MATCHING PURSUITS BASED MULTIPLE DESCRIPTION VIDEO CODING FOR LOSSY ENVIRONMENTS

Thinh Nguyen and Avideh Zakhor

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley, CA   94720
{thinhq, avz}@eecs.berkeley.edu

Invited Paper

ABSTRACT

Multiple description coding (MDC) is an error resilient source coding scheme that creates multiple descriptions of the source with the aim of providing an acceptable reconstruction quality when only one description is received, and improved quality as more descriptions become available. Recently, we developed a matching pursuit multiple description video coder (MP-MDVC) based on a three loop structure originally proposed by Reibman and her colleagues (MTDC). While the MP-MDVC outperforms the discrete cosine transform based MTDC, it is not optimized for lossy environments. In this paper, we extend MP-MDVC by considering the network loss characteristics when coding multiple descriptions. In particular, we propose a fast steepest descent algorithm for creating multiple descriptions that results in minimum expected distortion, given network outage probability, bandwidth constraints, and maximum allowable distortion for each description. Analytical and experimental results show that by taking network loss characteristics into account, our approach outperforms existing MP-MDVC techniques.

1. INTRODUCTION

Multiple description coding (MDC) is an error resilient source coding scheme that generates multiple encoded bitstreams of the source with the aim of providing an acceptable reconstruction quality of the source when only one description is received, and improved quality when multiple descriptions are available. This is different from layered coding which requires the presence of the base layer for enhancement layer to be useful. As such, MDC is useful for delay limited applications in lossy environments such as wireless communication where outage can last several seconds or longer, rendering retransmission less useful. MDC can also be useful in emerging peer-to-peer (P2P) systems [1] on the Internet. In P2P applications such as Kazaa [1], each peer has the ability to limit the bandwidth for file transfer to others. As such, a single 400kbps connection from one receiving peer to another serving peer does not provide sufficient bandwidth for streaming a 700kbps video. This situation can be remedied by multiple serving peers stream the video simultaneously to the receiving peer. However, the frequent joins and leaves of peers can potentially disrupt the smooth video streaming on the order of several seconds to minutes. The solution to the dynamics of peers joining and leaving is to use multiple descriptions of video in which, each serving peer streams an independently decodable video description to the receiving peer. MDC can also be used in conjunction with our recently proposed Path Diversity with Forward error correction (PDF) system for delay limited applications, in which packets are sent simultaneously on multiple paths created via relay nodes [2].

Practical MDC schemes have been developed for video, and their performance over lossy channels have been characterized [3]

[4][5]. Recently, an MDC technique based on matching pursuits (MP) signal decomposition, called MP-MDVC [6][5], has been shown to outperform Discrete Cosine Transform (DCT) based MDC [3]. In [4], the author applies the ROPE framework [7] to the DCT based MDC in [3] in order to optimize the video quality in lossy environments. In this paper, we extend MP-MDVC [5] to optimize it for lossy environments. We choose MP-MDVC for the following reasons. First, MP-MDVC has been shown to outperform DCT based multiple description video coder[5]. Second, the rate distortion analysis for MP is simple and elegant as compared to orthogonal transforms where quantization is done after the transform is completed [8]. In particular, since the number of bits per atom[1] is more or less constant, rate control and rate prediction is straightforward in MP coding systems. The same is true for distortion prediction as adding an atom reduces distortion by the square of its magnitude. In contrast with MP, the rate and distortion in DCT based video coders are typically controlled by the quantization step size. As a result, for a given quantization step size, the rate and distortion are not known until the all DCT coefficients are coded for the entire frame. These factors result in a natural framework and ease of analysis for extending MP-MDVC to lossy environments, that is absent in DCT formulation.

The motivation for incorporating loss characteristics into multiple description coding technique is as follows. In high loss environments, most likely only one description at a time is received, hence it is desirable to achieve high reconstruction quality from a single description. On the other hand, in low loss environments, two descriptions are likely to be received at a time, hence it is desirable to achieve as high of a reconstruction quality as possible from two descriptions. Since for a given total bit rate, there is an inherent trade-off between the reconstruction qualities of one and two descriptions, depending on the level of loss, one should strike a balance between the single and multiple description qualities. In this paper, we formulate and solve this problem of how to achieve the above mentioned quality trade-off for matchting pursuits based MDC. In particular, we propose a fast, steepest descent algorithm for computing the optimum number of atoms in each description based on the outage probabilities of the two channels so as to achieve minimum expected distortion, given bandwidth constraints, and maximum allowable distortion for each description. Our approach is based on a simple rate distortion analysis of MP, resulting in a fast practical implementation.

