In MLMDC, each description is constructed by combining BC and EC coefficients; the latter is controlled by a parameter c. In the central decoder, the combined coefficients are separated and similar to CGS decoding, the coefficients are superimposed. In the side decoder, the base coefficient is estimated by the conditional expected value of all the possible values. We also showed two possible combination methods, where BC and EC can be from one DCT coefficient, denoted as *CaseI,* or from two different DCT coefficients, denoted as *CaseII*. Both had almost the same estimation distortion and redundancy but with a much different side distortion, due to the correlation which exists between the coefficients in *CaseI*. Compared to other MDC schemes, it was observed that our method is more efficient for DCT coefficients with not too large a Laplace distribution parameter. The efficiency of MLMDC was also verified by experimental results, where a good side and central quality was observed, which is not possible with the existing MDC schemes. The improved side performance leads to higher average video quality at the receiver in lossy networks such as the Internet, in particular for four-description coding.

APPENDIX

For a Laplacian source and quantization rule described in (1) for each quantization level, N, we have

$$\begin{aligned} N>0 \quad &\Rightarrow \quad -fQ \le x_e < (1-f)\,Q \quad \& \quad f_N(x_e) \\ &= \frac{\lambda}{2} e^{-\lambda QN} e^{-\lambda x_e} \\ N<0 \quad &\Rightarrow \quad -(1-f)\,Q < x_e \le fQ \quad \& \quad f_N(x_e) \\ &= \frac{\lambda}{2} e^{\lambda QN} e^{\lambda x_e} \\ N=0 \quad &\Rightarrow \quad -(1-f)\,Q < x_e < (1-f)\,Q \quad \& \quad f_0(x_e) \\ &= \frac{\lambda}{2} e^{-\lambda |x_e|} \quad . \end{aligned} \tag{36}$$

Now, to obtain $f^{II}(x_e)$, we must sum up of the quantization errors in all quantization regions

$$\begin{aligned} &fQ \le x_e < (1-f)\,Q \;\Rightarrow \\ &f^{II}(x_e) = \sum_{N=0}^{\infty} f_N(x_e) = \frac{\lambda}{2} e^{-\lambda x_e} \sum_{N=1}^{\infty} e^{-\lambda QN} + \frac{\lambda}{2} e^{-\lambda x_e} \\ &\qquad = \frac{\lambda}{2} e^{-\lambda x_e} \left(\frac{e^{\lambda Q}}{e^{\lambda Q}-1} \right) \end{aligned} \tag{37}$$

and

$$\begin{aligned} &0 < x_e < fQ \;\Rightarrow \\ &f^{II}(x_e) = \sum_{N=-\infty}^{\infty} f_N(x_e) = \frac{\lambda}{2} e^{-\lambda x_e} \sum_{N=1}^{\infty} e^{-\lambda QN} + \\ &\frac{\lambda}{2} e^{-\lambda x_e} + \frac{\lambda}{2} e^{\lambda x_e} \sum_{N=-\infty}^{-1} e^{\lambda QN} = \frac{\lambda}{2} e^{-\lambda x_e} \left(\frac{e^{\lambda Q}}{e^{\lambda Q}-1} \right) + \\ &\frac{\lambda}{2} e^{\lambda x_e} \left(\frac{1}{e^{\lambda Q}-1} \right) = \frac{\lambda}{2} e^{-\lambda x_e} + \lambda \cosh(\lambda x_e) \left(\frac{1}{e^{\lambda Q}-1} \right) \cong \\ &\qquad \frac{\lambda}{2} e^{-\lambda x_e} + \frac{\lambda}{e^{\lambda Q}-1}. \end{aligned} \tag{38}$$

Clearly, function $f^{II}(x_e)$ is symmetric around zero and thus we have

$$f^{II}_{x_e}(x_e) = \begin{cases} \dfrac{\lambda}{2} e^{-\lambda |x_e|} + \dfrac{\lambda}{e^{\lambda Q}-1} & |x_e| \le fQ \\ \dfrac{\lambda}{2} \dfrac{e^{\lambda Q}}{e^{\lambda Q}-1} e^{-\lambda |x_e|} & fQ < |x_e| \le (1-f)\,Q. \end{cases} \tag{39}$$

REFERENCES

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

[2] C. Lee, J. Kim, Y. Altunbasak, and R. M. Mersereau, "Layered coded versus multiple description coded video over error-prone networks," *Signal Process. Image Commun.*, vol. 18, no. 5, pp. 337–356, May 2003.

