Multiple Description Coding With Prediction Compensation

Guoqian Sun, Upul Samarawickrama, *Student Member, IEEE*, Jie Liang, *Member, IEEE*, Chao Tian, *Member, IEEE*, Chengjie Tu, *Member, IEEE*, and Trac D. Tran, *Senior Member, IEEE*

***Abstract*—A new multiple description coding paradigm is proposed by combining the time-domain lapped transform, block level source splitting, linear prediction, and prediction residual encoding. The method provides effective redundancy control and fully utilizes the source correlation. The joint optimization of all system components and the asymptotic performance analysis are presented. Image coding results demonstrate the superior performance of the proposed method, especially at low redundancies.**

***Index Terms*—Image coding, lapped transform, linear estimation, multiple description coding (MDC), rate distortion theory.**

I. Introduction

MULTIPLE DESCRIPTION coding (MDC) [1] is an attractive technique of combating transmission errors, in which several compressed bit streams (descriptions) are generated and can be transmitted via different paths. The descriptions are designed such that the reconstruction quality degrades gracefully when some of them are lost, making MDC suitable for transmitting multimedia signals, where some transmission errors can be tolerated.

A central issue in practical MDC designs is how to introduce a controlled amount of redundancy into the descriptions which the decoder can conveniently utilize, and many methods have been proposed. In [2], a multiple description scalar quantizer (MDSQ) is developed by using a central quantizer and an index assignment, which generates two side quantizers such that each of them alone produces an acceptable side distortion, whereas their combination yields the finer central quantizer. The MDSQ is asymptotically near optimal [3], and has been employed in, e.g., [4] and [5], after the DCT or wavelet transform. However, the MDSQ index assignment is difficult to design and implement, and its redundancy is not easy to adjust.

In [6], a modified MDSQ (MMDSQ) with the same asymptotical performance as the MDSQ is developed, in which two staggered scalar quantizers are used to generate the first layer of each description. Another scalar quantizer is used to further partition the joint bins of the first-layer quantizers, and its output is split into the two descriptions. The MMDSQ avoids the index assignment and can easily adjust the redundancy. It also outperforms other MDSQ-based methods in MD image coding. However, both MDSQ and MMDSQ do not perform well at low redundancy regime, which is a desired property of good MDC schemes [7, pp. 365].

Another family of MDC schemes is based on the source splitting approach pioneered by Jayant in [8] and [9], where a signal is split into even and odd samples, and DPCM is used to encoded each description. If one description is lost, the missing data are predicted from their neighbors in the other description, using the source correlation. However, the prediction errors of the missing data are tied to the source correlation, which cannot be controlled. In [10], DPCM is used before splitting, and the prediction in the DPCM is designed to preserve some source correlations. Therefore, the redundancy between the descriptions can be adjusted to some extent. Although the method reduces the interdescription prediction error, the remaining error still limits the side decoder performance, especially at high rates.

In [11], the transform coefficients are split into two parts. Each part is quantized into one description. To introduce redundancy, each description also includes a coarsely quantized version of the other part, which helps the decoding when the other description is lost. The optimal redundancy rate allocation is studied. A similar approach is developed in [12] using the SPIHT algorithm. Recently, this method is applied to the JPEG 2000 framework in [13] under the name of RD-MDC, in which each JPEG 2000 code-block is encoded at two rates, one in each description, and the rate allocation is determined by Lagrangian optimization. However, to get balanced descriptions and optimal performance, the RD-MDC needs to classify all code-blocks into two subsets, such that any code-block in one subset has similar characteristics to another code-block in the other subset. This procedure is quite time-consuming. In addition, the side distortion of the RD-MDC at low redundancies is not satisfactory, as shown in Section V.

The pairwise correlating transform (PCT) [7] represents another method of introducing redundancy, where a set of 2×2 correlating transforms is applied to the uncorrelated coefficients

Manuscript received February 19, 2007; revised December 03, 2008. First published March 24, 2009; current version published April 10, 2009. This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada under Grants RGPIN312262-05, EQPEQ330976-2006, and STPGP350416-07. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Antonio Ortega.

G. Sun was with Simon Fraser University, Burnaby, BC V5A 1S6 Canada. He is now with Cisco Systems, Inc., Vancouver, BC V7X 1J1 Canada (e-mail: sung99@cisco.com).

U. Samarawickrama and J. Liang are with the Simon Fraser University, Burnaby, BC V5A 1S6 Canada (e-mail: usamaraw@sfu.ca; jiel@sfu.ca).

C. Tian is with AT&T Labs-Research, Florham Park, NJ 07932 USA (e-mail: tian@research.att.com).

C. Tu is with Microsoft Corporation, Redmond, WA 98052 USA (e-mail: chentu@microsoft.com).

T. D. Tran is with the Electrical and Computer Engineering Department, The Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: trac@jhu.edu).

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/TIP.2009.2013068

after the DCT. The outputs of each PCT are split into two descriptions. If one coefficient is lost, it is estimated from its counterpart in the other description.

The PCT framework has some inherent drawbacks. First of all, although the PCT has good low redundancy performance in theory, its practical application could not fully achieve this, because the PCT can only be applied to coefficients with large variances relative to the quantization error [7]. Other coefficients are directly split into the two descriptions. In the side decoder, these low-variance coefficients are simply estimated as zero, which limits the side decoder performance at low redundancies. Second, similar to [8], the PCT does not perform well at high redundancies because of the prediction residual [14]. In this case, the decoded image of the side decoder in [7] can be 2 dB lower than the MDSQ.

Third, the PCT system only uses the correlation it inserts between the two parts of a block, but does not exploit the rich correlation among neighboring blocks. Finally, the practical implementation of the PCT system is not easy. Given L coefficients, the method needs to pair the two coefficients with the kth and $(L-k)$th largest variances, and the optimal PCTs depend on the coefficient variances. To estimate these variances, in [7] all image blocks are classified into four classes. Coefficient variances of each class are then calculated and sorted for PCT designs. In addition, existing entropy coding for single description coding cannot be used for the PCT outputs due to different statistics and block sizes.

A generalized PCT (GPCT) is proposed in [14], which is a hybrid method. At low redundancies, it is simply the PCT. At high redundancies, in addition to the PCT, each description also encodes the prediction residual of the other half of PCT outputs. It is shown in [14] that the GPCT side distortion is 3 dB away from that of the MDSQ and MMDSQ. However, the GPCT only improves the PCT at high redundancies, but does not solve other problems of the PCT. In addition, there has not been any practical application using the GPCT.

As a summary, the methods in [2], [5]–[7], and [11]–[14] offer good redundancy control, but do not fully exploit the source correlation. Although the schemes in [8] and [9] can achieve good performance at low redundancies by utilizing the source correlation, they cannot adjust the redundancy of the MDC system. The algorithm in [10] uses the source correlation to adjust the redundancy, but suffers from the prediction error. Therefore, the goal of this paper is to develop an MDC scheme that can simultaneously achieve the following properties: providing effective control of the redundancy, taking full advantage of the rich correlation of the source in the MD encoding and decoding, and achieving satisfactory performance at all rates and redundancies.

One of the motivations of investigating such a scheme is Shannon's comments on reliable communications [15, pp. 75]: *"Any redundancy in the source will usually help if it is utilized at the receiving point. In particular, If the source already has a certain redundancy ... a sizable fraction of the letters can be received incorrectly and still reconstructed by the context."* Since MDC can be understood as a joint source channel coding problem, one would expect similar effect exists that the inherent dependence in the source can be utilized.

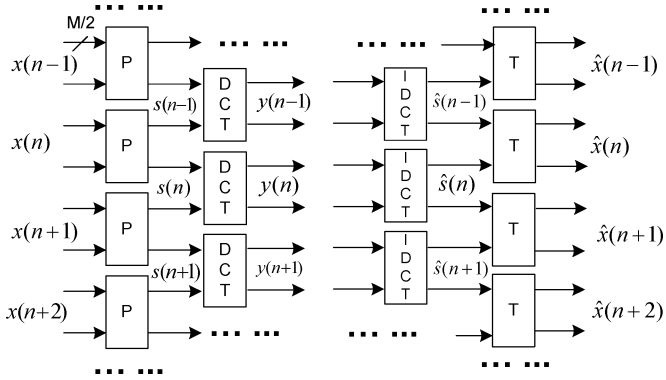


Fig. 1. Forward and inverse time-domain lapped transforms (TDLT).

The potential benefit of utilizing the source correlation in practical MD encoding and decoding suggests a design that deviates from the traditional transform paradigm, for which the error resilient design of the lapped transform [16] is a suitable platform. In [17] and [18], the lapped orthogonal transform output is split at the block level for transmission. The transform is designed to introduce some correlations to help estimating lost blocks at the decoder. In [19], a new family of transforms called the time-domain lapped transform (TDLT) [20] is used, which simplifies the design. The Wiener filter is applied in [21] to further improve the performance. However, these methods sacrifice too much coding efficiency to get good error resilience, and they all suffer from the prediction residual at high rates. In addition, they have to change the transform to adjust the redundancy, which is not convenient in practice.

