MULTIPLE DESCRIPTION IMAGE CODING WITH PREDICTION COMPENSATION

Guoqian Sun[1]*, Upul Samarawickrama*[1]*, Jie Liang*[1]*, Chengjie Tu*[2]*, and Trac D. Tran*[3]

[1]School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada, V5A 1S6
[2]Microsoft Corporation, Redmond, WA, USA, 98052
[3]ECE Department, The Johns Hopkins University, Baltimore, MD, USA 21218

ABSTRACT

A new multiple description image coding paradigm is presented in this paper by combining the lapped transform, block level source splitting, inter-description prediction, and coding of the prediction residual. Jointly optimal designs of all system components are discussed. Compared with the best multiple description image coding algorithm in the literature, the new method can achieve significant improvement when one description is lost, given the same bit rate and the same central distortion.

***Index Terms*—** Image coding, Image communication, Information theory, Estimation

1. INTRODUCTION

As an attractive diversity technique for combating transmission errors, the multiple description coding (MDC) [1] generates more than one compressed bit stream (description), which can be transmitted via different paths. Judiciously designed redundancies are introduced in all bit streams such that the reconstruction quality degrades gracefully when some of them are lost.

The rate-distortion bounds for MDC have been established in, for example, [2]. The two main practical mechanisms of approaching these bounds are based on quantization and transform, respectively. The first approach is pioneered by the multiple description scalar quantizer (MDSQ) [3], which is asymptotically near-optimal. However, it requires complicated index assignment. Recently, an elegant two-stage modified MDSQ (MMDSQ) with the same asymptotic performance as the MDSQ for stationary signals is developed [4], in which the first layers of the two descriptions are generated by two uniform scalar quantizers with staggered bins, and another uniform quantizer is used to further partition the joint bins of the two first layer quantizers. The quantization result of the second-stage quantizer is evenly split into two parts to form the second layers of the two descriptions. The application of this method in wavelet-based image coding yields, to the best of our knowledge, the best multiple description image coding performance in the literature.

Various transform-based MDC algorithms have also been studied. In [5, 6], the lapped orthogonal transform is used to add redundancy to each description. The transformed image is split at block level to form multiple descriptions. When some descriptions are lost, the lost blocks are filled by averaging neighboring blocks in [5]. In [6], the missing areas are concealed by imposing a smoothness constraint. In [7], the time domain lapped transform [8] is used, which simplifies the problem formulation. The Wiener filter is further applied in [9] to estimate the lost blocks, leading to significant improvement over [5–7]. However, its performance is still below that of [4].

Another transform based MDC algorithm is the pairwise correlating transform [10], which introduces redundancy between two independent coefficients before splitting them into two descriptions. If one coefficient is lost, it is estimated from its counterpart in the other description. Compared with MDSQ, this method can yield a lower redundancy range, but it has worse performance at high rates [10]. It is shown in [11] that this is caused by the inherent prediction residual of the linear prediction. To resolve this problem, it is proposed in [11] to encode the prediction residual in each description. However, no image coding result is reported in [11].

In this paper, we present a new MDC framework by combining the time domain lapped transform, block level splitting, inter-description prediction, and prediction compensation. The coding of the prediction residual resolves the problem in [5–7, 9] and enables the system to easily achieve different tradeoffs between the central and side distortions. Since the prediction and the coding of the residual operate at the block level, the design and implementation of the method are also simpler than the coefficient-level method in [10, 11]. Image coding results show that at the same bit rate and central distortion, our method can outperform the method in [4] by up to 6 dB if only one description is received. The performance of our method can be further improved, as will be discussed in the end of this paper.

2. PROBLEM FORMULATION AND OPTIMAL DESIGN

In this paper, we only consider MDC with two balanced descriptions. Fig. 1 illustrates the generation of one description by the proposed method. The other description is obtained similarly. The $M \times M$ prefilter $\mathbf{P}$ at block boundaries and the M-point DCT $\mathbf{C}$ generate the time domain lapped transform [8] (M is the block size), whose compression performance is comparable to JPEG 2000 [12]. In what follows, we use $\mathbf{x}(i)$, $\mathbf{s}(i)$, $\mathbf{y}(i)$ and $\mathbf{q}(i)$ to denote the i-th block of prefilter input, DCT input, DCT output and quantization noise, respectively. The prefiltered blocks are split into even-indexed blocks and odd-indexed blocks, which we call the *intra-description blocks* of each description or *intra blocks* for short.

