

MULTIPLE DESCRIPTION AND MATCHING PURSUIT CODING FOR VIDEO TRANSMISSION OVER THE INTERNET

Hsi-Tzeng Chan and Chung-Lin Huang
Electrical Engineering Department
National Tsing-Hua University
Hsin-Chu, Taiwan
e-mail: clhuang@ee.nthu.edu.tw

ABSTRACT

This paper proposes a video communication scheme combining the multiple description coding (MDC) using scalar quantizers with the matching pursuit (MP) to transmit the video over the Internet. The parameters obtained after MP, called atoms, are encoded into two balanced descriptions, which are transmitted over two separate channels. Simulation results show that our proposed method surpasses MP single description coding (SDC) and H.263 with FEC as well as MP MDC using share atoms.

Keywords: Multiple Description Coding (MDC), Matching Pursuits (MP).

1. INTRODUCTION

Recently, there has been considerable interest in transmitting of coded video sequences over the Internet. However, coded video data are extremely sensitive to channel errors. Diversity is commonly used to enhance the reliability of the communication systems. The probability of receiving some information from at least one of the channels is greatly increased. Multiple description coding (MDC) addresses the problem of coding a source for transmission over a communication system with diversity. MDC encodes the source into two descriptions. If both descriptions are correctly received, a high-quality reconstruction can be decoded while, if either one is lost, a lower-quality, but acceptable, reconstruction can be decoded from the other one.

MDC methods [1-4] are divided into two main categories: multiple description scalar quantization (MDSQ) [1] and pairwise correlating transforms (MDCT) [4]. Quantizer methods are used to produce two complementary descriptions of the same scalar quantity using two similar quantizers offset from each other. Transform methods maintain correlation among the transformed coefficients. Wang et al. [4] shows that transform methods can reach a lower redundancy range, but it does not perform as well as quantizer method.

The block-based discrete cosine transform (DCT) has been used to encode the residual images is efficient but it introduces undesirable blocking artifacts at low bit rates. Matching pursuits (MP) residual coding [5], does

not suffer from blocking artifacts because a over-complete basis set is used to match the residual, and the area with the largest energy value is first encoded.

Tang and Zakhor [6] propose a MP MDC scheme that the residual is coded into two set of atoms, F1 and F2, to be sent over two channels. The first L atoms found during MP iterations are shared by both sets and subsequent atoms are alternatively put into the two sets. The correlation between these two sets of atoms is controlled by the number of the shared atoms. Because the redundancy is large for the shared-atom method[6], we propose a more efficient scheme by using MDSQ to encode the first L atoms. Input frame sequence is encoded by MP video coding, and the atoms obtained after MP are encoded into two balanced descriptions, which are transmitted over two separate channels.

2. MATCHING PURSUITS VIDEO CODING

The concept of MP is 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 dictionary[7]. MP provides extremely flexible signal representations since the selection of the dictionaries is arbitrary. With such a flexible choice of the basis functions, MP can be used to encode the residual images. To encode the motion residual signal, we must first extend the method to the discrete 2-D domain with the proper choice of a basis dictionary. The dictionary consists of an over-complete collection of 2-D separable Gabor functions of which the 1-D discrete function is a scaled, modulated Gaussian function as

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

where $i \in \{1, 2, \dots, N\}$ and $\vec{\alpha} = (s, \xi, \phi)$ is a triplet consisting of a positive scale, a modulation frequency, and a phase shift. If we consider B as the set of all such triples $\vec{\alpha}$, then we can specify our 2-D separable Gabor functions as $G_{\vec{\alpha}, \vec{\beta}}(i, j) = g_{\vec{\alpha}}(i)g_{\vec{\beta}}(j)$ where $i, j \in \{0, 1, 2, \dots, N-1\}$, and $\vec{\alpha}, \vec{\beta} \in B$.

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 set. The MP coding scheme needs to examine each 2-D dictionary structure at all possible integer-pixel locations in the image and compute all of the resulting inner products. We assume that the image is sparse, and it contains pockets of energy at locations where the motion prediction model was inadequate. The location of “high-energy” pockets can be used as an initial estimate for the inner-product search. The motion residual image to be coded is first divided into blocks, and the sum of the squares of all pixel intensities is computed for each block. The center of the block with the largest energy value is selected as an initial estimate for the inner product search. The dictionary is then exhaustively matched to a window around the initial estimate. An atom consists of five parameters shown in the following table.

Table 1 The parameters define an atom.

