A New Error Resilient Video Coding Using Matching Pursuit and Multiple Description Coding

Hsi-Tzeng Chan, Chih-Ming Fu, and Chung-Lin Huang

***Abstract*—This paper proposes a new error resilient video coding scheme using the matching pursuit (MP) residual coding and the multiple description coding. A conventional discrete cosine transform-based video coding scheme introduces undesirable blocking artifacts at low bit rates. MPs for residual coding has been developed to overcome the disadvantage. The parameters obtained after MP, called atoms, are encoded into two balanced descriptions using scalar quantizers. The two descriptions are respectively transmitted over two separate channels. If both descriptions are received, a high-quality reconstruction can be obtained, however, if either description is lost, a low-quality, but acceptable, reconstruction can be acquired.**

***Index Terms*—Matching pursuits (MP), Multiple description coding (MDC).**

I. Introduction

RECENTLY, there has been considerable interest in video streaming over the Internets. However the encoded videos are extremely sensitive to channel errors. Diversity is commonly used to enhance the reliability of the communication system which transmits information through different channels simultaneously from the source to the users. Therefore, if a channel breaks down, an alternate path may be available. Multiple description coding (MDC) methods [1]–[5] encode a video source into two descriptions for transmission over two different channels.

MDC methods are divided into two main categories: multiple description scalar quantization (MDSQ) [1] and pairwise correlating transform [5]. The former produces two complementary descriptions of the same scalar quantity using two similar ordinary-working quantizers offset from each other, whereas the latter maintains correlation among the transformed coefficients. Various MDC methods for transmitting the encoded video over wireless channels have also been studied [6]. The basic objective of MDC is to degrade the quality of reconstruction gracefully with the loss of any one description.

Current DCT-based video encoding schemes introduce undesirable blocking artifacts at low bit rates. Matching pursuit (MP) residual coding [7] does not cause these artifacts because a over-complete basis set is used to match the residual, and the area with the largest energy value is first encoded. Other MP functionalities, such as peak signal-to-noise (PSNR) scalability and arbitrary shape coding technique with a new search and atom position coding technique, have also been proposed [10]. Various MP modulus quantizer design strategies under target bit rate to improve performance have been developed, such as atom modulus quantization [11], [12], generalized bit-planes [13], and gain-shape vector quantization (VQ) [14]. However, most of the MP video coding methods do not consider the video transmission over the noisy channel.

A robust MP video transmission method using error resilient entropy coding using the Bluetooth air interface standard is developed [15]. Tang *et al.* [16] propose the MP shared atom scheme in which the residual is coded into two sets of atoms. The first L atoms are shared by both sets and the subsequent atoms are alternatively put into the two sets. Tradeoff between the quality of reconstruction from two descriptions and the quality of reconstruction from one description can be controlled by the number of the shared atoms between two atom sets.

Different from the MP shared atom method [16], we propose a more efficient method by encoding the MP atoms into multiple descriptions using MDSQ. The MP atoms are encoded into two balanced descriptions. The two descriptions are transmitted over two separate channels, respectively. If both descriptions are received, a high-quality reconstruction can be obtained through the central decoder. Once only one description is received, a lower-quality, but acceptable, reconstruction can be acquired through the side decoder. Simulation results show that our method outperforms the single description coding (i.e., MP coding and H.263) with forward error correction (FEC) and the MP shared atom method [16] over lossy channels.

II. MP Video Coding

The MP is used to represent a signal with an over-complete basis which is dense for all finite energy function. The set of over-complete basis is called MP dictionary [7] which consists of a general family of time-frequency atoms generated by scaling, translating and modulating a single function. MP provides extremely flexible signal representations since the selection of the dictionaries is arbitrary.