In this paper, we present a new MDC scheme by combining the time-domain lapped transform, block level source splitting, Wiener filter based prediction, and the encoding of the prediction residual. We formulate the joint optimization of all components in the system, analyze its asymptotic performance, and give various design results. Our scheme resolves the problems of previous methods, and outperforms the MMDSQ, RD-MDC, and PCT in MD image coding, especially at low redundancies. In addition to the performance gain, the proposed method also has lower complexity than many existing methods such as the PCT, GPCT, and RD-MDC.

II. Problem Formulation and Optimal Design

In this section, after a brief introduction of the time-domain lapped transform (TDLT) developed in [20], we present the proposed MDC scheme and discuss its advantages over other methods. We then formulate the joint design of various components of the system and derive the optimal solution.

A. Time-Domain Lapped Transform

Fig. 1 shows the block diagram of the TDLT developed in [20], which is a low-cost extension of the DCT, but with competitive performance compared to JPEG 2000 [22]. At the encoder, an $M \times M$ prefilter $\mathbf{P}$ is employed at the boundary of two blocks (M is the block size). The M-point DCT $\mathbf{C}$ is then applied to each block, creating basis functions that cover two blocks. In the decoder, the inverse DCT and postfilter $\mathbf{T}$ at block boundaries

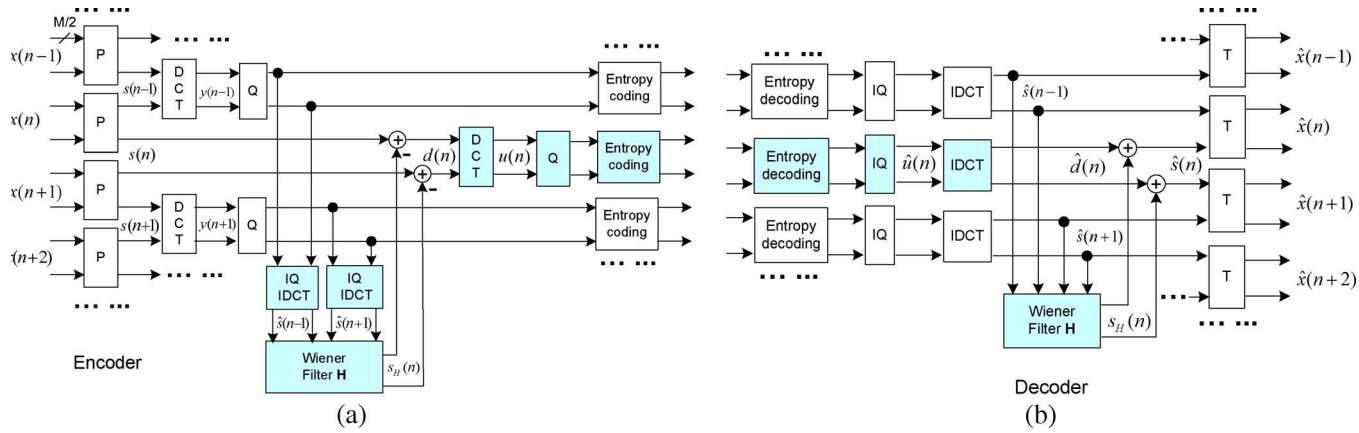


Fig. 2. Block diagrams of encoding (a) and decoding (b) one description in the proposed method.

are applied. Matrices $\mathbf{P}$ and $\mathbf{T}$ have the following structures to yield near-optimal linear-phase lapped transform [20]:

$$\mathbf{P} = \mathbf{W}\mathrm{diag}\{\mathbf{I}, \mathbf{V}\}\mathbf{W} \tag{1}$$

$$\mathbf{T} = \mathbf{P}^{-1} = \mathbf{W}\mathrm{diag}\{\mathbf{I}, \mathbf{V}^{-1}\}\mathbf{W} \tag{2}$$

$$\mathbf{W} = \frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{I} & \mathbf{J} \\ \mathbf{J} & -\mathbf{I} \end{bmatrix} \tag{3}$$

where $\mathbf{I}$ and $\mathbf{J}$ are $M/2 \times M/2$ identity matrix and counteridentity matrix, respectively. Matrix $\mathbf{V}$ is an $M/2 \times M/2$ invertible matrix that can be optimized for different purposes. In this paper, $\mathrm{diag}\{\mathbf{A}, \mathbf{B}\}$ denotes a block diagonal matrix with matrices $\mathbf{A}$ and $\mathbf{B}$ on the diagonal and zeros elsewhere.

Denote $\mathbf{P}_0$ and $\mathbf{P}_1$ as the first and the last $M/2$ rows of the prefilter $\mathbf{P}$, and $\mathbf{T}_0$ and $\mathbf{T}_1$ as the first and the last $M/2$ columns of $\mathbf{T}$, respectively, i.e.,

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_0^T & \mathbf{P}_1^T \end{bmatrix}^T \tag{4}$$

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_0 & \mathbf{T}_1 \end{bmatrix} \tag{5}$$

the $M \times 2M$ forward lapped transform $\mathbf{F}$ and $2M \times M$ inverse transform $\mathbf{G}$ can be written as

$$\mathbf{F} = \mathbf{C}\mathrm{diag}\{\mathbf{P}_1, \mathbf{P}_0\} \tag{6}$$

$$\mathbf{G} = \mathrm{diag}\{\mathbf{T}_1, \mathbf{T}_0\}\mathbf{C}^T = \mathbf{T}_{21}\mathbf{C}^T. \tag{7}$$

B. Overview of the Proposed MDC Scheme and Its Advantages

In this paper, we modify the TDLT framework to generate two descriptions. Fig. 2 illustrates the encoding and decoding of one description by the proposed method. The other description is obtained in a symmetric manner. In this paper, we use $\mathbf{x}(k)$, $\mathbf{s}(k)$, and $\mathbf{y}(k)$ to denote the kth block of prefilter input, DCT input, and DCT output. The reconstruction of a variable is denoted by the hat operator.

In Fig. 2, the prefiltered blocks $\{\mathbf{s}(k)\}$ are split into even-indexed blocks and odd-indexed blocks. After DCT, quantization, and entropy coding, the two groups form the base layers of the two descriptions, respectively. Since the base layer only includes half of the input, each description also contains an enhancement layer to help reconstructing the other half, if one description is lost. To fully exploit the source correlation, we use a Wiener filter $\mathbf{H}$ to predict a missing block using its two reconstructed neighboring blocks, as in [21]. More precisely, $\mathbf{s}(n)$ in Fig. 2 is predicted using the nearest N samples from $\hat{\mathbf{s}}(n-1)$ and N samples from $\hat{\mathbf{s}}(n+1)$, where $1 \le N \le M$. The derivation of the Wiener filter is given in the Appendix. Different from [21], in this paper, the prediction residuals $\{\mathbf{d}(k)\}$ are further DCT-transformed, quantized and entropy-coded to form the enhancement layer of each description.

At the decoder side, if one description is lost, the missing blocks are first estimated from the received base layer blocks by Wiener filter. The decoded enhancement layer blocks are then added to the estimation before postfiltering. When both descriptions are available, only the decoded base layers from the two descriptions are fed to the postfilter for reconstruction. The enhancement layer in each description is simply discarded. Therefore, the enhancement layer is the redundancy introduced by our method, which can be easily controlled by adjusting the quantization step size.

Our scheme enjoys various advantages over existing methods. Compared to the MDSQ/MMDSQ, RD-MDC, PCT/GPCT, it uses Wiener filter to exploit the source correlation between neighboring blocks in MD encoding and decoding. The benefit of utilizing the source correlation can be seen in Section V, where even a simple system with DCT and linear interpolation based prediction can outperform other methods in some cases, especially at low redundancies. Compared with the MDSQ and [8]–[10], our method offers a more flexible control of the redundancy. Our method is also superior to [11]–[13] by using predictive coding, as it is well established that for correlated sources, predictive coding has better rate-distortion performance than direct encoding [23, pp. 113], i.e., it achieves lower distortion at the same bit rate.

The proposed scheme also avoids other limitations of the PCT and GPCT. First, the block level splitting has much better coding efficiency than coefficient level splitting, as the structure within each block is intact. Second, prediction in the spatial domain allows the residual to be coded even at low redundancies, where important image edge information can still survive the excessive quantization after the prediction and the DCT. In contrast, most coefficients in the PCT/GPCT are split directly in this case, which are uncorrelated and cannot be predicted from each

others. Third, as shown in Section III, at high redundancies, the theoretical performance of our method is as good as the GPCT. In practical image coding, our method outperforms not only the PCT/GPCT, but also the MMDSQ and RD-MDC, as our method is well suited for nonstationary signals. Finally, our scheme can be easily optimized and implemented. Only one prefilter and one Wiener filter are needed, and the prediction residuals are processed in the same way as base layer blocks.