$\tilde{\alpha}, \tilde{\beta}$	The best match structure element from dictionary
x, y	Location of the best match in residual image.
p	Value of the largest inner product

3. MULTIPLE DESCRIPTION CODING

A block diagram of the operation of the MDSQ is shown in Figure 1. A source sample x is mapped by the quantizer $q(\cdot)$ to the central reconstruction \hat{x}^0 and the side reconstructions \hat{x}^1 and \hat{x}^2 from the codebooks $X^0, X^1, X^2 \in \mathcal{C}$, where $X^0 = \{x^0_{ij}, (i, j) \in H\}$, $X^1 = \{x^1_i, i \in H\}$, $X^2 = \{x^2_j, j \in J\}$, $I = \{1, 2, \dots, K\}$, $J = \{1, 2, \dots, K\}$, $H \subseteq I \times J$, and the output of scalar quantizer is an index l . Then the index assignment $s(\cdot)$ map l to a pair of indices $(i, j) \in H$. The indices i and j are separately transmitted over the two channels. 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 $e_1 : \mathcal{C} \rightarrow I$ and $e_2 : \mathcal{C} \rightarrow J$, where \mathcal{C} is the set of the reconstruction level of the codebook. The encoder produces two indices i and j , respectively. The three decoders of the MDSQ can be represented by the mapping $\mathbf{r}_0 : H \rightarrow \mathcal{C}$ (central decoder), $\mathbf{r}_1 : I \rightarrow \mathcal{C}$ and $\mathbf{r}_2 : J \rightarrow \mathcal{C}$ (side decoders), and the output of the three decoders are the reconstruction levels with indices i, j and (i, j) from the codebooks X^1, X^2, X^0 , respectively. In conclusion, the MDSQ is constructed by the partition \mathbf{P} and the three decoders $\mathbf{r} = (\mathbf{r}_0, \mathbf{r}_1, \mathbf{r}_2)$.

Here, we consider the balanced descriptions, so that the two descriptions are equal rate descriptions, and identical average distortions when either description is received. The MDSQ can be treated as a constrained optimization problem of which the Lagrangian function is defined as

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

We use an iterative descent algorithm to determine an optimal $(\mathbf{P}^*, \mathbf{r}^*)$ for given $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$. We find the encoder partition \mathbf{P}^* so that $L(\mathbf{P}, \mathbf{r}^*, \lambda_1, \lambda_2)$ is minimized over all \mathbf{P} , and find the decoder \mathbf{r}^* so that $L(\mathbf{P}^*, \mathbf{r}, \lambda_1, \lambda_2)$ is minimized over all \mathbf{r} . The process is repeated until the value of the Lagrangian function $L(\mathbf{P}^*, \mathbf{r}^*, \lambda_1, \lambda_2)$ is small enough.

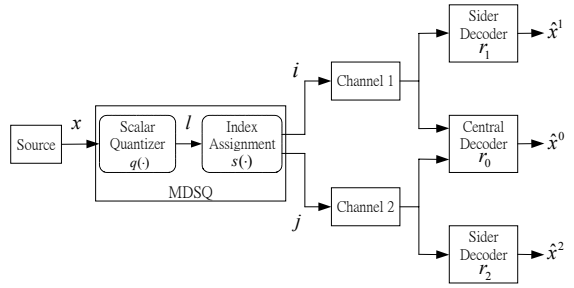


Figure 1 The MDSQ system with two channels.

The index assignment is a mapping from the index l of each central cell to a codeword pair $(i, j) \in H$. The index l is put into a matrix, and then the row and column entries (i, j) are the two descriptions, which are transmitted over two channels. In [1], it is shown that the performance of an index assignment is determined by a parameter, which is called the *spread* of an index assignment. Spread means the difference between the maximum and the minimum number in each projection of the matrix. We can find that if the spread is large, the side distortion will become large. Thus, the spread should be minimized for a good set of index assignment.

4. EXPERIMENTATION AND COMPARISON

For video packet transmission over the Internet, the packets are either received correctly or lost. These losses are mainly caused by network congestion and queuing delay. A packet loss model determines the probability of the event that a packet is lost. The conventional packet loss model is the two-state Markov model [9], in which, the loss process is modeled as a discrete-time Markov chain with two states. The current state S_i of the stochastic process depends only on the previous value S_{i-1} . The model has two states, good state (0) and bad state (1). For a MD source delivered over a joint link, the 4-state channel model is proposed [10]. There are four states which express whether both descriptions are correctly received (state 00), only one description is correctly received and another description is lost (state 01 and state 10), and both descriptions are lost (state 11).