MP can be extended to the discrete two-dimensional (2-D) domain with the proper choice of a basis dictionary. The dictionary set consists of an over-complete collection of 2-D separable Gabor functions of which the one-dimensional (1-D) discrete Gabor function is a scaled, modulated Gaussian function as

$$g_{\vec{\alpha}}(i) = K_{\vec{\alpha}} \cdot g\left(\frac{i - \frac{N}{2} + 1}{s}\right) \cdot \cos\left(\frac{2\pi\xi\left(i - \frac{N}{2} + 1\right)}{16} + \phi\right) \quad (1)$$

where $i \in \{0, 1, \ldots, N-1\}$ and $\vec{\alpha} = (s, \xi, \phi)$ is a triplet consisting of a positive scale, a modulation frequency, and a

Manuscript received May 7, 2003; revised December 1, 2004. This paper was recommended by Associate Editor O. Al-Shaykh. This work was supported by the Department of Energy, Taiwan, R.O.C., under Project 93-EC-17-A-02-S1-032.

The authors are with the Electrical Engineering Department, National Tsing-Hua University, Hsinchu 30010, Taiwan, R.O.C. (e-mail: clhuang@ee.nthu.edu.tw).

Digital Object Identifier 10.1109/TCSVT.2005.852396

phase shift. The constant $K_{\overrightarrow{\alpha}}$ is chosen such that the resulting sequence is of unit norm. Let B be the set of all such triples $\overrightarrow{\alpha}$, then we can specify the 2-D Gabor functions as

$$G_{\overrightarrow{\alpha},\overrightarrow{\beta}}(i,j) = g_{\overrightarrow{\alpha}}(i)g_{\overrightarrow{\beta}}(j)$$
$$i,j \in \{0,1,\ldots,N-1\}, \quad \overrightarrow{\alpha},\overrightarrow{\beta} \in B. \quad (2)$$

In practice, a finite set of 1-D basis functions is chosen and all separable products of these 1-D functions are allowed to exist in the 2-D dictionary, which forms an over-complete basis set for the residual image. This basis set consists of all of the shape in the dictionary placed at all possible integer pixel locations in the residual image. Since this element can be placed at any integer pixel location, it alone forms the standard basis for the residual image.

Since the residual image is sparse, we "prescan" the residual image for high-energy pockets. The location of such pockets can be used as an initial estimate for the inner-product search. The motion residual image is first divided into blocks, and the energy of each block is computed. The center of the block with the largest energy value is selected as an initial estimate for the inner product search. The largest inner product, along with the corresponding dictionary structure and image locations, forms a set of five parameters (i.e., $\vec{\alpha}, \vec{\beta}, x, y$, and p), which define an atom. The $\vec{\alpha}$ and $\vec{\beta}$ indicate the best match structure element from the dictionary. The x and y denote the location of best match in residual image. The p is the value of largest inner product which is the projection of image data at (x,y) onto $G_{\vec{\alpha},\vec{\beta}}(i,j)$.

III. Multiple Description Scalar Quantizers (MDSQ)

In MDSQ systems [1]–[3], a source sample x is mapped by the quantizer to the central reconstruction $\hat{x}^0$ and the side reconstructions $\hat{x}^1$ and $\hat{x}^2$ from the codebooks $\hat{X}^0, \hat{X}^1, \hat{X}^2 \in \boldsymbol{C}$, where $\hat{X}^0 = \{\hat{x}^0_{\mathrm{ij}}, (i,j) \in H\}, \hat{X}^1 = \{\hat{x}^1_i, i \in I\}, \hat{X}^2 = \{\hat{x}^2_j, j \in J\}, I = \{1,2,\ldots,K\}, J = \{1,2,\ldots,K\}, H \subseteq I \times J$, and the output of scalar quantizer is an index l. The index assignment $s(\cdot)$ maps l to a pair of indexes $(i,j) \in H$. The indexes i and j are separately transmitted over the two channels. If both are received, the central decoder will reconstruct the source sample. If only one (i or j) is received, the side decoder may reconstruct the source sample. The correlation between the two descriptions arises from the mapping of the quantization index by the index assignment $s(\cdot) : l \rightarrow (i,j) \in H$.