C. Joint Optimal Design of the System Components

Let R [in bits/pixel (bpp)] represent the overall bit rate of the two descriptions, i.e., the ratio between the total bits of the two descriptions and the number of input samples. Let R_0 and R_1 denote the average bits for each base layer sample and each enhancement layer sample, respectively. The bit rate for each description is thus $(1/2)(R_0+R_1)$ bpp/description, and the total rate is $R = R_0 + R_1$.

As shown below, given the target bit rate R, the probability p of losing one description, the matrix $\mathbf{V}$ in prefilter $\mathbf{P}$, and the size of the Wiener filter, we can find the closed-form expressions of the corresponding optimal Wiener filter $\mathbf{H}$ and bit allocation R_0 and R_1 that minimize the expected distortion at the receiver. To further find the optimal matrix $\mathbf{V}$ with the minimal expected distortion, an unconstrained numerical optimization program, such as the function *fminunc* in MATLAB, can be used, by treating all entries of $\mathbf{V}$ as unknown variables. Next, we give the derivation of the first step, i.e., the optimal R_0 and R_1 for a given $\mathbf{V}$. The derivation of the corresponding Wiener filter is given in the Appendix.

Let D_0 and D_1 be the central distortion and side distortion, i.e., the mean squared error (MSE) when the decoder receives two and one description, respectively. The expected distortion D is defined as

$$D = (1-p)^2 D_0 + 2p(1-p)D_1. \tag{8}$$

In the proposed MDC scheme, each description contains half base layer blocks and half enhancement layer blocks. Only base layer blocks are used if two descriptions are received. If one of them is lost, half of the input is reconstructed via base layer and the other half is via enhancement layer. Therefore

$$\begin{aligned} D_0 &= D_{B,M} \\ D_1 &= \frac{1}{2}(D_{B,M} + D_{E,M}) \end{aligned} \tag{9}$$

where $D_{B,M}$ and $D_{E,M}$ are the MSE caused by the subband quantization noise in base layer and prediction-compensated enhancement layer, respectively. The subscript M denotes the block size (we use $M = 1$ to represent the DPCM case, as studied in Section III). Substituting into (8), we have

$$\begin{aligned} D &= (1-p)^2 D_{B,M} + p(1-p)(D_{B,M} + D_{E,M}) \\ &= (1-p)D_{B,M} + p(1-p)D_{E,M}. \end{aligned} \tag{10}$$

We first find the optimal expressions of $D_{B,M}$ and $D_{E,M}$ for given R_0 and R_1, under the optimal bit allocation within each block. Since the inverse lapped transform is generally not orthogonal, the reconstruction error $D_{B,M}$ or $D_{E,M}$ is the weighted combination of subband quantization noises, and the weighting parameters are the norms of the inverse transform filters. Let the quantization noise of $\mathbf{y}(k)$ be $\mathbf{q_y}(k)$. After the inverse TDLT, the reconstruction error becomes $\mathbf{G}\mathbf{q_y}(k)$. As usual, we assume the quantization noises of different subbands are uncorrelated. The MSE of the reconstruction is thus

$$D_{B,M} = \frac{1}{M}\sum_{i=0}^{M-1} \|\mathbf{g}_i\|^2 \sigma^2_{q_y(i)} \tag{11}$$

where $\sigma^2_{q_y(i)}$ is the variance of the ith entry of $\mathbf{q_y}(k)$, and $\mathbf{g}_i$ is the ith column of $\mathbf{G}$. Under the assumptions of high rates and i.i.d. sources, $\sigma^2_{q_y(i)}$ can be written as [23, pp. 108]

$$\sigma^2_{q_y(i)} = \epsilon \sigma^2_{y(i)} 2^{-2R_{0i}} \tag{12}$$

where ϵ is a constant that depends on the input statistics and the quantization scheme. R_{0i} is the bits allocated to the ith entry of a base layer block, and $(1/M)\sum_{i=0}^{M-1} R_{0i} = R_0$. $\sigma^2_{y(i)}$ is the variance of the ith entry of $\mathbf{y}(k)$, which is the ith diagonal element of the autocorrelation matrix $\mathbf{R_{yy}}$

$$\mathbf{R_{yy}} = \mathbf{F}\mathbf{R}_{\mathbf{x}_2\mathbf{x}_2}\mathbf{F}^T \tag{13}$$

and $\mathbf{x}_2 = [\mathbf{x}^T(k), \mathbf{x}^T(k+1)]^T$.

Upon optimal bit allocation of R_{0i} [24], the minimal value for (11) is given by

$$D_{B,M} = \epsilon \left(\prod_{i=0}^{M-1} \|\mathbf{g}_i\|^2 \sigma^2_{y(i)} \right)^{\frac{1}{M}} 2^{-2R_0} \triangleq \epsilon \sigma^2_{B,M} 2^{-2R_0}. \tag{14}$$

This is, in fact, the objective function of the single description coding. For block transforms, the minimum value of (14) is achieved by the Karhunen–Loève transform (KLT). For lapped transforms and longer filter banks, there is no closed-form solution, but numerical optimization method can be used to find the solution that minimizes (14).

In single description coding, a performance measure called the *coding gain* is defined based on $\sigma^2_{B,M}$

$$\text{coding gain} = 1/\left(\prod_{i=0}^{M-1} \|\mathbf{g}_i\|^2 \sigma^2_{y(i)} \right)^{\frac{1}{M}} = 1/\sigma^2_{B,M}. \tag{15}$$

When $M = 8$, the coding gain of the DCT and the optimized TDLT in [20] is 8.83 and 9.62 dB, respectively, for first-order Gaussian–Markov inputs with correlation $r = 0.95$.

We now look at $D_{E,M}$, the MSE caused by enhancement layer blocks in the side decoder. It can be seen from Fig. 2 that the prediction residual is

$$\mathbf{d}(n) = \mathbf{s}(n) - \mathbf{s}_H(n) \tag{16}$$

where $\mathbf{s}_H(n)$ is the Wiener filter-based prediction of $\mathbf{s}(n)$ from $\hat{\mathbf{s}}(n-1)$ and $\hat{\mathbf{s}}(n+1)$. The derivation of $\mathbf{s}_H(n)$ for the given prefilter is given in the Appendix. At the decoder, the reconstruction of a predictively coded block $\mathbf{s}(n)$ is

$$\hat{\mathbf{s}}(n) = \hat{\mathbf{d}}(n) + \mathbf{s}_H(n) \tag{17}$$

where $\hat{\mathbf{d}}(n)$ is the reconstruction of $\mathbf{d}(n)$. From (16) and (17), we have the following:

$$\hat{\mathbf{s}}(n) - \mathbf{s}(n) = \hat{\mathbf{d}}(n) - \mathbf{d}(n). \tag{18}$$

In other words, the reconstruction error of $\mathbf{s}(n)$ is equal to that of the prediction residual $\mathbf{d}(n)$. This is, indeed, a property of any differential coding system [23, pp. 113].

As in Fig. 2, let $\mathbf{u}(n)$ be the DCT transform of $\mathbf{d}(n)$, we have

$$\hat{\mathbf{d}}(n) - \mathbf{d}(n) = \mathbf{C}^T \mathbf{q}_{\mathbf{u}}(n) \tag{19}$$

where $\mathbf{q}_{\mathbf{u}}(n)$ is the quantization noise of $\mathbf{u}(n)$. After post-filtering, the reconstruction error becomes $\mathbf{T}_{21}\mathbf{C}^T\mathbf{q}_{\mathbf{u}}(n) = \mathbf{G}\mathbf{q}_{\mathbf{u}}(n)$, and its MSE is

$$D_{E,M} = \frac{1}{M}\sum_{i=0}^{M-1} \epsilon \|\mathbf{g}_i\|^2 \sigma_{u(i)}^2 2^{-2R_{1i}} \tag{20}$$

where R_{1i} is the bits allocated to the ith entry of $\mathbf{u}(n)$. $\sigma_{u(i)}^2$ is the variance of the ith entry of $\mathbf{u}(n)$, given by the ith diagonal element of the autocorrelation matrix $\mathbf{R}_{\mathbf{uu}}$. The expression of $\mathbf{R}_{\mathbf{uu}}$ is given by (44) in the Appendix.