Different from [5–7, 9], each description also encodes the prediction residuals of blocks in the other description. The prediction for each block is obtained by Wiener filtering the two reconstructed neighboring blocks from the other description. In analogy to the temporal prediction-based inter frames in video coding, we call the spatial prediction residuals the *inter-description blocks* or *inter blocks*.

The DCT, quantization and entropy coding are then applied to all

This work is supported by NSERC Discovery Grant, NSERC RTI Grant and SFU President's Research Grant. Author emails: guoqians@sfu.ca, usamaraw@sfu.ca, jiel@sfu.ca, chentu@microsoft.com, trac@jhu.edu.

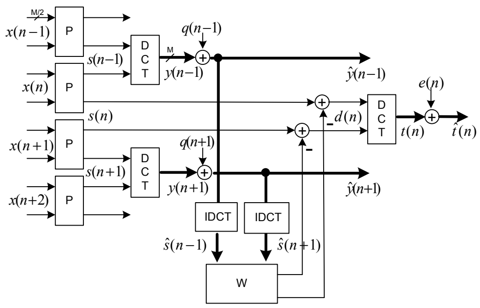


Fig. 1. Block diagram for generating one description.

blocks in each description, but with one quantization step for intra blocks and another step for inter blocks.

In the decoder side, if both descriptions are received, the decoded intra blocks from the two descriptions are combined to obtain the reconstructed signal. The inter block data are discarded. If only one description is received, the missing intra blocks are first predicted from the received intra blocks. The final reconstruction of the missing blocks is the sum of the prediction and the received inter blocks. In other words, the redundant information in our method is used to reduce the side distortion. This is different from the MMDSQ in [4], where the second layer bits reduce the central distortion.

Let $\mathbf{d}(i)$, $\mathbf{t}(i)$ and $\mathbf{e}(i)$ denote the i-th block of DCT input, DCT output and quantization noise of the prediction residual part, as shown in Fig. 1. Define

$$\begin{aligned}\mathbf{s}_2 &= \begin{bmatrix} \mathbf{s}^T(n-1) & \mathbf{s}^T(n+1) \end{bmatrix}^T, \\ \hat{\mathbf{s}}_2 &= \begin{bmatrix} \hat{\mathbf{s}}^T(n-1) & \hat{\mathbf{s}}^T(n+1) \end{bmatrix}^T,\end{aligned} \quad (1)$$

the Wiener filter for $\mathbf{s}(n)$ from the two neighboring blocks can be shown to be [9]

$$\mathbf{W} = \mathbf{R}_{\mathbf{s}(n)\mathbf{s}_2}\mathbf{R}^{-1}_{\mathbf{s}_2\mathbf{s}_2}, \quad (2)$$

where $\mathbf{R}$ denotes the correlation matrix between the two subscript signals. To get (2), we assume the input is an AR(1) signal. As in [9], we ignore the quantization error in $\mathbf{s}_2$.

We assume that each description is either completely lost with a probability of p or perfectly received with probability $1-p$. Our objective is to find the optimal prefilter $\mathbf{P}$ and the optimal bit allocation between intra and inter blocks that minimize the expected distortion. This makes it easier to study the effect of the description loss probability p than the objective functions in [5–7, 9]. We use D_0, D_1, R, R_0 and R_1 to denote the central distortion, the side distortion, the total bit rate of the system, the intra block bit rate and the inter block bit rate, respectively, where $R_0 + R_1 = R$, and R_0 is usually greater than R_1. As in many MDC systems, the expected distortion D is thus defined as

$$D = (1-p)^2 D_0 + 2p(1-p)D_1. \quad (3)$$

Since each description contains intra-coded blocks and inter-coded blocks, we have

$$\begin{aligned}D_0 &= D_{intra}, \\ D_1 &= \frac{1}{2}(D_{intra} + D_{inter}), \\ D &= (1-p)^2 D_{intra} + p(1-p)(D_{intra} + D_{inter}) \\ &= (1-p)D_{intra} + p(1-p)D_{inter},\end{aligned} \quad (4)$$

where D_{intra} and D_{inter} are the average quantization error of the intra-coded blocks and inter-coded blocks, respectively. Under the assumptions of Gaussian sources and optimal bit allocation for the given R_0 and R_1, we have

$$D_{intra} = 2^{-2R_0}\prod_{i=0}^{M-1}(\sigma^2_{y_i}||f_i||^2)^{\frac{1}{M}} \triangleq 2^{-2R_0}\sigma_0^2, \quad (5)$$

where $\sigma^2_{y_i}$ is the variance of the i-th subband of the intra part, $||f_i||^2$ is the norm of the i-th synthesis basis function.