The encoder of the MDSQ can be regarded as the partition $\mathbf{P} = \{P_{ij}, (i,j) \in H\}$. where $P_{ij} = \{x : e_1(x) = i, e_2(x) = j\}$ are called the central cells, and the mapping $e_1 : \boldsymbol{C} \rightarrow I$ and $e_2 : \boldsymbol{C} \rightarrow J$, where $\boldsymbol{C}$ is the set of the reconstruction level of the codebook. The encoder produces two indexes i and j. The three decoders of the MDSQ can be represented by the mapping $\boldsymbol{r}_0 : H \rightarrow \boldsymbol{C}$ (central decoder), $\boldsymbol{r}_1 : I \rightarrow \boldsymbol{C}$ and $\boldsymbol{r}_2 : J \rightarrow \boldsymbol{C}$ (side decoders), and the output of the three decoders are the reconstruction levels with indexes i, j and (i,j) from the codebooks $\hat{X}^1, \hat{X}^2$, and $\hat{X}^0$.

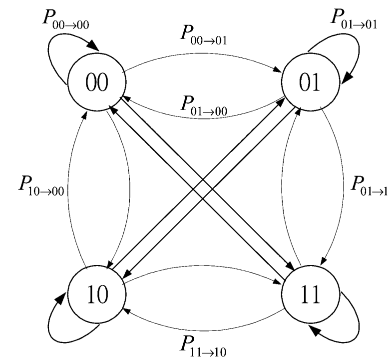


Fig. 1. Four-state channel model [9].

Here, we consider the balanced descriptions, which create an identical average distortion when either description is received. The MDSQ can be treated as a constrained optimization problem [1] of which the Lagrangian function is defined as

$$L(\mathbf{P},\mathbf{r},\lambda_1,\lambda_2) = E(d_0(X,\hat{X}^0))$$
$$+\lambda_1(E(d_1(X,\hat{X}^1)) - D_1) + \lambda_2(E(d_2(X,\hat{X}^2)) - D_2) \quad (3)$$

We use an iterative descent algorithm to determine an optimal $(\mathbf{P}', \mathbf{r}')$ for given $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$. To derive conditions for the encoder partition $\mathbf{P}$ that minimizes $L(\mathbf{P},\mathbf{r},\lambda_1,\lambda_2)$ for given $\mathbf{r}, \lambda_1 \geq 0$ and $\lambda_2 \geq 0$, we can rewrite the Lagrangian function in integral form as

$$L(\mathbf{P},\mathbf{r},\lambda_1,\lambda_2) = \int_{-\infty}^{\infty} G(x,\lambda_1,\lambda_2)p(x)dx \quad (4)$$

where $G(x,\lambda_1,\lambda_2) = d_0(x,\hat{x}^0_{\mathrm{ij}}) + \lambda_1(d_1(x,\hat{x}^1_i) - D_1) + \lambda_2(d_2(x,\hat{x}^2_j) - D_2)$. To minimize the integral, one should map each sample of x to an index pair (i,j) so that $G(x,\lambda_1,\lambda_2)$ is minimized.

Next, we may derive conditions for the decoder $\mathbf{r}$ that minimizes $L(\mathbf{P},\mathbf{r},\lambda_1,\lambda_2)$ for given $\mathbf{P}$, $\lambda_1 \geq 0$, and $\lambda_2 \geq 0$. The optimum decoders are given by

$$\mathbf{r}_0(i) = \arg\min_{t \in C} E(d_0(X,t)|(i,j)), \quad (i,j) \in H \quad (5)$$
$$\mathbf{r}_1(i) = \arg\min_{t \in C} E(d_1(X,t)|i), \quad i \in I \quad (6)$$
$$\mathbf{r}_2(i) = \arg\min_{t \in C} E(d_2(X,t)|j), \quad j \in J \quad (7)$$

where $\boldsymbol{C}$ is the set of the reconstruction level of the codebook. The crucial step in MDSQ is determining the central partition. The determination of the central partition can be viewed as a problem of determining the extreme points of a convex set.

IV. Error Resilient Coding Scheme

Different from the shared atom scheme [16], the MP_MDC scheme applies the MDSQ to encode the MP atom parameters into two descriptors. To evaluate the error-resilient capability, we simulate the video transmission over the lossy channels using two different channel models [8], [9].