Since (20) has the same format as (11), the derivation from (11) to (14) can be applied here, and the minimal value of $D_{E,M}$ after optimal bit allocation is, therefore

$$D_{E,M} = \epsilon \left(\prod_{i=0}^{M-1} \|\mathbf{g}_i\|^2 \sigma_{u(i)}^2 \right)^{\frac{1}{M}} 2^{-2R_1} \triangleq \epsilon \sigma_{E,M}^2 2^{-2R_1}. \tag{21}$$

The remaining bit allocation issue is to find the optimal R_0 and R_1 for the given $\mathbf{P}$ that minimize the expected distortion D in (10). Substituting (14) and (21) into (10), the problem can be written as

$$\begin{aligned} &\arg \min_{R_0,R_1} (1-p)\epsilon\sigma_{B,M}^2 2^{-2R_0} + p(1-p)\epsilon\sigma_{E,M}^2 2^{-2R_1} \\ &\text{s.t.} \quad R_0 + R_1 = R. \end{aligned} \tag{22}$$

This can be solved using the Lagrangian method, and the optimal bit allocation can be found to be

$$\begin{aligned} R_0 &= \min\left(R, \frac{R}{2} + \frac{1}{4}\log_2 \frac{\sigma_{B,M}^2}{p\sigma_{E,M}^2}\right) \\ R_1 &= \max\left(0, \frac{R}{2} - \frac{1}{4}\log_2 \frac{\sigma_{B,M}^2}{p\sigma_{E,M}^2}\right). \end{aligned} \tag{23}$$

At high rates, i.e., if R_1 is not forced to 0 in (23), substituting this into (14) and (21), we have

$$\begin{aligned} D_{B,M} &= \epsilon\sqrt{p}\sigma_{B,M}\sigma_{E,M}2^{-R} \\ D_{E,M} &= \epsilon\frac{1}{\sqrt{p}}\sigma_{B,M}\sigma_{E,M}2^{-R} = D_{B,M}/p. \end{aligned} \tag{24}$$

In this case, the minimal objective function in (22) becomes

$$D_{min} = 2(1-p)D_{B,M} = 2\epsilon\sqrt{p}(1-p)\sigma_{B,M}\sigma_{E,M}2^{-R}. \tag{25}$$

Finally, plugging (24) into D_0 and D_1 in (9) yields the following distortion product D_0D_1:

$$D_0 D_1 = \frac{1}{2}(1+p)\epsilon^2\sigma_{B,M}^2\sigma_{E,M}^2 2^{-2R} \tag{26}$$

which is further discussed in the DPCM case in Section III.

The following remarks are in order.

Remark 1: Equation (23) shows that more bits should be allocated to the prediction residual when the loss probability p is higher or when $\sigma_{E,M}^2$ is larger (the data are more difficult to predict). Notice that $R_1 = 0$ when $R < (1/2)\log_2(\sigma_{B,M}^2/p\sigma_{E,M}^2)$. In this case, the method reduces to our previous approach in [21]. However, it should be emphasized that this threshold is derived based on the first-order Gauss–Markov model. For nonstationary signals like natural images, we show in Section V that sending prediction residual is beneficial even at very low bit rates and low redundancies, because these bits are spent at regions with strong edges, and can thus significantly improve the reconstruction quality.

Remark 2: Equation (25) shows that the optimal TDLT for the proposed MDC scheme needs to minimize $\sigma_{B,M}\sigma_{E,M}$, whereas the optimal single description transform should minimize $\sigma_{B,M}^2$ in (14). Since $\sigma_{B,M}$ and $\sigma_{E,M}$ are dependent, the optimal TDLT for the proposed MDC is different from the single description case, although the difference is not much, as shown in Section IV.

Remark 3: Equation (25) also shows that when the loss probability p changes, we always need to minimize $\sigma_{B,M}\sigma_{E,M}$. Therefore, the optimal transform is independent of p. This is desired in practice. If R_1 is forced to 0, (25) would be invalid, and the optimal transform would be a function of p. However, our optimization results in Section IV show that the optimal transform is not sensitive to p.

Remark 4: The DCT can be viewed as a special case of the TDLT when the prefilter is $\mathbf{P} = \mathbf{I}$. In this case, the derivation above is still valid. The performance of our method in the DCT case is also studied in Sections IV and V.

III. Asymptotic Performance of the Proposed Method

We now analyze the asymptotic performance of our transform-based method. This is achieved by studying the DPCM case, which has the same performance as transform coding when the block size goes to infinity [24]. In the DPCM case, we split the data into even and odd samples, and use DPCM to encode each group. Each description also predicts the samples in the other group and encodes the residual as the enhancement layer. Since the block size is $M = 1$, no block transform or lapped transform is used.

This special case of our method is indeed an improvement of Jayant's DPCM based method in [8] and [9]. As discussed in Section I, Jayant's method splits the source at sample level and uses linear prediction between the two descriptions, but it does not encode the prediction residual, which is handled by the enhancement layer in our method.

A. Optimization in the DPCM Case

Assume the input follows a first-order Gauss–Markov model with correlation coefficient r. After splitting, each part is still

a first-order Gauss–Markov signal, but with correlation coefficient r^2. Let x_{n-1} and x_{n+1} be two consecutive samples in one description. If the DPCM is used in each description, the optimal prediction of x_{n+1} from $\hat{x}_{n-1}$ is given by $\tilde{x}_{n+1} = r^2\hat{x}_{n-1}$. At high rates, the variance of the residual $e_{n+1} = x_{n+1} - \tilde{x}_{n+1}$ is

$$(1-r^4)\sigma_x^2 + r^4\sigma_q^2 \approx (1-r^4)\sigma_x^2 \triangleq \sigma_{B,1}^2 \tag{27}$$

where σ_q^2 is the quantization noise variance of x_{n-1} and is negligible at high rates.

Similar to (18), the DPCM system also satisfies $\hat{x}_{n+1} - x_{n+1} = \hat{e}_{n+1} - e_{n+1}$ [23]. Therefore, the reconstruction error of $x(n+1)$ after DPCM with bit rate R_0 is

$$D_{B,1} = \epsilon\sigma_{B,1}^2 2^{-2R_0} = \epsilon(1-r^4)\sigma_x^2 2^{-2R_0}. \tag{28}$$

This can be viewed as the counterpart for (14).

When one description is lost, each missing sample x_n is first predicted from $\hat{x}_{n-1}$ and $\hat{x}_{n+1}$, i.e.,

$$\tilde{x}_n = \mathbf{h}[\hat{x}_{n-1}, \hat{x}_{n+1}]^T \triangleq \mathbf{h}\hat{\mathbf{x}}_2 \tag{29}$$

where the optimal solution for $\mathbf{h}$ is the Wiener filter. At high rates, it is approximately [9]

$$\mathbf{h} \approx R_{x_n\mathbf{x}_2}R_{\mathbf{x}_2\mathbf{x}_2}^{-1} = \frac{r}{1+r^2}[1\ 1]. \tag{30}$$

In this case, the variance of the residual $x_n - \tilde{x}_n$ is

$$\sigma_{E,1}^2 = \frac{1-r^2}{1+r^2}\sigma_x^2 + \frac{2r^2}{(1+r^2)^2}\sigma_{q0}^2 \approx \frac{1-r^2}{1+r^2}\sigma_x^2. \tag{31}$$

In Jayant's method, this error exists even at high rates. In our method, the prediction residual is further encoded at rate R_1 in the enhancement layer in each description, leading to a reduced distortion of

$$D_{E,1} = \epsilon\sigma_{E,1}^2 2^{-2R_1} = \epsilon\frac{1-r^2}{1+r^2}\sigma_x^2 2^{-2R_1} \tag{32}$$

which is the counterpart for (21). Since (28) and (32) have the same format as (14) and (21), respectively, the bit allocation solution (23) to (22) can be applied here as well. From (27) and (31), we know that $\sigma_{B,1}^2/\sigma_{E,1}^2 = (1+r^2)^2$, so (23) becomes

$$\begin{aligned} R_0 &= \min\left(R, \frac{1}{2}R + \frac{1}{4}\log_2\left(\frac{(1+r^2)^2}{p}\right)\right) \\ R_1 &= \max\left(0, \frac{1}{2}R - \frac{1}{4}\log_2\left(\frac{(1+r^2)^2}{p}\right)\right). \end{aligned} \tag{33}$$

B. Asymptotic Performance of the Proposed Method

It is shown in [3] (see also [6]) that under the high rate assumption, the product of the side distortion and the central distortion of an MDC scheme for a stationary source satisfies

$$D_0D_1 \geq \frac{1}{4}\epsilon^2 P_x^2 2^{-2R} \tag{34}$$

where P_x is the entropy power of the source [23, pp. 95]. This property of the distortion product has been widely used as a performance measure for MDC.