By the property of differential coding, the reconstruction error of $\mathbf{s}(n)$ equals that of $\mathbf{d}(n)$. From this we can get

$$D_{inter} = 2^{-2R_1}\prod_{i=0}^{M-1}(\sigma^2_{t_i}||f_i||^2)^{\frac{1}{M}} \triangleq 2^{-2R_1}\sigma_1^2, \quad (6)$$

where $\sigma^2_{t_i}$ is the variance of the i-th subband of the prediction residual, and can be found from the i-th diagonal element of autocorrelation matrix $\mathbf{R}_{\mathbf{t}(n)\mathbf{t}(n)}$, which is given by

$$\mathbf{R}_{\mathbf{t}(n)\mathbf{t}(n)} = \mathbf{C}\{\mathbf{R}_{\mathbf{s}(n)\mathbf{s}(n)} - \mathbf{W}\mathbf{R}_{\mathbf{s}_2\mathbf{s}(n)}\}\mathbf{C}^T. \quad (7)$$

The optimal values of R_0 and R_1 can be found by defining the Lagrangian cost function

$$L = (1-p)2^{-2R_0}\sigma_0^2 + p(1-p)2^{-2R_1}\sigma_1^2 + \lambda(R_0 + R_1 - R), \quad (8)$$

and the solution is

$$\begin{aligned}R_0 &= \min\left(R, \frac{R}{2} + \frac{1}{4}\log_2\frac{\sigma_0^2}{p\sigma_1^2}\right) \\ R_1 &= \max\left(0, \frac{R}{2} - \frac{1}{4}\log_2\frac{\sigma_0^2}{p\sigma_1^2}\right),\end{aligned} \quad (9)$$

where the min( ) and max( ) operators are to ensure that $R_0 \leq R$ and $R_1 \geq 0$. Eq. (9) shows that more bits are needed for the prediction residual when the description loss probability p is higher or when the correlation in the signal is weaker. Notice that $R_1 = 0$ when $R < \frac{1}{2}\log_2\left(\frac{\sigma_0^2}{p\sigma_1^2}\right)$. This is the threshold below which the prediction compensation will not be useful at all. In this case our method reduces to our previous approach in [9]. It should be noted that the threshold is for stationary Gaussian AR(1) signals. For nonstationary signals like images, we will show later that sending prediction residual is helpful at almost all bit rates.

The optimal bit allocation above is for a given prefilter $\mathbf{P}$. We can then use a Matlab optimization program to find the optimal prefilter $\mathbf{P}$ that minimizes the expected distortion in (3).

It can be shown that as in the method in [11], the asymptotic performance of the proposed method is roughly 3 dB below the MDSQ and MMDSQ for stationary signals. However, the proposed method is especially suitable for nonstationary signals, and it can significantly outperform the MMDSQ, as shown in the next section.

3. DESIGN EXAMPLES AND IMAGE CODING RESULTS

We denote the proposed method as multiple description lapped transform with prediction compensation (*MDLT-PC*) and the prediction-only method in [9] as *MDLT-P*. Fig. 2 compares the relationship between the optimized coding gain and the loss probability p in the two methods. For fair comparison, the curve for the old method is obtained with the objective function (3) and $R_1 = 0$. It can be seen

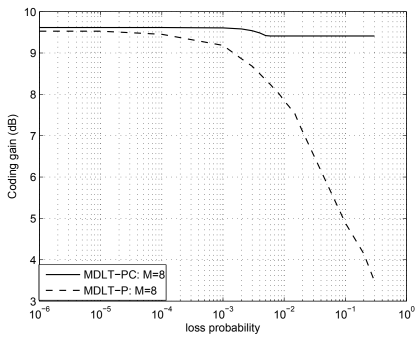


Fig. 2. The relationship between coding gain and loss probability.

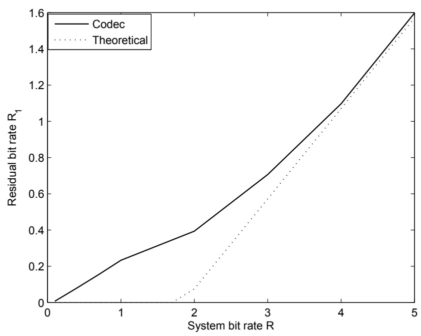


Fig. 3. The theoretical and actual optimal R_1 for image Lena.

that the coding gain of MDLT-P in [9] (as well as [5–7]) drops continuously as the increase of p. In contrast to this, the coding gain of the MDLT-PC only degrades slightly. When the block size is 8, the coding gain of the best lapped transform in [8] is 9.62 dB, whereas the coding gain of our proposed MDC system stabilizes at 9.42 dB when $p > 0.005$. Therefore when the prediction residual is encoded, the multiple description lapped transform is not sensitive to the loss probability, making it possible to fix the transform and still achieve near-optimal performance over a wide range of operating scenarios.