TABLE I
NUMBER OF BITS FOR ENCODING ATOM PARAMETERS

Atom parameters	The number of bits for original atom parameters		Two balanced descriptions (i, j) after MDSQ (bits)		Overhead of MDSQ approach
Largest inner product	8		(4, 4)	(4, 4)	0
Dictionary indices	$\bar{\alpha}$: 5 $\bar{\beta}$: 5	subtotal 10	(3, 3) (3, 3)	subtotal 12	2
Locations	x : 8 y : 8	subtotal 16	(7, 7) (7, 7)	subtotal 28	12
Total bits	34		48		14 (41.17%)

A. Channel Models

When data packets are delivered over the Internet, usually a packet is either received correctly or lost caused by network congestion. The channel model for transmitting the MD can be either the two-state Markov model or the four-state channel model.

Two-State Markov Model [8]: In channel model, the loss process is modeled as a discrete-time Markov chain with two states. The current state of the stochastic process depends only on the previous state. The model has two states, good state (0) and bad state (1). A packet is received correctly when the channel is in the good state and errors occur when the channel is in the bad state. P_{01} and P_{10} are the state transition probabilities from good to bad state and from bad to good state, respectively. The transition probabilities can be calculated from the channel characteristics, such as average-burst-error-length $(\mathrm{ABEL}) = 1/P_{10}$ and average packet error rate $(\mathrm{PER}) = P_{01}/(P_{01}+P_{10})$.

Four-State Channel Model [9]: For MD video transmission over two-path channels, there are three different types of links which correspond to three subpaths: 1) joint link along path-1 and path-2; 2) disjoint link along path-1; and 3) disjoint link along path-2. We model a two-path diversity system through three subpaths as either the joint or disjoint portions of each path. For single description (SD), distortion depends critically on the burst loss length. Unlike SD, distortion of MD depends critically on whether loss afflicts both descriptions at the same time, rather than the burst loss length on any SD. Therefore, an appropriate model for a MD source over a joint link is the 4-state channel model (shown in Fig. 1). There are four states which express whether both descriptions are correctly received (state 00), only one description is correctly received and the other description is lost (state 01 and state 10), or both descriptions are lost (state 11).

B. MP_MDC *Scheme*

The MP video coding performance will saturate after some atoms have been encoded [7]. Therefore, the MP_MDC scheme encodes 50 atoms into two sets of descriptions by using MDSQ to ensure the baseline video quality. To obtain the optimal encoder and decoder for each atom parameter, we need to collect the statistics of each atom parameter. Based on the statistics, we can determine a central partition for each atom parameter, use (5)–(7) to obtain the optimal decoder, and then calculate the Lagrangian function L_1 using (4). We can determine another

TABLE II
INDEX ASSIGNMENT TABLES FOR (a) $\{4,3\}$, (b) $\{5,2\}$, AND (c) $\{6,1\}$

(a)

1	3						
2	4						
		5	6				
		7	8				
				9	11		
				10	12		
						13	14
						15	16

(b)

1	3		
2	4		
		5	6
		7	8

(c)