In the DPCM case of our method, substituting the bit allocation into (28), (32) and (9), we can get

$$D_0D_1 = \frac{1}{2}\epsilon^2\sigma_{B,1}^2\sigma_{E,1}^2 2^{-2R} = \frac{1}{2}\epsilon^2(1+p)(1-r^2)^2\sigma_x^4 2^{-2R}. \tag{35}$$

This corresponds to (26) for $M = 1$. To gain more insight, notice that the variance of the innovation sequence of a first-order Gauss–Markov signal is $\sigma_w^2 = (1-r^2)\sigma_x^2$, which is also the entropy power of the signal, i.e., $P_x = \sigma_w^2$. Therefore, the distortion product in (35) can be written as

$$D_0D_1 = \frac{1}{2}\epsilon^2(1+p)P_x^2 2^{-2R}. \tag{36}$$

Comparing (36) and (34), the performance of the proposed method is away from the theoretical bound by a factor of $2(1+p)$, or about 3 dB for small values of p. This is similar to the performance of the GPCT in [14] at high redundancies. However, it should be noted that this result is for the first-order Gauss–Markov signal, which is stationary. We show in Section V that for nonstationary signals such as natural images, the proposed method performs equally well as other MDC algorithms at high redundancies. For low redundancies, it significantly outperforms other methods, because the block-level prediction compensation is quite suitable for nonstationary signals.

Fig. 3 compares the expected distortion D of our method and Jayant's method for $p = 0.1$. D_0 and D_1 are also included. The input is a first-order Gauss–Markov source with $r = 0.95$. The constant ϵ is chosen as 1. The figure shows that the side SNR of Jayant's method could not be improved at high rates, due to the existence of prediction error (31), which dominates the expected distortion at high rates. This problem is resolved in our method by encoding the prediction error. Thus, both the side SNR and the expected SNR can be improved as the increase of the bit rate. However, this is achieved at the price of reduced central SNR. Fig. 3 also shows that when the bit rate is below about 2.6 bits/sample, no residual is encoded, and our method reduces to Jayant's method. However, as shown in Section V, for natural images, it is beneficial to encode the residuals even at very low bit rates.

IV. Optimization Results and Design Examples

In this section, we show various optimized distortion products and filters for the proposed method. Table I summarizes the optimized distortion product D_0D_1 under different configurations of the transform, block size M, description loss probability p, and Wiener filter size. The input is chosen as a first-order Gauss–Markov signal with $\sigma_x^2 = 1$ and $r = 0.95$. As in Fig. 3, $\epsilon = 1$ is selected. The bit rate is $R = 4$ bpp. The source codes to generate the results in this paper can be downloaded from [25].

Three configurations of the proposed algorithm are compared with the DPCM in Table I. The first one jointly optimizes the prefilter and the Wiener filter to minimize the expected distortion in (10). We denote this method as multiple description lapped transform with prediction compensation (*MDLT-PC*). The second one is denoted as *TDLT-PC*, which uses the best prefilter for single description coding, i.e., by optimizing (14). The last one is denoted as *DCT-PC*, which only uses the DCT,

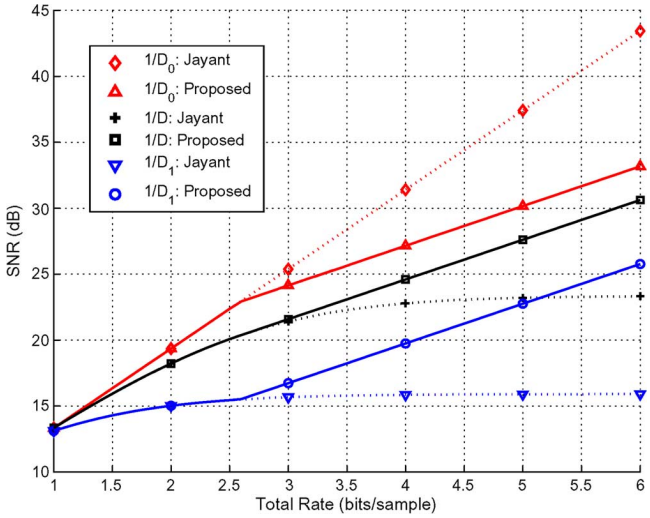


Fig. 3. Comparison of the expected distortion D of the DPCM case of our method and Jayant's method. The central and side distortions are also included for reference ($p = 0.1$, $\sigma_x^2 = 1$, $r = 0.95$, and $\epsilon = 1$).

TABLE I
DISTORTION PRODUCT D_0D_1 ($\times 10^{-5}$) OF DIFFERENT CONFIGURATIONS (WITH $\sigma_x^2 = 1$, $r = 0.95$, $\epsilon = 1$, AND $R = 4$ bpp)

Transform	M	N	p = 0.01	p = 0.05	p = 0.1	p = 0.2
DPCM	1	1	1.88	1.95	2.04	2.23
MDLT-PC	16	16	1.89	1.97	2.06	2.25
	8	8	1.97	2.05	2.14	2.34
	8	2	2.00	2.08	2.18	2.38
	8	1	2.02	2.10	2.20	2.40
	4	4	2.29	2.38	2.49	2.72
TDLT-PC	8	8	2.06	2.14	2.24	2.45
	8	2	2.09	2.17	2.28	2.49
	8	1	2.11	2.19	2.30	2.50
DCT-PC	8	1	2.23	2.32	2.43	2.65

i.e., no prefilter. In TDLT-PC and DCT-PC, Wiener filter and residual encoding are still used.

In Table I, MDLT-PC already approaches the performance of DPCM when $M = 16$. For $M = 8$, the distortion product of the three configurations is about 5%, 10%, and 20% inferior to the DPCM, respectively. Reducing Wiener filter size increases the distortion, but the change is less than 3%. For $M = 8$, even MDLT-PC with $N = 1$ has better performance than TDLT-PC with $N = 8$, showing the advantage of joint optimization. Note also that MDLT-PC with $M = 4$ is worse than DCT-PC with $M = 8$ by about 2.5%. Finally, as shown in [21], Wiener filter with $N = 1$ is optimal for the DCT. In this case, it actually reduces to linear interpolation. Despite its simplicity, we show in Section V that the DCT-PC can still achieve better results than the MMDSQ and RD-MDC in some cases.

We show in Section II that the optimal transform is independent of p when $R_1 > 0$. Simulation results also show that the coding gain of the optimized transform does not change much when R_1 is forced to 0. The optimal result is also not sensitive to R and N. For $M = 8$, when the rate R and the error probability p vary in a large range, the coding gain of the optimized TDLT only changes between 9.41 and 9.61 dB. This makes it possible to fix the transform and still achieves near optimal performance for all practical scenarios.

Two design examples are used in the image coding in the next section. The first one is optimized for $M = 8$, $R = 1$ bpp, $p = 0.2$ and $N = 8$. The coding gain of the result is 9.53 dB, the product D_0D_1 is 0.00164, and the corresponding optimized matrix $\mathbf{V}$ in the TDLT prefilter is given by

$$\mathbf{V} = \begin{bmatrix} 0.8787 & 0.6591 & 0.2426 & 0.1521 \\ -0.5619 & 0.8044 & 0.5009 & 0.1444 \\ 0.1165 & -0.3914 & 0.9813 & 0.2933 \\ -0.0383 & 0.0129 & -0.1641 & 1.0875 \end{bmatrix}. \quad (37)$$

The second example is optimized for $M = 8$, $R = 1$ bpp, $p = 0.2$ and $N = 1$, with a coding gain of 9.54 dB and $D_0D_1 = 0.00167$. The optimized matrix $\mathbf{V}$ is

$$\mathbf{V} = \begin{bmatrix} 0.9424 & 0.6840 & 0.1942 & 0.1046 \\ -0.5389 & 0.8658 & 0.5061 & 0.1037 \\ 0.1126 & -0.3688 & 1.0344 & 0.2778 \\ -0.0402 & -0.0001 & -0.1323 & 1.1183 \end{bmatrix}. \quad (38)$$

In this case, the corresponding 8×2 Wiener filter is simply

$$\begin{bmatrix} 0.67 & 0.63 & 0.59 & 0.54 & 0.46 & 0.41 & 0.37 & 0.33 \\ 0.33 & 0.37 & 0.41 & 0.46 & 0.54 & 0.59 & 0.63 & 0.67 \end{bmatrix}^T.$$

V. PERFORMANCE IN MD IMAGE CODING

In this section, we evaluate the performance of the proposed method in the MD coding of natural images. The source codes and testing parameters for all results can be downloaded from [25]. Six 512×512 standard test images with different characteristics are used. The block size M is chosen to be 8. Two descriptions are generated by partitioning the transformed blocks in a checkerboard pattern. All base layer blocks in each description are grouped together to form a 256×512 subimage, which is then encoded by the embedded entropy coding in [22]. Similarly, all enhancement layer blocks in each description are also grouped together and encoded by the same entropy coding algorithm.

The 1-D Wiener filter is applied to 2-D images in a separable manner. First, each row of a block is Wiener-predicted using the co-located rows in the left and right neighboring blocks. Second, each column of the block is estimated using the co-located columns in the top and bottom neighboring blocks. After that, the average of the row and the column predictions is used as the final prediction of the 2-D block. The 2-D prediction residual is then calculated and encoded in the enhancement layer.