The rest of our paper is organized as follows. In Section 2, we present overviews of MP and MP-MDVC. Next in Section 3, we describe our proposed optimization algorithm for lossy environments. In Section 4, we provide experimental results showing improved video quality over MP-MDVC[5]. Finally, we conclude in Section 5.

---

[1]an atom is a basis function together with the position information and magnitude

2.1. Overview of MP

In many video coding standards, residual images are encoded using block-based DCT coding, thus introducing noticeable blocking artifacts at low bit rates. In matching pursuits video coding, motion compensated residual frames are decomposed into an overcomplete set of basis functions that is much larger than the complete basis set in DCT. Residual frame is coded using an iterative, greedy algorithm in which, at each iteration, the basis function with the largest inner product is subtracted from the residual frame. Since the magnitude of the inner product between the residue and the chosen function corresponds to distortion reduction at that iteration, this iterative, greedy method ensures that most important features are coded first. More details on matching pursuits video coding can be found in [9].

2.2. Overview of MP-MDVC

In this section, we briefly present the three loop structure originally proposed in [3] and subsequently used in MP-MDVC [5]. Figure 1 shows three loop MP-MDVC structure for generating two descriptions. In the central prediction loop, a new frame is first motion compensated from its prediction based on both descriptions, while in the side loops, the new frame is motion compensated based on only one description. This approach is employed in order for the decoder to track the encoder states when both descriptions are received, or when one of the descriptions is lost. Residue encoder applies MP decomposition to the motion compensated residue as described in Section 2.1, to result in a set of atoms. Let $F1(F2)$ denote the set of atoms generated by central loop for channel 1(2) as shown in Figure 1. In [5], the first L atoms found during MP decomposition in the central loop are shared by both sets $F1$ and $F2$, and subsequent atoms are alternately assigned into the two sets. Since atoms are found in decreasing order of magnitude, this results in the central loop atoms in $F1$ and $F2$ to be of approximately equal importance. Similarly, the sets $G1$ and $G2$ containing atoms from the side loops for channels 1 and 2 respectively, are found using MP decomposition. For convenience, we denote $F = F1 \cup F2$, $G = G1 \cup G2$, and the atom type by the name of its set, e.g. $F1$ atom. $F1$ and $G1$ atoms are sent on channel 1 while $F2$ and $G2$ atoms are sent on channel 2. Motion vectors, frame headers, and intra-coded (I) frames are duplicated and sent through both channels. Energy of residue $R1(R2)$ from motion compensation based on one description is first reduced by exploiting its correlation with coded residue from the central loop $F1(F2)$ through pixel-wise subtraction. Also, to save bits for residual coding, the same motion vectors are used in both side loops and central loop. More details on MP-MDVC can be found in [5].
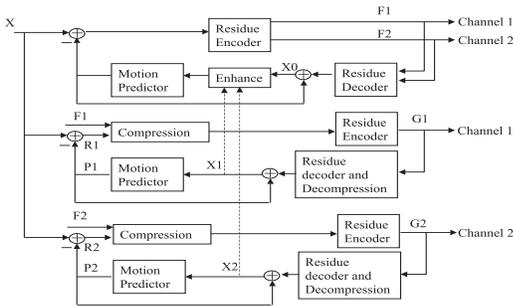

Fig. 1. *Three loop structure of MP-MDVC.*

3. FAST ALGORITHM MP-MDVC

In this section, we extend the MP-MDVC approach in [5] to optimally select the number of atoms in the central and side loops so as to match the outage probability of the channels.