We next demonstrate the performance of the proposed method in the coding of natural images. The block size is selected to be 8, and the two descriptions are generated by partitioning the transformed blocks in a checkerboard pattern. To predict a lost block, the average of the horizontal Wiener prediction and vertical Wiener prediction is used. The embedded entropy coding in [12] is applied to encode the intra part and the inter part of each description independently. This is indeed another advantage of the block level splitting over the coefficient level splitting in [10,11], *i.e.*, existing entropy coding can be applied directly.

We first show that sending prediction residual is helpful even for smooth images and low bit rates. This can be seen from the result for the image Lena in Fig. 3, where the theoretical curve is based on the AR(1)-based bit allocation formulas in (9) with $p = 0.1$, and the actual curve is obtained by varying the quantization steps of the intra and inter parts until the minimal expected distortion is achieved. It can be seen that (9) is only accurate at high rates, and better result can be obtained by sending the prediction error even at very low rates.

Fig 4 compare the proposed method and the MDLT-P in [9] with the images Lena, Boat and Barb. The best design example P21 in [9] is used for the MDLT-P method. The transform for the MDLT-PC method is optimized with the assumption of $p = 0.1$. Although the central PSNR of the MDLT-P method is slightly higher than the proposed method, its side PSNR deteriorates drastically. In fact, the side PSNR of the MDLT-P is almost constant even at high rates, due to the MSE of the Wiener filter. In the new method, the side PSNR improves steadily as the increase of the bit rate. As a result, graceful degradation from the central PSNR to the side PSNR is achieved.

Although the AR(1)-based bit allocation is not accurate at low rates, the optimized transform by the AR(1) model yields better MDC performance than the transform in [8] at all rates. It also outperforms other methods. In Fig. 5, the proposed method is compared with the wavelet-based MMDSQ in [4], which has the best MDC performance in the literature. As in [3, 4] and many other MDC papers, the tradeoff between the central PSNR and the side PSNR is used as the performance measure. In our method, this is achieved by varying the quantization steps of the intra and inter parts. The total bit rate R is fixed at 1 bit/pixel. It can be seen that for most images, our method outperforms the MMDSQ by a large margin. For example, when the central PSNR is 37.5 dB, the side PSNR improvement of our method over the MMDSQ is around 6 dB for the image Barb. For smooth images like Lena, our method only loses to MMDSQ slightly. Fig. 6 shows some examples from Fig. 5 when one description is lost. The two methods are compared at the same bit rate (1bpp) and same central PSNR. Our method achieves an improvement of 5.2 dB and 3.0 dB for Barb and Boat, respectively. The visual quality of our method is also clearly superior.

4. CONCLUSION

This paper presents an improved MDC paradigm by integrating time domain lapped transform, block level splitting, linear prediction and compensation. Image coding results show that it outperforms the quantization based MDC significantly. The proposed framework can be further improved. For example, the side information can be reduced by refining the entropy coding for the inter blocks, and other prediction methods can be applied to reduce the prediction residual.

5. ACKNOWLEDGMENT

The authors are grateful to Dr. Chao Tian for providing the source code of the multiple description coding algorithm in [4].

6. REFERENCES

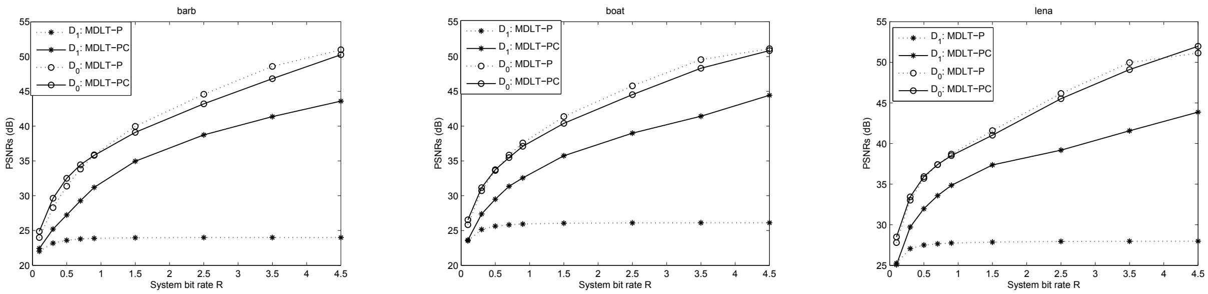
[1] V. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 74–93, Sept. 2001.