We first study the tradeoff between D_0 and D_1, the central PSNR and the side PSNR. This is related to the distortion product D_0D_1 and has been used as a performance measure in, e.g., [3], [5], [6], and [13]. In our method, this is easily achieved by varying the quantization step sizes of the two layers. In Figs. 4 and 5, our method is compared with the wavelet and Tarp filter-based MMDSQ in [6], and the JPEG 2000-based RD-MDC in [13] (source code at [26]), which represent the state of the art in MDC. The total bit rate is $R = 1$ and $R = 0.25$ bpp, respectively. Four configurations of the proposed method are tested, namely, MDLT-PC with $N = 8$ and $N = 1$ (37), (38),

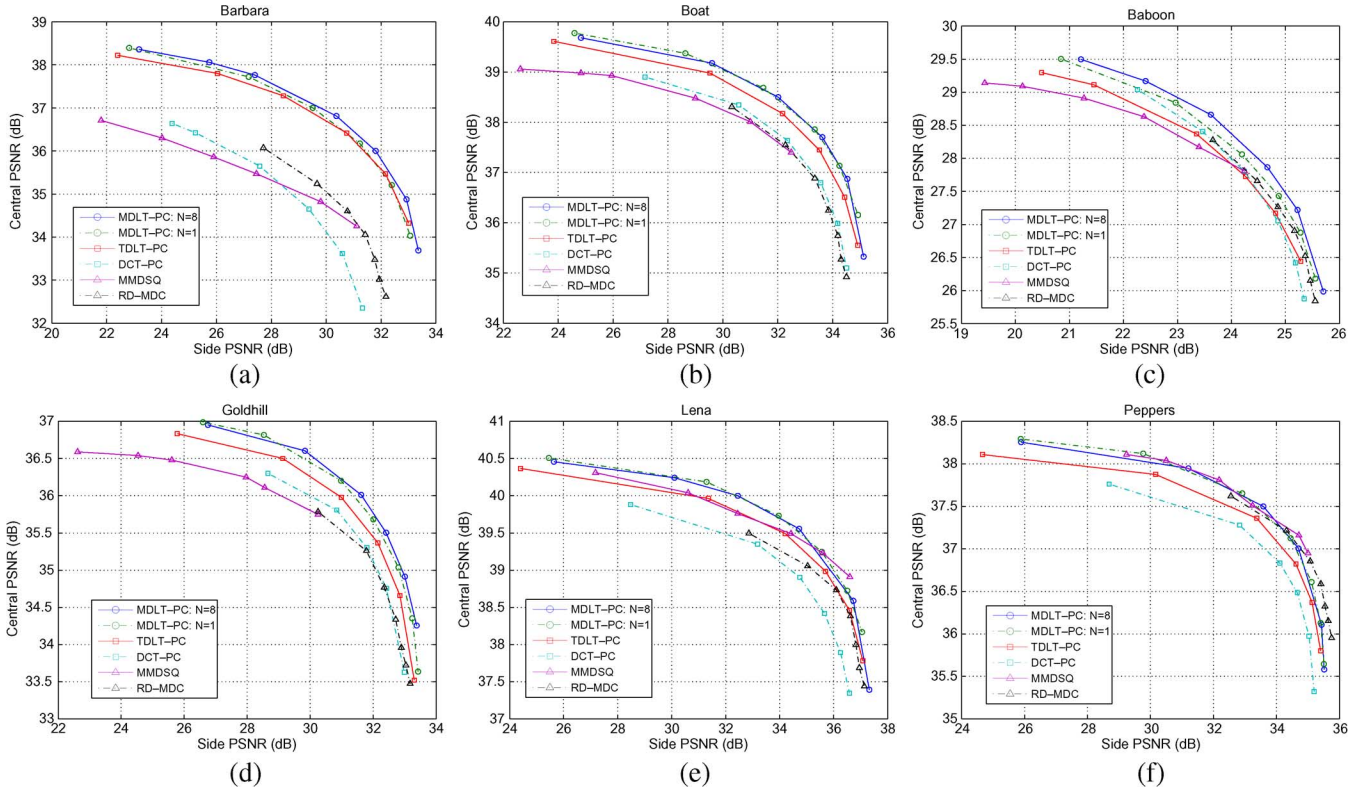


Fig. 4. Performances of MDLT-PC, MMDSQ and RD-MDC at $R = 1$ bit/pixel. (a) Barbara. (b) Boat. (c) Baboon. (d) Goldhill. (e) Lena. (f) Peppers.

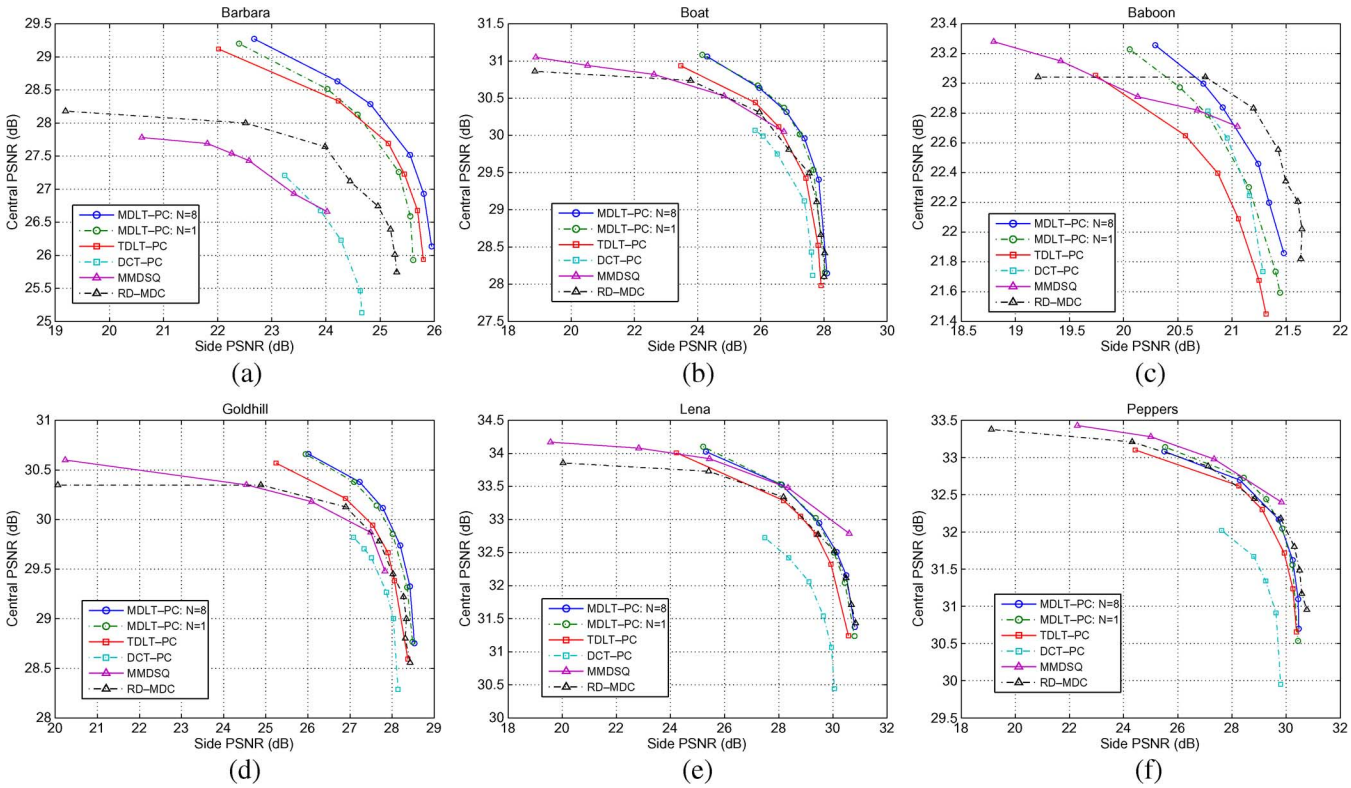


Fig. 5. Performances of MDLT-PC, MMDSQ and RD-MDC at $R = 0.25$ bit/pixel. (a) Barbara. (b) Boat. (c) Baboon. (d) Goldhill. (e) Lena. (f) Peppers.

DLT-PC with $N = 8$ and DCT-PC with $N = 1$. The average of the two side MSEs is used to calculate the side PSNR in these figures.