3.1. Problem Formulation

Using the notation in Table 1, and assuming independent loss of descriptions, we wish to find a fast algorithm to compute f_1, g_1, f_2,

$p_{1(2)}$	Probability of losing description 1(2)
D_0	Distortion induced by receiving both descriptions
$D_{1(2)}$	Distortion induced by receiving single description 1(2)
D'	Distortion induced by losing both descriptions
D^*	Maximum allowable distortion of any description
$R_{1(2)}$	Bits per frame of description 1(2)
$R_{1(2)}^*$	Maximum allowable bits per frame of description 1(2)
$f_1(f_2)$	Number of F atoms for description 1(2) per frame
$g_1(g_2)$	Number of G atoms in description 1(2) per frame

Table 1. *Notation*

and g_2 in order to minimize expected distortion, J, given by:

$$
\begin{aligned}
J &= (1-p_1)p_2 D_1(f_1, g_1) + (1-p_2)p_1 D_2(f_2, g_2) \\
&+ (1-p_1)(1-p_2)D_0(f_1, f_2, g_1, g_2) + p_1 p_2 D' \quad (1)
\end{aligned}
$$

subject to following constraints:

$$
\max(D_1(f_1, g_1), D_2(f_2, g_2)) \leq D^* \quad (2)
$$
$$
R_1(f_1, g_1) \leq R_1^*
$$
$$
R_2(f_2, g_2) \leq R_2^*
$$

In solving this problem, we are inherently assuming that D^*, R_1^*, and R_2^* are chosen in such a way that the set of feasible solutions is non-empty. The last term in Equation (1) can be ignored since it does not contain free variables. This is a general non-linear convex optimization with non-linear constraints. However, we can simplify the problem to arrive at a practical solution by simplifying the constraints. First, we remove the distortion constraint in (2) and add a penalty term P to J and call this sum J' to get

where
$$
J' = P + J \quad (3)
$$

$$
P = (|D^* - \max(D_1, D_2)| + \max(D_1, D_2) - D^*)^2 \quad (4)
$$

The rationale for this is that if the distortion constraint in (2) is not satisfied, namely, $\max(D_1, D_2) > D^*$ then $P = 4(\max(D_1, D_2) - D^*)^2$ is four times the square of the distortion difference. This is a large value resulting in large J', thus making the corresponding tuple (f_1, g_1, f_2, g_2) an unlikely solution. On the other hand, if the distortion constraint is satisfied, then $P = 0$, resulting in the minimum values of J' and J to be identical.

Second, in matching pursuits implementation [9], the number of bits used to code an atom is approximately constant, making the number of bits per frame directly proportional to number of atoms per frame. In addition, since higher bit rate results in lower distortion, any algorithm that minimizes the distortion should use all the available bandwidth. Based on these observations, the rate constraints for description i in (2), can be rewritten as $f_i + g_i = \frac{R_i^*}{B}$ where B is number of bits per atom.

As a result, we simplify our optimization problem into the problem of finding f_1, g_1, f_2, and g_2 in order to minimize

$$
\begin{aligned}
J' &= (1-p_1)p_2 D_1(f_1, g_1) + (1-p_2)p_1 D_2(f_2, g_2) \\
&+ (1-p_1)(1-p_2)D_0(f_1, f_2, g_1, g_2) + P \quad (5)
\end{aligned}
$$

subject to following constraints:

$$
f_1 + g_1 = \frac{R_1^*}{B} \quad \text{and} \quad f_2 + g_2 = \frac{R_2^*}{B} \quad (6)
$$

3.2. Solution to Optimization Problem

The above problem can be solved fast using steepest descent method with a two pass algorithm. In the first pass, the central atoms are coded in the central loop. The resulting energies of the central atoms are then used in the second pass to code the atoms in the side loops.

Before describing the algorithm in details, we first provide several observations on which our algorithm is based.

Observation one: MP decomposition codes atoms in decreasing order of magnitude, and the distortion reduction due to an atom is roughly square of its magnitude.

Observation two: Using the three-loop structure in Figure 1, there is an inherent PSNR trade-off between the frames in the central and side loops for a given total bit rate. As the number of F atoms in the central loop increases, the number of G atoms in the side loop has to decrease in order to satisfy the constraints in Equation (6). Arguably, the side loop could potentially benefit from central loop atoms, since central loop atoms are used to predict side loop residues. However, these central loop atoms, especially the small magnitude ones, do not describe the side loop frame as well as the side loop atoms do. The reason is that the central loop atoms result from coding the frame predicted from the central loop which can be different from the frames predicted by the side loop. Hence, it is more efficient to use the same number of bits for side loop atoms to represent the side frame than central loop ones.