The MP MDC scheme encodes the first 50 atoms by using MDSQ, and the subsequent atoms are alternatively put into the two descriptions. To obtain optimal encoders and decoders for each atom parameter, we need to collect the statistics of each atom parameter. According to the statistics, we can determine a central partition for each atom parameter, obtain the optimal decoders, and then calculate the Lagrangian function L_I . The process is repeated to choose the encoders and decoders with the smallest value of the Lagrangian function L_{min} to produce the multiple descriptions for each atom.

Table 2 shows the number of bits for atom parameters before and after MDSQ. The total number of bits is 34 bits for a MP atom, which includes the largest inner product (8 bits), the corresponding dictionary indexes (10 bits) and the locations (16 bits). For the largest inner product, the sign bit is copied and put into the two descriptions and the absolute value of the largest inner product is encoded by MDSQ. The two descriptions of the largest inner product require 8 bits. For the dictionary indices, the two indexes are quantized by two MDSQs respectively. Totally we need 12 bits for quantizing the two dictionary indices into two descriptions respectively. The MP-coded images are very sensitive to the errors of the locations of the atoms. Therefore, to encode the locations, we copy the first 4 bits of each direction into the two descriptions and the remaining bits of each direction are encoded by MDSQ into two descriptions respectively.

MDSQ applied on location x has created two 7 bits 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, and use 28 bits for the two descriptions of the locations. Totally, we require 48 bits for encoding an atom in two descriptions. The increased redundant bits of an atom after MDSQ are 14 bits. Most of the redundant bits are used for encoding the locations.

Table 2. The number of bits for encoding atoms.

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

In the MP shared-atom scheme[2], 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 SDC scheme of MP and H.263, and use FEC to protect MP atoms and DCT coefficients with $RS(n, k)$ code. To combat the burst errors occurring on a lossy channel, we use interleaving and rearrange packets for atoms as shown in Figure 2. The first atoms of every 12 frames are gathered and packetized together, and the second atoms of every 12 frames are gathered and packetized together, and so on. Interleaving will cause additional delay, but it does not increase bandwidth and a burst losses on the channel will be transformed into a sequence of isolated losses by interleaving.

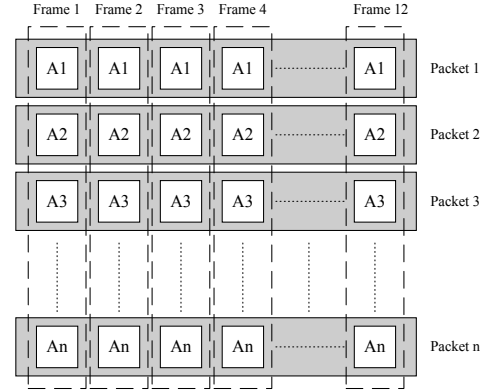


Figure 2 Interleaving and Packetization for atoms.

Here, we compare the MP MDC scheme with three other schemes: (1) MP shared-atom scheme, (2) MP SDC with FEC, (3) H.263 SDC with FEC. Under the condition of the same total bit rate, we run the simulations for all of the four different schemes on different channel conditions. The performance evaluation is done in the situation when the video sequences are coded at 36k bits/s, 10 frames/s, with QCIF format, and the total bit rate after MDSQ for MDC and FEC for SDC is 45k bits/s. To prevent the error propagation, I frames are inserted every 12 frames. We also assume that I frames and motion vectors are received without errors. We use the video sequence Akiyo in our simulation.

Figure 3(a) shows the PSNR of our MP MDC scheme and the other schemes under the two-state model with average burst error length (ABEL) being 5 and average packet error rate from 0 to 40%. Figure 3(b) shows the PSNR of the four schemes under the two-state model with ABEL = 8 and average packet error rate from 0 to 40 %. Figure 3(c) shows the PSNR under the two-state model when the two channels of the MDC schemes are unbalanced (with different error rates). Figure 3(d) shows the PSNR under the four-state channel model. In these figures, we can see that our MP MDC scheme always outperforms the MP shared-atom scheme and H.263 using FEC on different channel conditions. Our MP MDC scheme also beats the MP with FEC scheme at high error rate. But at low error rate, the

performance of MP with FEC is better.