It can be seen that the performances of the MMDSQ and RD-MDC are quite similar in many cases. The MMDSQ performs better for smooth images like Lena and Peppers, whereas

Fig. 6. Side decoder results with total rate of 1 bpp in (a)–(c) and 0.25 bpp in (d)–(f). The central PSNR isincluded in the parentheses. (a) Barbara by MMDSQ: 24.52 dB (36.09 dB). (b) Barbara by RD-MDC: 27.60 dB (36.07 dB). (c) Barbara by MDLT-PC: 31.68 dB (36.07 dB). (d) Goldhill by MMDSQ: 24.58 dB (30.35 dB). (e) Goldhill by RD-MDC: 24.46 dB (30.35 dB). (f) Goldhill by MDLT-PC: 27.22 dB (30.35 dB).

the RD-MDC is better for images with more textures, such as Barbara.

At the same central PSNR, the side PSNR of our method outperforms the MDSQ and RD-MDC in most cases of Barbara, Boat, Baboon, and Goldhill, many with large margins, especially at low redundancies. The improvement can be more than 8 and 6 dB for $R = 1$ bpp and $R = 0.25$ bpp, respectively. For smooth images like Lena and Peppers, our method still achieves better performance in many low redundancy cases. At high redundancies, the three methods behave similarly, especially for smooth images.

The results also show that the proposed method is not sensitive to the size of the Wiener filter. In most cases, the low-cost MDLT-PC with $N = 1$ has similar performance to the MDLT-PC with $N = 8$, making it a good candidate for practical applications, due to its simpler Wiener filter.

When the single description optimized TDLT is used (TDLT-PC), the curves are lower than the jointly optimized MDLT-PC curves by less than 0.5 dB in all cases. If only the DCT is used (DCT-PC), the curves can be up to 2 dB lower, with more degradations at low rates. These relationships agree with Table I for Gauss–Markov sources. Note that even the simple DCT-PC with linear interpolation can achieve similar or better performance than the MMDSQ and RD-MDC at some low redundancy experiments, demonstrating the benefit of utilizing source correlation in these cases.

The first points of all curves of our method correspond to $R_1 = 0$. In this case, our method reduces to the prediction-only method in [21], and reasonable side PSNR is still obtained. In contrast, the side decoding performance of the MMDSQ with low redundancies is not satisfactory, because most received bits are in the second layer and cannot be used. Similar defect also exists in the RD-MDC at low redundancies, because there is little information about half of the code-blocks.

Fig. 6 shows some decoding results with one description. The three methods are compared at the same total bit rate and same central PSNR. Clearly, our method achieves significant improvement in both the side PSNR and the visual quality.

Finally, we compare our method and the PCT. In Fig. 9 of [7], the D_0 of image Lena is kept at 35.78 dB. At this D_0, the JPEG-based single description coder in [7] needs a rate of 0.60 bpp, whereas the TDLT codec in [7] only needs 0.346 bpp, due to the improved transform and entropy coding. The PCT achieves a side PSNR of 27.94 and 29.63 dB, with redundancy of 15% and 22% over 0.60 bpp, respectively. To get the same D_1, our MDLT-PC with $N = 8$ only needs a redundancy of 9.88% and 16.02% over 0.346 bpp, respectively. In addition, the lowest PCT redundancy in Fig. 9 of [7] is about 10%, with $D_1 \approx 25$ dB. When $R_1 = 0$, our method can achieve a redundancy of 4.38%, and the corresponding D_1 is 25.45 dB. Therefore, our method can achieve the same D_1 with less redundancy than the PCT, and it also reaches a lower redundancy range. The comparison is not very rigorous because of the codec differences, but even if the PCT can be implemented in the TDLT framework, it is still difficult to achieve similar performance to our method due to its aforementioned limitations.

VI. Conclusion

This paper presents an MDC paradigm by integrating time-domain lapped transform, block level splitting, linear prediction and compensation. The method provides effective redundancy control while simultaneously takes advantage of the source correlation. Image coding results show that it outperforms existing methods significantly.

The findings in this paper can be useful for other applications such as MD video coding. Moreover, the proposed method can be further improved. For example, the enhancement layer bit rate can be reduced by refining the entropy coding for the residual, and 2-D prediction and adaptive filter can be applied to reduce the prediction residual, as in [27].

The method can also be generalized to create K descriptions, where $K > 2$. A direct generalization is to split the image into K subimages after prefiltering. Each subimage is coded as the base layer of one description. Each description also encodes the prediction residuals of all other subimages as the enhancement layer. Further improvement of this approach is our ongoing research.

Appendix

In this section, we derive the Wiener filter $\mathbf{H}$ in Fig. 2. In [21], all data in the two neighboring blocks are used to estimate a lost block. Since the performance is not sensitive to the size of the Wiener filter, in this paper we use N $(1 \le N \le M)$ neighboring samples from each side to estimate the lost block, where N can be chosen to obtain a desired tradeoff between complexity and performance.

In Fig. 2, the prediction of $\mathbf{s}(n)$ is $\mathbf{s}_H(n) = \mathbf{H}\hat{\mathbf{s}}_{2,N}$, where $\mathbf{H}$ is the $M \times 2N$ prediction filter, and $\hat{\mathbf{s}}_{2,N}$ is a $2N \times 1$ vector containing $2N$ nearest neighboring samples next to $\mathbf{s}(n)$, N samples from $\hat{\mathbf{s}}(n-1)$ and N samples from $\hat{\mathbf{s}}(n+1)$. That is

$$\hat{\mathbf{s}}_{2,N} = \begin{bmatrix} \hat{\mathbf{s}}_B^T(n-1) & \hat{\mathbf{s}}_B^T(n+1) \end{bmatrix}^T \tag{39}$$

where

$$\begin{aligned} \hat{\mathbf{s}}_B(n-1) &= [\hat{s}_{M-N}(n-1) \quad \dots \quad \hat{s}_{M-1}(n-1)]^T \\ \hat{\mathbf{s}}_B(n+1) &= [\hat{s}_0(n+1) \quad \dots \quad \hat{s}_{N-1}(n+1)]^T. \end{aligned} \tag{40}$$

The autocorrelation of the prediction residual is

$$\mathbf{R}_{\mathbf{dd}} = E\left\{ (\mathbf{H}\hat{\mathbf{s}}_{2,N} - \mathbf{s}(n)) (\mathbf{H}\hat{\mathbf{s}}_{2,N} - \mathbf{s}(n))^T \right\}. \tag{41}$$

The Wiener filter that minimizes the MSE $\mathrm{trace}\{\mathbf{R}_{\mathbf{dd}}\}/M$ is $\mathbf{H} = \mathbf{R}_{\mathbf{s}\hat{\mathbf{s}}_{2,N}} \mathbf{R}^{-1}_{\hat{\mathbf{s}}_{2,N}\hat{\mathbf{s}}_{2,N}}$, where $\mathbf{R}_{\mathbf{s}\hat{\mathbf{s}}_{2,N}}$ is the correlation between $\mathbf{s}(n)$ and $\hat{\mathbf{s}}_{2,N}$, and $\mathbf{R}_{\hat{\mathbf{s}}_{2,N}\hat{\mathbf{s}}_{2,N}}$ is the autocorrelation $\hat{\mathbf{s}}_{2,N}$. At high rate, the quantization noise can be ignored [21], [23, pp. 114], and the Wiener filter can be approximated as

$$\mathbf{H} = \mathbf{R}_{\mathbf{s}\mathbf{s}_{2,N}} \mathbf{R}^{-1}_{\mathbf{s}_{2,N}\mathbf{s}_{2,N}}. \tag{42}$$

The matrices involved can be obtained from the structure of the lapped transform in Fig. 1. Define

$$\begin{aligned} \mathbf{s}_3 &= \begin{bmatrix} \mathbf{s}^T(n-1) & \mathbf{s}^T(n) & \mathbf{s}^T(n+1) \end{bmatrix}^T \\ \mathbf{x}_4 &= \begin{bmatrix} \mathbf{x}^T(n-1) & \mathbf{x}^T(n) & \mathbf{x}^T(n+1) & \mathbf{x}^T(n+2) \end{bmatrix}^T. \end{aligned} \tag{43}$$

As shown in Fig. 1, $\mathbf{s}_3 = \mathbf{P}_{34}\mathbf{x}_4$. Thus, $\mathbf{R}_{\mathbf{s}_3\mathbf{s}_3} = \mathbf{P}_{34}\mathbf{R}_{\mathbf{x}_4\mathbf{x}_4}\mathbf{P}_{34}^T$, where $\mathbf{P}_{34} = \mathrm{diag}\{\mathbf{P}_1, \mathbf{P}, \mathbf{P}, \mathbf{P}_0\}$, and $\mathbf{R}_{\mathbf{x}_4\mathbf{x}_4}$ is the autocorrelation matrix of $\mathbf{x}_4$. In this paper, $\mathbf{R}_{\mathbf{x}_4\mathbf{x}_4}$ is obtained by assuming the input follows a first-order Gauss–Markov model with correlation coefficient $r = 0.95$. Matrices $\mathbf{R}_{\mathbf{s}\mathbf{s}_{2,N}}$ and $\mathbf{R}_{\mathbf{s}_{2,N}\mathbf{s}_{2,N}}$ in (42) can then be obtained from the appropriate sub-matrices of $\mathbf{R}_{\mathbf{s}_3\mathbf{s}_3}$.