1	3
2	4

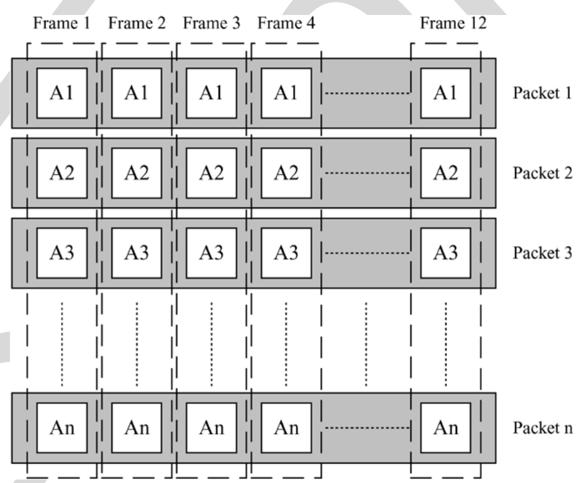
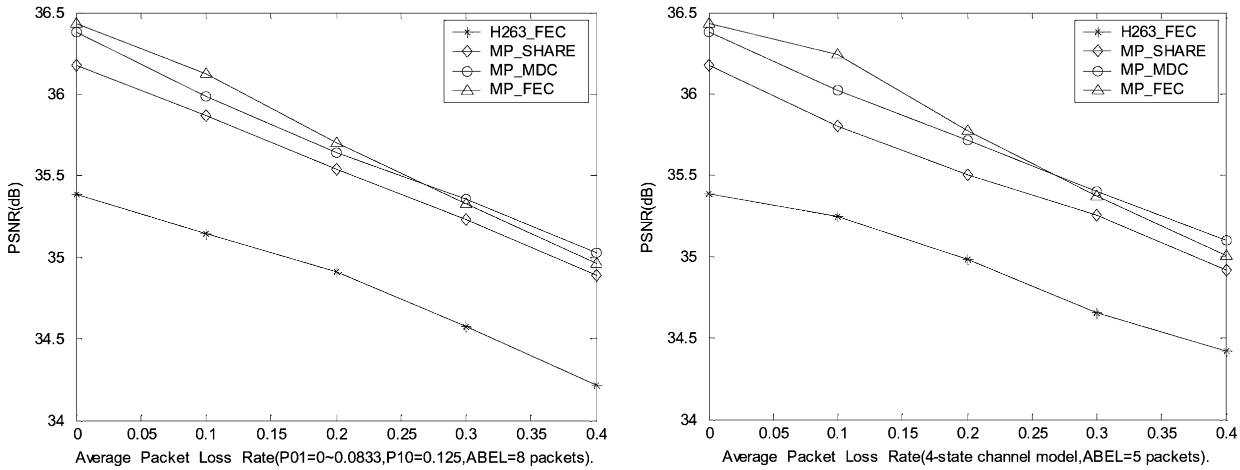


Fig. 2. Interleaving and packetization for encoded atoms.

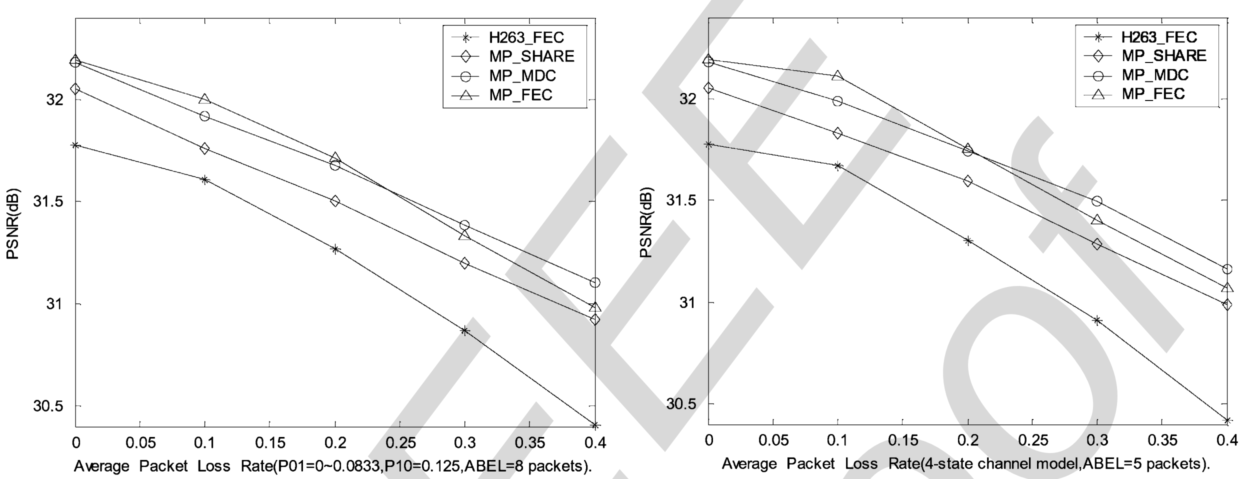
central partition to find the optimal decoder for this central partition, and then calculate the Lagrangian function L_2 using (4). The process is repeated for choosing the encoders and decoders with the smallest value of the Lagrangian function $L_{\min}$ to produce multiple descriptions for each atom parameter.