[3] J. Chakareski, S. Han, and B. Girod, "Layered coding versus multiple descriptions for video streaming over multiple paths," *Multimedia Syst.*, vol. 10, no. 4, pp. 275–285, Jan. 2005.

[4] A. Reibman, Y. Wang, X. Qiu, Z. Jiang, and K. Chawla, "Transmission of multiple description and layered video over an (EGPRS) wireless network," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2000, pp. 136–139.

[5] R. Singh, A. Ortega, L. Perret, and W. Jiang, "Comparison of multiple description coding and layered coding based on network simulations," *Proc. SPIE Conf. Visual Commun. Image Process.*, vol. 3974, pp. 929–939, Apr. 2000.

[6] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[7] R. Bemardini, M. Durigon, R. Rinaldo, L. Celetto, and A. Vitali, "Polyphase spatial subsampling multiple description coding of video streams with H.264," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2004, pp. 3213–3216.

[8] M. Gallant, S. Shirani, and F. Kossentini, "Standard-compliant multiple description video coding," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2001, pp. 946–949.

[9] T. Tillo and G. Olmo, "Data-dependent pre- and postprocessing multiple description coding of images," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1269–1280, May 2007.

[10] J. G. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," *Proc. SPIE Conf. Visual Commun. Image Process.*, vol. 4310, pp. 392–409, Jan. 2001.

[11] C. Zhu and M. Liu, "Multiple description video coding based on hierarchical B pictures," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 4, pp. 511–521, Apr. 2009.

[12] H. Bai, Y. Zhao, and C. Zhu, "Multiple description video coding using adaptive temporal subsampling," in *Proc. Int. Conf. Multimedia Expo*, 2007, pp. 1331–1334.

[13] A. Reibman, H. Jafarkhani, W. Yao, and M. Orchard, "Multiple description video using rate-distortion splitting," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1. Oct. 2001, pp. 978–981.

[14] K. R. Matty and L. P. Kondi, "Balanced multiple description video coding using optimal partitioning of the DCT coefficients," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 7, pp. 928–934, Jul. 2005.

[15] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 821–834, May 1993.

[16] O. Campana, R. Contiero, and G. A. Mian, "An H.264/AVC video coder based on a multiple description scalar quantizer," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 2, pp. 268–272, Feb. 2008.

[17] T. Tillo, E. Baccaglini, and G. Olmo, "Multiple descriptions based on multirate coding for JPEG 2000 and H.264/AVC," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1756–1767, Jul. 2010.

[18] V. Parameswaran, A. Kannur, and B. Li, "Adapting quantization offset in multiple description coding for error resilient video transmission," *Visual Commun. Image Represent.*, vol. 20, no. 7, pp. 491–503, Oct. 2009.

[19] Y. Wang, M. Orchard, and A. Reibman, "Multiple description image coding for noisy channels by pairing transform coefficients," in *Proc. IEEE 1st Workshop Multimedia Signal Process.*, Jun. 1997, pp. 419–424.