5. CONCLUSIONS

This paper proposes a video scheme based on MP and MDC. Performance comparison is made between our MP MDC scheme and the other schemes. Experimental results show that our MP MDC scheme surpasses the other schemes with the same total rate on different channel conditions, except that MP using FEC would perform better at low error rate.

REFERENCE

- [1] V. A. Vaishampayan, "Design of Multiple Description Scalar Quantizers," IEEE Trans on IT, Vol. 39, No. 3, pp. 821-834, May 1993.
- [2] V. K. Goyal, "Multiple Description Coding: Compression Meets the Network," IEEE Signal Processing Magazine, Vol.18, pp. 74-93, Sept. 2001.
- [3] V. A. Vaishampayan; and S. John, "Balanced Interframe Multiple Description Video Compression," IEEE Trans on IP, Vol. 3, pp. 812-816, 1999.
- [4] Y. Wang, M. T. Orchard, V. A. Vaishampayan; and A. R. Reibman, "Multiple Description Coding Using Pairwise Correlating Transforms," IEEE Trans on IP, Vol. 10, No. 3, pp. 821-834, March 2001.
- [5] R. Neff and A. Zakhor, "Very Low Bit-rate Video Coding Based on Matching Pursuit," IEEE Trans on CAS for VT, Vol. 7, No.1, pp.158-171, Feb 1997.
- [6] X. Tang and A. Zakhor, "Matching Pursuit Multiple Description Coding for Wireless Video," IEEE Trans on CAS for VT, Vol. 12, pp. 566-575, 2002.
- [7] S. Mallant and Z. Zhang, "Matching Pursuit with Time-Frequency Dictionary," IEEE Trans on SP, Vol. 41, pp. 3397-3415, Dec 1993.
- [8] Telecommunication Standardization Sector of ITU, "ITU-T Recommendation H.263," Feb 1998.
- [9] M. Yajnik, S. Moon, and J. Kurose, "Measurement and Modeling of the Temporal Dependence in Packet Loss," INFOCOM, pp. 78-98, 1999.
- [10] J. Apostolopoulos; W. Tan; S. Wee; and G. W. Wornell, G.W., "Modeling Path Diversity for Multiple Description Video Communication," IEEE ICASSP 2002, pp. 2161-2164.

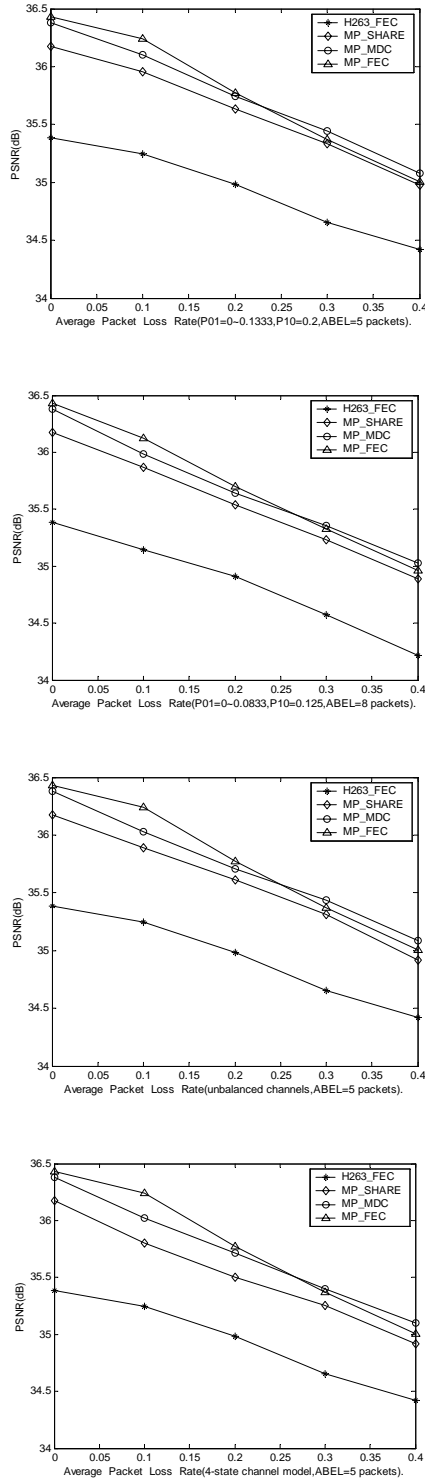


Fig 3 Akiyo sequence with 45 kbits/s, 10 frames/s.