[2] R. Zamir, "Gaussian codes and shannon bounds for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2629–2636, Nov. 1999.

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

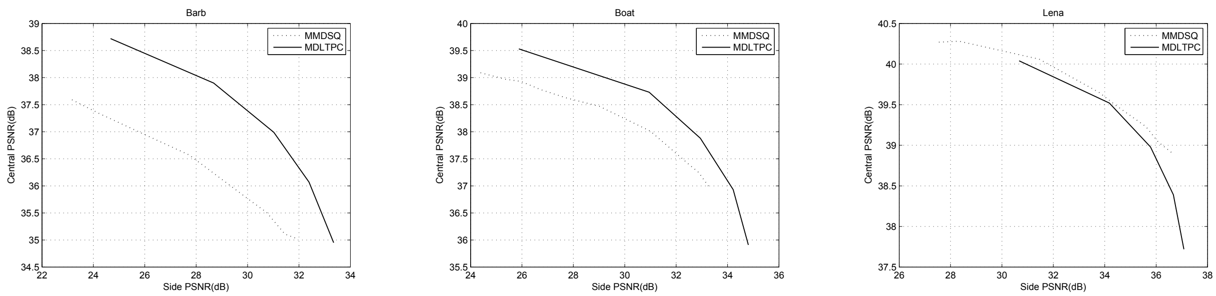
[4] C. Tian and S. Hemami, "A new class of multiple description scalar quantizer and its application to image coding," *IEEE Signal Process. Lett.*, vol. 12, no. 4, pp. 329–332, Apr. 2005.

[5] S. S. Hemami, "Reconstruction-optimized lapped transforms for robust image transmission," *IEEE Trans. Circ. Syst. for Video Tech.*, vol. 6, no. 2, pp. 168–181, Apr. 1996.



(a) (b) (c)

Fig. 4. Performance comparison between MDLT-PC and MDLT-P in [9] for different images. (a) Barb; (b) Boat; (c) Lena.



(a) (b) (c)

Fig. 5. Performances of MDLT-PC and MMDSQ at $R = 1$ bit/pixel. (a) Barb; (b) Boat; (c) Lena.

(a) (b) (c) (d)

Fig. 6. Decoding examples by MMDSQ and TDLT-PC with one description lost. The two methods are compared at the same rate (1bpp) and central PSNR. The side PSNR and the central PSNR of each case are provided. (a) Barb by MMDSQ (25.8 dB / 37 dB); (b) Barb by MDLT-PC (31.0 dB / 37 dB). (c) Boat by MMDSQ (28.8 dB / 38.5 dB); (d) Boat by MDLT-PC (31.8 dB / 38.5 dB).

[6] D. Chung and Y. Wang, "Multiple description image coding using signal decomposition and reconstruction based on lapped orthogonal transforms," *IEEE Trans. Circ. Syst. for Video Tech.*, vol. 9, no. 6, pp. 895–908, Sept. 1999.

[7] C. Tu, T. D. Tran, and J. Liang, "Error resilient pre-/post-filtering for DCT-based block coding systems," *IEEE Trans. Image Proc.*, vol. 15, no. 1, pp. 30–39, Jan. 2006.

[8] T. D. Tran, J. Liang, and C. Tu, "Lapped transform via time-domain pre- and post-processing," *IEEE Trans. Signal Proc.*, vol. 51, no. 6, pp. 1557–1571, Jun. 2003.

[9] 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 Proc.*, to appear, Feb. 2007.

[10] Y. Wang, M. Orchard, V. Vaishampayan, and A. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Trans. Image Proc.*, vol. 10, no. 3, pp. 351–366, Mar. 2001.

[11] Y. Wang, A. Reibman, M. Orchard, and H. Jafarkhani, "An improvement to multiple description transform coding," *IEEE Trans. Signal Proc.*, vol. 50, no. 11, pp. 2843–2854, Nov. 2002.

[12] C. Tu and T. D. Tran, "Context based entropy coding of block transform coefficients for image compression," *IEEE Trans. Image Proc.*, vol. 11, no. 11, pp. 1271–1283, Nov. 2002.