For all of the atom parameters, we use the modified nested index assignment [1]. Table I shows the number of bits for atom parameters before and after MDSQ. The total number of bits is 34 b for each MP atom, which includes the largest inner product (8 b), the corresponding dictionary indexes (10 b) and the locations (16 b). For the largest inner product, the sign bit is copied and put into the two descriptions, and the absolute value of the



(a) (b)

Fig. 3. Akiyo sequence encoded with 45 kb/s and 10 frames/s.



(a) (b)

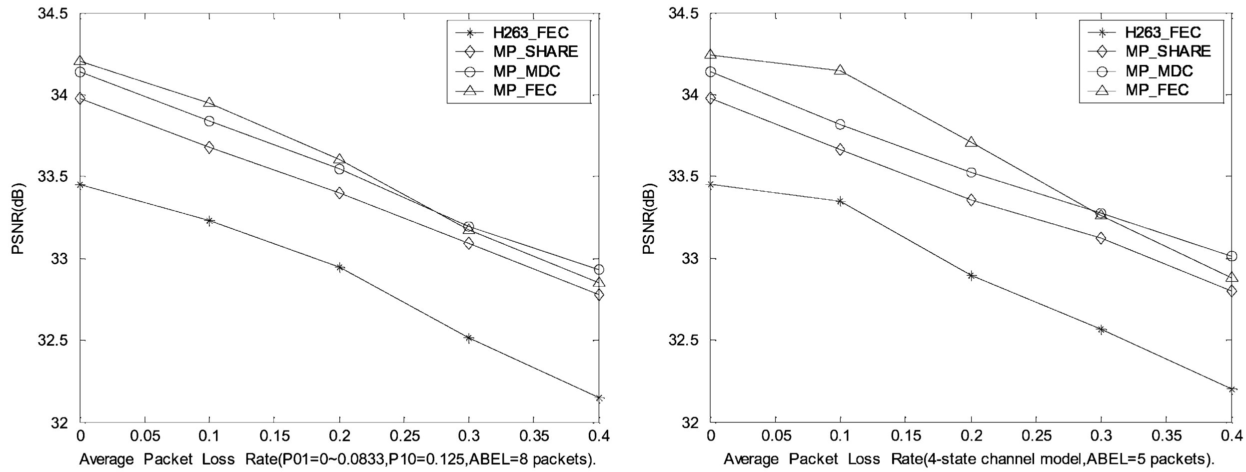
Fig. 4. Hall sequence encoded with 45 kb/s and 10 frames/s.

largest inner product is encoded by MDSQ. The two descriptions of the largest inner product are encoded using 8 b. For the dictionary indices, the two indexes are quantized by two MDSQs, respectively. So we need 12 b for quantizing the two dictionary indexes into two descriptions, respectively.

The MP-encoded video is very sensitive to the location description errors of the atoms. The performance of decoded frame will be degraded by several dB if there are quantization errors in the locations (x and y). Therefore, we need to use more bits to encode the locations using MDSQ, which is mentioned as follows: the first 4 b of each direction x (or y) are copied and put into the two descriptions and the remaining bits of each direction are re-quantized by MDSQ into two descriptions, respectively. Since the first four significant bits are copied to both descriptions, the quantization errors will be tolerable. If the number of repetition bits is increased, then more error-resilient redundancy will be added. On the other hand, if the number of repetition bits is decreased, then the quantization errors of side-decoder will be increased dramatically.