Observation three: Suppose a frame I_0 is obtained from MP coding the original frame I_{orig} using N atoms, frame $I_{partial}$ is the resulting frame from randomly removing $M \leq N$ atoms from I_0, and frame I_{diff} is obtained by pixel-wise subtraction of $I_{partial}$ from I_{orig}. When I_{diff} is MP coded using $K \leq M$ atoms, the distortion reduction by these K atoms is roughly the sum of squares of magnitudes of K largest atoms belonging to the set of M removed atoms. Intuitively, the reason is as follows: I_{diff} contains the high energy regions which are used to be represented by the removed atoms. Since MP decomposition finds the regions with largest energies to code first, it is likely to find and code the largest atoms that have been previously removed in frame I_0.

We now propose a fast algorithm to solve the optimization problem in Equation (5) based on the above observations.
Step 1: Assign the number of central loop atoms to the maximum allowable by bit rate constraint, namely, $|F| \leftarrow \lfloor \frac{R_1^*}{B} \rfloor + \lfloor \frac{R_2^*}{B} \rfloor$.
Step 2: Code the frame in the central loop using F atoms, and record their corresponding magnitudes. Based on observation one, these magnitudes are used to compute the distortion in later steps.
Step 3: Assign the central loop atoms alternately in the order of decreasing magnitude into two sets of atoms $F1$ and $F2$ until either $f_1 > \lfloor \frac{R_1^*}{B} \rfloor$ or $f_2 > \lfloor \frac{R_2^*}{B} \rfloor$; then assign the remaining atoms into the other unfilled set. We have experimentally verified this initial allocation of the F atoms to be robust to different loss probabilities of the descriptions. We set L, the number of shared atoms between the two descriptions in central loop in [5] to 0. The reasoning is based on observation three: When coding the side frame, the resulting G atoms from the side loops are likely to more or less correspond to the atoms assigned to the other description; hence the two descriptions are likely to share similar energy atoms resulting from side loops.
Step 4: *Steepest Descent steps:* From observation two on PSNR trade-off, for a fixed bit rate, increasing number of G atoms requires decreasing number of F atoms, and therefore, increases(decreases) the PSNR of frames in the side (central) loop. Thus, given a total number of atoms, the problem is optimum allocation of atoms between F and G sets, i.e. central and side loops. Furthermore, since the total bits per description remains fixed at $\lfloor \frac{R_i^*}{B} \rfloor$, we only allow allocation of atoms within a description, e.g. adding a side loop atom from a given description requires removing a central loop atom from the same description. Since our algorithm begins with the maximum number of central loop atoms, there are three options to descent to the minimum value: (a) decrease f_1 and increase g_1, (b) decrease f_2 and increase g_2, and (c) stop, the allocation is optimal. Starting from the smallest energy central loop atoms found in step 1, we choose one of the above options so as to lower J'. We repeat the process for the next central loop atoms in the order of increasing energy until removing a central loop atom and adding a side loop atom no longer results in a smaller value of J'.

3.3. Practical Implementation

Even though, in each step of the steepest descent algorithm, we need to compute J', we have not shown how to compute it efficiently. A simple method would be to use the three-loop structure in Figure 1 to directly code three separate frames, one for the central loop and two for the side loops, and compute the resulting distortions and J' in each step of the steepest descent algorithm. However, this simple method is computationally expensive since the two side loop frames have to be coded again every time a side loop atom is added in step 4 of the algorithm described in Section 3.2.

To reduce computational complexity, we propose a fast method to approximate the distortion of the side loop frames. Rather than coding the entire side loop frame each time a side loop atom is added in order to compute the resulting distortion, we approximate the distortion reduction by using the atom's magnitude. The key to this approximation is based on observation three. Recall that the central loop atoms are alternately assigned to each description; hence, each description is missing the central loop atoms from the other description. When coding the side loop frame for each description, the MP decomposition is likely to choose the K side loop atoms to be similar to the largest K missing central loop atoms for that description. Since the magnitude of central loop atoms are already computed and recorded in step 2, there is no need to re-code the side loop frame to compute the resulting distortion. Instead, the magnitude of these missing central loop atoms are used to approximate the distortion for each side loop frame. Also, since J' is non-increasing and there are at most $|F|$ atoms, the steepest descent converges in at most $|F|$ steps.