[20] M. Orchard, Y. Wang, V. Vaishampayan, and A. Reibman, "Redundancy rate distortion analysis of multiple description image coding using pairwise correlating transforms," in *Proc. Int. Conf. Image Process.*, Oct. 1997, pp. 608–611.
[21] Y. Wang, M. T. Orchard, V. Vaishampayan, and A. R. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Trans. Image Process.*, vol. 10, no. 3, pp. 351–366, Mar. 2001.
[22] Y. Wang, A. R. Reibman, M. T. Orchard, and H. Jafarkhani, "An improvement to multiple description transform coding," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2843–2854, Nov. 2002.
[23] V. K. Goyal and J. Kovacevic, "Generalized multiple description coding with correlating transforms," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2199–2224, Sep. 2001.
[24] M. T. Lu, J. C. Wu, K. J. Peng, P. Huang, J. J. Yao, and H. H. Chen, "Design and evaluation of a P2P IPTV system for heterogeneous networks," *IEEE Trans. Multimedia*, vol. 9, no. 8, pp. 1568–1579, Dec. 2007.
[25] C. W. Hsiao and W. J. Tsai, "Hybrid multiple description coding based on H.264," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 1, pp. 76–87, Jan. 2010.
[26] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 1737–1744, Nov. 1996.
[27] R. Puri and K. Ramchandran, "Multiple description source coding through forward error correction codes," in *Proc. 33rd Asilomar Conf. Signals Syst. Comput.*, Oct. 1999, pp. 342–346.
[28] C.-S. Kim and S.-U. Lee, "Multiple description coding of motion fields for robust video transmission," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 9, pp. 999–1010, Sep. 2001.
[29] T. Tillo, M. Grangetto, and G. Olmo, "Redundant slice optimal allocation for H.264 multiple description coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 59–70, Jan. 2008.
[30] H.-T. Chan, C.-M. Fu, and C.-L. Huang, "A new error resilient video coding using matching pursuit and multiple description coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 1047–1052, Aug. 2005.
[31] D.-M. Chung and Y. Wang, "Multiple description image coding using signal decomposition and reconstruction based on lapped orthogonal transforms," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 6, pp. 895–908, Sep. 1999.
[32] Y. Wang, A. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proc. IEEE*, vol. 93, no. 1, pp. 57–70, Jan. 2005.
[33] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1984.
[34] G. Yovanof and S. Liu, "Statistical analysis of the DCT coefficients and their quantization error," in *Proc. 30th Asilomar Conf. Signals Syst. Comput.*, vol. 1. 1996, pp. 601–605.
[35] E. Lam and J. Goodman, "Modeling DCT coefficients for fast video encoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 4, pp. 608–616, Jun. 1999.
[36] Y. Altunbasak and N. Kamaci, "An analysis of the DCT coefficient distribution with the H.264 video coder," in *Proc. IEEE ICASSP*, vol. 3. 2004, pp. III-177–III-80.
[37] X. Li, N. Oertel, A. Hutter, and A. Kaup, "Laplace distribution based Lagrangian rate distortion optimization for hybrid video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 2, pp. 193–205, Feb. 2009.
[38] G. J. Sullivan, "Efficient scalar quantization of exponential and Laplacian random variables," *IEEE Trans. Inform. Theory*, vol. 42, no. 5, pp. 1365–1374, Sep. 1996.
[39] G. J. Sullivan and S. Sun, "On dead-zone plus uniform threshold scalar quantization," in *Proc. SPIE Visual Commun. Image Process.*, Jul. 2005, pp. 1041–1052.
[40] G. Sullivan, "Adaptive quantization encoding technique using an equal expected-value rule," presented at the 14th Meeting Joint Video Team, JVT-N011, Hong Kong, China, Jan. 2005.
[41] I. Radulovic, P. Frossard, Y. K. Wang, M. M. Hannuksela, and A. Hallapuro, "Multiple description video coding with H.264/AVC redundant pictures," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 1, pp. 144–148, Jan. 2010.
[42] J. G. Apostolopoulos and M. D. Trott, "Path diversity for enhanced media streaming," *IEEE Commun. Mag.*, vol. 42, no. 8, pp. 80–87, Aug. 2004.

Mohammad Kazemi received the B.Sc. degree from the Isfahan University of Technology, Isfahan, Iran, in 2003, and the M.Sc. degree from the Sharif University of Technology (SUT), Tehran, Iran, in 2005, both in electrical engineering. He is currently pursuing the Ph.D. degree from the Department of Electrical Engineering, SUT.

His current research interests include the areas of image/video processing, coding, and transmission, with emphasis on error resilient video transmission over lossy channels.

Khosrow Haj Sadeghi received the B.Sc. degree in electrical engineering from the Sharif University of Technology (SUT), Tehran, Iran, in 1984, and the M.S. and Ph.D. degrees from the University of California, Irvine, in 1987 and 1990, respectively.

He is currently an Assistant Professor with the Department of Electrical Engineering, SUT. His current research interests include signal processing, integrated system-on-a-chip, analog and digital circuits design.

Shervin Shirmohammadi (M'04–SM'04) received the Ph.D. degree in electrical engineering from the University of Ottawa, Ottawa, ON, Canada.

He is currently an Associate Professor with the School of Information Technology and Engineering, University of Ottawa, and an Adjunct Professor with the School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran. He is the Associate Director of the Distributed and Collaborative Virtual Environment Research Laboratory and the Co-Director of the Multimedia Communications Research Laboratory. The results of his research have led to more than 170 publications, over a dozen technology transfers to the private sector, and a number of awards and prizes. His current research interests include multimedia systems and networking, specifically adaptive P2P video streaming, mobile video, 3-D video, video analysis, and gaming and virtual environments.

Dr. Shirmohammadi is an Associate Editor of the *ACM Transactions on Multimedia Computing, Communications, and Applications*, the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, and *Springer's Journal of Multimedia Tools and Applications*. He also chairs or serves on the program committees of a number of conferences in multimedia, virtual environments, and games. He is an IEEE Distinguished Lecturer, a University of Ottawa Gold Medalist, a Licensed Professional Engineer in Ontario, and a Professional Member of the ACM.