From $\mathbf{u}(n) = \mathbf{C}\mathbf{d}(n) = \mathbf{C}(\mathbf{s}(n) - \mathbf{H}\hat{\mathbf{s}}_{2,N})$ and the Wiener filter $\mathbf{H}$ given in (42), we have

$$\mathbf{R}_{\mathbf{uu}} = \mathbf{C}\left\{\mathbf{R}_{\mathbf{ss}} - \mathbf{H}\mathbf{R}_{\mathbf{s}_{2,N}\mathbf{s}}\right\}\mathbf{C}^T. \tag{44}$$

In this paper, this is used in (20) to obtain the distortion of the enhancement layer.

As in [21], we normalize the Wiener filter to have unit row sums. In addition, two $M \times N$ Wiener filters are used at the boundary to predict a block from only the top or the bottom neighboring block. Their derivations are similar to (42).

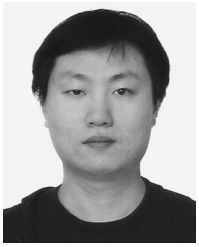
Acknowledgment

The authors would like to thank the anonymous reviewers for their suggestions that have significantly enhanced the presentation of this paper.

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] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 821–834, May 1993.
[3] V. A. Vaishampayan and J.-C. Batllo, "Asymptotic analysis of multiple description quantizers," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 278–284, Jan. 1998.
[4] J.-C. Batllo and V. A. Vaishampayan, "Asymptotic performance of multiple description transform codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 703–707, Mar. 1997.
[5] S. D. Servetto, K. Ramchandran, V. A. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," *IEEE Trans. Image Process.*, vol. 9, no. 5, pp. 813–826, May 2000.
[6] C. Tian and S. S. Hemami, "A new class of multiple description scalar quantizer and its application to image coding," *IEEE Signal Process. Lett.*, vol. 12, no. 4, pp. 329–332, Apr. 2005.
[7] Y. Wang, M. T. Orchard, V. A. 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.
[8] N. S. Jayant, "Subsampling of a DPCM speech channel to provide two self-contained half-rate channels," *Bell Syst. Tech. J.*, vol. 60, pp. 501–509, Apr. 1981.
[9] N. S. Jayant and S. W. Christensen, "Effects of packet losses in waveform coded speech and improvements due to an odd-even sample-interpolation procedure," *IEEE Trans. Commun.*, vol. COM-29, no. 2, pp. 101–109, Feb. 1981.
[10] A. Ingle and V. A. Vaishampayan, "DPCM system design for diversity systems with applications to packetized speech," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 1, pp. 48–58, Jan. 1995.
[11] W. Jiang and A. Ortega, "Multiple description coding via polyphase transform and selective quantization," in *Proc. SPIE Conf. Visual Communication Image Processing*, Feb. 1999, vol. 3653, pp. 998–1008.
[12] A. C. Miguel, A. E. Mohr, and E. A. Riskin, "SPIHT for generalized multiple description coding," in *Proc. IEEE Conf. Image Processing*, Oct. 1999, vol. 3, pp. 842–846.
[13] T. Tillo, M. Grangetto, and G. Olmo, "Multiple description image coding based on lagrangian rate allocation," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 673–683, Mar. 2007.
[14] 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.

[15] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Champaign, IL: Univ. Illinois Press, 1998.
[16] H. S. Malvar, *Signal Processing With Lapped Transforms*. Norwood, MA: Artech House, 1992.
[17] S. S. Hemami, "Reconstruction-optimized lapped transforms for robust image transmission," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 2, pp. 168–181, Apr. 1996.
[18] D. Chung and Y. Wang, "Lapped orthogonal transform designed for error resilient image coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 9, pp. 752–764, Sep. 2002.
[19] C. Tu, T. D. Tran, and J. Liang, "Error resilient pre-/post-filtering for DCT-based block coding systems," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 30–39, Jan. 2006.
[20] T. D. Tran, J. Liang, and C. Tu, "Lapped transform via time-domain pre- and post-processing," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1557–1571, Jun. 2003.
[21] J. Liang, C. Tu, L. Gan, T. D. Tran, and K.-K. Ma, "Wiener filter-based error resilient time domain lapped transform," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 428–441, Feb. 2007.
[22] C. Tu and T. D. Tran, "Context based entropy coding of block transform coefficients for image compression," *IEEE Trans. Image Process.*, vol. 11, no. 11, pp. 1271–1283, Nov. 2002.
[23] D. S. Taubman and M. W. Marcellin, *JPEG 2000: Image Compression Fundamentals, Standards, and Practice*. Boston, MA: Kluwer, 2002.
[24] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
[25] MDLT-PC Source Code [Online]. Available: http://www.ensc.sfu.ca/~jiel/MDLTPC.html
[26] RD-MDC Source Code [Online]. Available: http://www.telematica.polito.it/sas-ipl
[27] J. Liang, X. Li, G. Sun, and T. D. Tran, "Two-dimensional Wiener filters for error resilient time domain lapped transform," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Toulouse, France, May 2006, vol. III, pp. 241–244.



Guoqian Sun received the B.S. and M.S. degrees from Southeast University, China, in 2000 and 2003, respectively. He is currently pursuing the M.A.Sc. degree at the School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada.

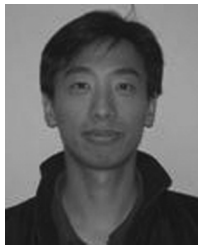
He is currently with the Service Provider Video Technology Group, Cisco Systems, Inc., Vancouver, BC. From 2003 to 2005, he was with ZTE Corporation, Nanjing, China. His research interests include multimedia communications, image/video compressions, and digital communications.

Upul Samarawickrama (S'05) received the B.Sc. degree in electronics and telecommunications engineering from the University of Moratuwa, Sri Lanka, in 2002, and the M.Sc. degree from the University of Manitoba, Canada, in 2005. He is currently pursuing the Ph.D. degree at the School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada.

His research interests include multiple description coding, joint source-channel coding, and digital communications.

Jie Liang (S'99–M'04) received the B.E. and M.E. degrees from Xi'an Jiaotong University, China, in 1992 and 1995, the M.E. degree from National University of Singapore (NUS) in 1998, and the Ph.D. degree from The Johns Hopkins University, Baltimore, MD, in 2003, respectively.

Since May 2004, he has been an Assistant Professor at the School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada. From 2003 to 2004, he was with the Video Codec Group of Microsoft Digital Media Division, Redmond,WA. His research interests include image and video coding, multirate signal processing, and joint source channel-coding.



Chao Tian (S'00–M'05) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2000, and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, in 2003 and 2005, respectively.

He was a postdoctoral researcher at Ecole Polytechnique Federale de Lausanne (EPFL) from 2005 to 2007. He joined AT&T Labs-Research, Florham Park, NJ, in 2007, where he is now a Senior Member of Technical Staff. His research interests include multiuser information theory, joint source-channel coding, quantization design and analysis, as well as image/video coding and processing.

Chengjie Tu (S'02–M'04) received the B.E. and M.E. degrees from University of Science and Technology of China in 1994 and 1997, respectively, and the M.S.E. and Ph.D. degrees from The Johns Hopkins University, Baltimore, MD, in 2002 and 2003, respectively.

Since 2004, he has been with the Video Codec Group, Microsoft Corporation, Redmond, WA. His current research interests include image/video compression, rich media presentation/communication, and multirate signal processing.

Trac D. Tran (S'94–M'98–SM'08) received the B.S. and M.S. degrees from the Massachusetts Institute of Technology, Cambridge, in 1993 and 1994, respectively, and the Ph.D. degree from the University of Wisconsin, Madison, in 1998, all in electrical engineering.

In July 1998, he joined the Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD, where he currently holds the rank of Associate Professor. His research interests are in the field of digital signal processing, particularly in compressed sensing, sampling, multirate systems, filter banks, transforms, wavelets, and their applications in signal analysis, compression, processing, and communications.

Prof. Tran was the Co-Director (with Prof. J. L. Prince) of the 33rd Annual Conference on Information Sciences and Systems (CISS'99), Baltimore, MD, in March 1999. In the summer of 2002, he was an ASEE/ONR Summer Faculty Research Fellow at the Naval Air Warfare Center-Weapons Division (NAWCWD), China Lake, CA. He has served as Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING as well as the IEEE TRANSACTIONS ON IMAGE PROCESSING. He currently serves as a member of the IEEE Technical Committee on Signal Processing Theory and Methods (SPTM TC). He received the NSF CAREER award in 2001 and the William H. Huggins Excellence in Teaching Award from The Johns Hopkins University in 2007.