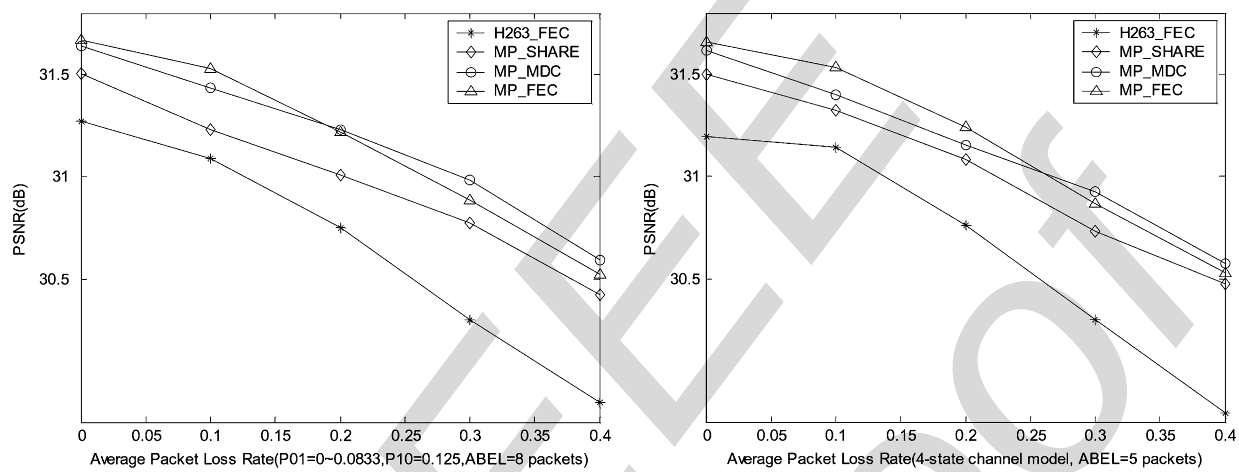
Repeating 4 b is the best choice for encoding the atom position of the MP-coded QCIF format video. MDSQ encodes the location x (or y) using two 7-b descriptions. Each description can be decomposed into two parts, i.e., $\{d_1 d_2\}$ where d_1 and d_2 represent the copied bits and the number of bits after MDSQ for one description, respectively. Here, we denote $\{4,3\}$ to represent the first 4 copied bits and 3 MDSQ bits in one description. The 7-b description can be decomposed of different combinations, such as $\{5,2\}$ and $\{6,1\}$. From Table II, we can see that the spread for all of the three cases is the same, therefore, the quantization errors for $\{4,3\}$, $\{5,2\}$, and $\{6,1\}$ are also the same.

If we use 6 b for one description and choose $\{4,2\}$ or $\{3,3\}$, it will result in higher side quantization errors and lower SNR performance because of the larger spread. Therefore, we choose $\{4,3\}$ to encode each location x (or y) into two descriptions, and generate 28 b for encoding the locations (x and y). Totally we require 48 b for encoding an atom into two descriptions. The increased redundant bits of an atom after MDSQ are 14 b (41%



(a) (b)

Fig. 5. Mom sequence encoded with 45 kb/s and 10 frames/s.



(a) (b)

Fig. 6. Foreman sequence encoded with 45 kb/s and 10 frames/s.

increment). The redundancy of the location descriptions is the largest, due to its noise-sensitivity is the highest.

For MP shared-atom scheme [16], the first 25 atoms are shared by both descriptions and the subsequent atoms are alternatively put into the two descriptions. In our simulation, we also implement the single description coding (SDC) scheme (i.e., MP and H.263). The SDC is cascaded with FEC, it is a **<DEFINE RS.>** RS (n, k) code which can correct $n - k$ errors at most. To reduce the effect of the burst errors occurring on a lossy channel, we use the interleaving scheme to rearrange packets for atoms as shown in Fig. 2. The first atoms of every 12 frames are put into the same packet, so are the second atoms of every 12 frames, and so on. Interleaving causes additional delay, but it does not increase bandwidth, and the packet burst loss will be scattered to different frames.

V. Experimental Results

In the experiment, we compare the performance of MDC with the SDC (protected by FEC). We may find that MDC outperforms SDC over lossy channel at larger loss rates, and at small loss rates the SDC outperforms MDC due to redundancy in MDC. Here, we compare the $\mathrm{MP\_MDC}$ scheme with three other schemes: 1) MP shared atom scheme; 2) MP with FEC; and 3) H.263 with FEC. Under the condition of the same total bit rate, we do the simulation of video transmission over various channel conditions. The QCIF video sequences are originally encoded at 36 k /s, 10 frames/s, and the total encoded bit rates (including the source coder and the FEC coder) of our method and the other three methods are the same, i.e., 45 kb/s. To prevent the error propagation, I-frames are inserted every 12 frames. We also assume that I-frames and motion vectors are highly protected and are decoded without errors. We test four video sequences including Akiyo, Hall Monitor, Mom and Daughter, and Foreman. The comparison experimental results are shown in the Figs. 3–6. We use the two-state Markov model and four-state channel model for our simulation.