This approximation assumes that the reference frames in the three loops are identical. In practice, these reference frames are not identical except for the first I-frame in a group of pictures (GOP). However, they are sufficiently similar to make this approximation work well in practice as shown in Section 4.

4. EXPERIMENTAL RESULTS

In this section, we use standard MPEG CIF sequences Foreman and Hall coded at 10fps with GOP size of 30 for our experiments. The total number of atoms for each description is 100, making the approximate bit rate per description for Foreman and Hall to be 140 and 52kbps, respectively. The outage probability for both descriptions is set to 0.1. We assume that during an outage, the entire description is lost. This is reasonable since outages in wireless and P2P environments are on the order of several seconds to minutes. The minimum PSNR constraints for Foreman and Hall are chosen typically by applications or users. In this experiment, we set reasonable minimum constraints for Hall and Foreman to 31 and 29dB, respectively.

We first show the PSNR results of proposed approximations in Section 3.3. Figure 2 shows the predicted PSNR using approximation in Section 3.3 and the actual PSNR by coding the side loop frame for Foreman and Hall sequences as a function of frame number. As seen, the predicted and actual PSNR for both sequences are very close. The algorithm predicts better for Hall than Foreman sequences. In general, the algorithm predicts better for low motion sequences. Also, prediction error seems to be slightly larger as the frame number increases. This error is possibly due to the larger difference of the reference frames in central and side loops for later frames in a GOP. Another observation is that for most frames, the actual PSNR is slightly larger than the predicted PSNR. This is intuitively plausible since side loop atoms are likely to represent the side loop frame better than central atoms, as discussed in observation two. Figure 3 shows the PSNRs of two description (PSNR0) and one description (PSNR1) for Foreman and Hall sequences. As seen, the PSNRs of either description is larger than the specified minimum constraints of 29dB and 31dB for Foreman and Hall, respectively, showing that the algorithm satisfies the distortion constraints. We now consider the how PSNR0 and PSNR1 vary as a function of outage probability. For simplicity, we assume that outage probability of the two channels are identical. Intuitively, for small outage probability, one would expect PSNR0 to be large, and PSNR1 to be small
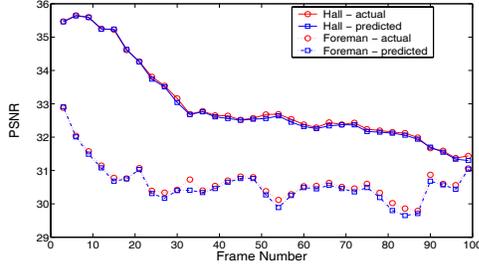
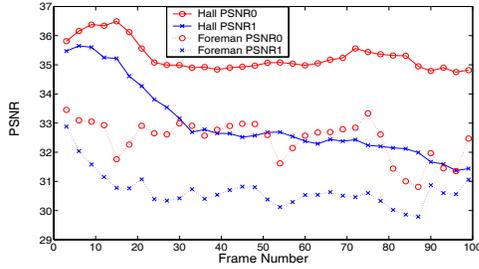**Fig. 2**. *Actual and predicted PSNR for Foreman and Hall.*

Fig. 3. *PSNR0 and PSNR1 for Foreman and Hall.*

since most of the time, the receiver receives two descriptions. On the other hand, when the outage probability is large, PSNR1 should be large since most of the time only one description is received. Figure 4 confirms this, showing PSNR0 decrease and PSNR1 increase with the outage probability. Figure 5 shows the corresponding average number of G atoms for each description as a function of outage probability. As expected, as outage probability increases, more G atoms are used, resulting in higher PSNR1. To compare the perfor-
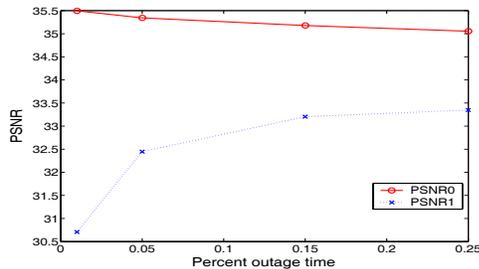