Figs. 3(a)–6(a) show the comparison of the PSNR of our $\mathrm{MP\_MDC}$ scheme and the other schemes under the condition that the two-state model with average burst error length $(\mathrm{ABEL}) = 8$ and average packet error rate $= 040\%$. Figs. 3(b)–6(b) show the comparison of the PSNR when the four-state

(a) (b)

Fig. 7. (a) Akiyo sequence is encoded with 45 kb/s, the average packet loss rate is 0.3, and the PSNR of frame #4 is 35.49 dB. (b) Foreman sequence is encoded with 45 kb/s, the average packet loss rate is 0.3, and the PSNR of frame #21 is 31.22 dB.

channel model is applied $(\mathrm{ABEL} = 5)$. The decoded QCIF image frames of Akiyo and Foreman with transmission loss are shown in Fig. 7.

From the simulation results, we show that the $\mathrm{MP\_MDC}$ scheme outperforms the MP shared-atom scheme and H.263 with FEC under different channel conditions. Under lower average packet loss rate, the performance of the $\mathrm{MP\_MDC}$ scheme is not as good as the MP with FEC, however, under higher average packet loss rate, the $\mathrm{MP\_MDC}$ will surpass the MP with FEC. Therefore, the $\mathrm{MP\_MDC}$ may provide a more robust error resilient capability than the FEC.

VI. Conclusion

Experimental results show that the $\mathrm{MP\_MDC}$ scheme outperforms the other three schemes under different channel conditions. Under low channel error rate, the $\mathrm{MP\_MDC}$ is not as good as MP with FEC, however, it is still an effective error-resilient scheme for transmitting video over lossy channels.

References

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

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

[3] **<PLEASE PROVIDE ISSUE AND MONTH.>**V. A. Vaishampayan and S. John, “Balanced iterframe multiple description video compression,” *IEEE Trans. Image Process.*, vol. 3, pp. 812–816, 1999.

[4] 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, Mar. 2001.

[5] 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. 821–834, Mar. 2001.

[6] A. R. 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 ICIP*, 2000, pp. 136–139.

[7] R. Neff and A. Zakhor, “Very low bit-rate video coding based on matching pursuit,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 158–171, Feb. 1997.

[8] M. Yajnik, S. Moon, and J. Kurose, “Measurement and modeling of the temporal dependence in packet loss,” *Proc. INFOCOM*, pp. 78–98, 1999.

[9] J. Apostolopoulos, W. T. Tan, S. Wee, and G. W. Wornell, “Modeling path diversity for multiple description video communication,” in *Proc. IEEE ICASSP*, vol. 3, 2002, pp. 2161–2164.

[10] O. K. Al-Shaykh, E. Miloslavsky, T. Nomura, R. Neff, and A. Zakhor, “Video compression using matching pursuits,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 123–143, Feb. 1999.

[11] R. Neff and A. Zakhor, “Modulus quantization for matching-pursuit video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 6, pp. 895–912, Sep. 2000.

[12] C. De Vleeschouwer and A. Zakhor, “Atom modulus quantization for matching pursuit video coding,” in *Proc. ICIP*, vol. 3, Jun. 2002, pp. III-681–III-684.

[13] R. Caetano, E. A. da Silva, and A. G. Ciancio, “Matching pursuits video coding using generalized bit-planes,” in *Proc. ICIP*, vol. 3, Jun. 2002, pp. III-677–III-680.

[14] Y. T. Chou, W. L. Hwang, and C. L. Huang, “Gain-shape optimized dictionary for matching pursuit video coding,” *Signal Process.*, vol. 83, pp. 1937–1943, Sep. 2003.

[15] M. F. Tariq, P. Czerepinski, A. Nix, D. Bull, and N. Canagarajah, “Robust and scalable matching pursuits video transmission using the Bluetooth air interface standard,” *IEEE Trans. Consum. Electron.*, vol. 46, pp. 673–681, Aug. 2000.

[16] **<PLEASE PROVIDE ISSUE NUMBER AND MONTH.>**X. Tang and A. Zakhor, “Matching pursuit multiple description coding for wireless video,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 566–575, 2002.