Fig. 4. *PSNR0 and PSNR 1 for Hall as a function of outage probability.*

mance of our proposed approach MP-MDVC [5], Figures 6(a) and (b) show the expected PSNR in Equation (1) for standard sequences Hall, Foreman, News, and Mom as a function of outage probability. In MP-MDVC[5], $f_1 = g_1 = f_2 = g_2 = 50$ independent of network characteristics. As seen, at low outage probability, our method results in over 1dB improvement over MP-MDVC. As the outage probability increases, the expected PSNR gap decreases. This is due to the fact that the chosen number of G atoms in MP-MDVC method is large enough, leading to higher PSNR1, and hence resulting in higher expected PSNR for high loss environments.

5. CONCLUSIONS

In this paper, we have proposed a fast steepest descent algorithm for matching pursuits multiple description video coder in lossy environments. Analytical and experimental results show that by taking network loss characteristics into account, our algorithm results in improved performance over MP-MDVC [5].
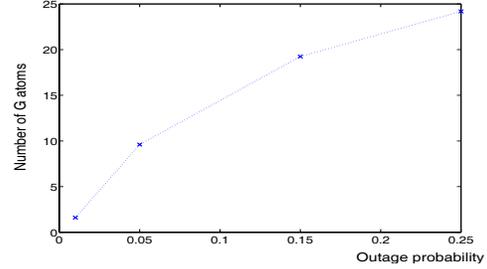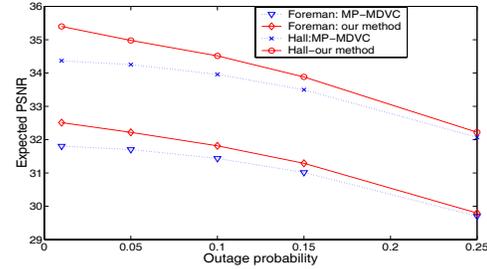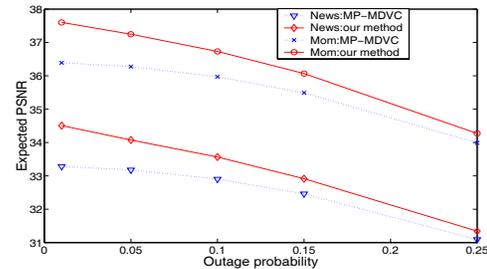

Fig. 5. *Average number of G atoms for Hall as a function of outage probability.*

(a)

(b)

Fig. 6. *Expected PSNR as a function of outage probability; (a) Foreman and Hall; (b) Mom and News.*

6. REFERENCES

[1] http://www.kazaa.com.

[2] T. Nguyen and A. Zakhor, "Path diversity with forward error corection (pdf) system for packet switched networks," in *INFOCOM*, San Francisco, CA, April 2003.

[3] A. Reibman, H. Jafarkhani, Y. Wang, M. Orchard, and R. Puri, "Multiple description coding for video using motion compensated prediction," in *Proceedings of International Conference on Image Processing (ICIP)*, October 1999, vol. 3, pp. 837–841.

[4] A. Reibman, "Optimizing multiple description video coders in a packet loss environment," in *Packet Video Workshop*, April 2002.

[5] X. Tang and A. Zakhor, "Matching pursuits multiple description coding for wireless video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 1, pp. 566–575, June 2002.

[6] X. Tang and A. Zakhor, "Matching pursuits multiple description coding for wireless video," in *Proceedings of 6th International Conference on Image Processing (ICIP)*, October 2001, vol. 1, pp. 926–929.

[7] R. Zhang, S. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 977–995, June 2000.

[8] R. Neff and A. Zakhor, "Matching pursuits video coding-part2:operational models for rate and distortion," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 1, pp. 27–39, January 2002.

[9] R. Neff and A. Zakhor, "Very low bit rate video coding based on matching pursuits," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 1, pp. 158–171, February 1997.