

# DEPENDENT BIT ALLOCATION IN MULTIVIEW VIDEO CODING

*Jae Hoon Kim, Jonathan Garcia and Antonio Ortega*

Signal and Image Processing Institute  
 Department of Electrical Engineering  
 University of Southern California  
 Los Angeles, CA 90089-2564

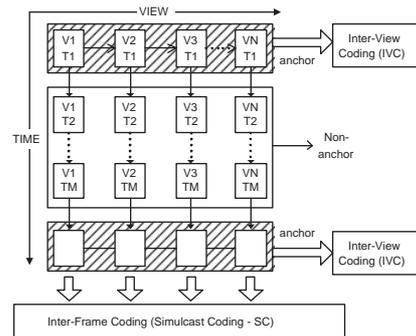
## ABSTRACT

We consider the bit allocation problem in multiview video coding (MVC). A dependent coding technique using trellis expansion and the Viterbi algorithm (VA) is proposed, which takes into account dependencies across time and views. We note that, typically, optimal quantizer choices have the following properties: i) quantization choices tend to be similar for frames that are consecutive (in time or in view), ii) better quantization tends to be used for frames closer to the root of the dependency tree. We propose a search algorithm to speed up the optimization of quantization choices. Our results indicate significant gains can be achieved by an appropriate selection of bit allocation across frames.

## 1. INTRODUCTION

Multiview video refers to sets of video sequences providing different perspectives of a common scene or object, and which have been captured simultaneously by several cameras at different locations. This is illustrated in Fig. 1, where each rectangle represents a frame captured at a given time and by a given camera. A straightforward approach for compression is simulcast coding (SC), where standard video coding techniques are applied to compress each view independently. This approach exploits temporal correlation, but does not take advantage of correlation across views. Inter-view coding (IVC) can be applied to further improve coding efficiency. For example, as illustrated in Fig. 1, we can select to insert intra-coded frames at the same time instant in all views. Then IVC can be applied to these frames (which in the sequel we will call “anchor” frames). Reconstructed anchor frames will then be used for inter-frame coding within each view. In this work, we use block-based predictive techniques for inter-view coding similar to the motion compensated predictive coding techniques used in MPEG-2, H.264/AVC, etc. Any such predictive coding techniques (across frames in a view, or across views) lead to dependencies, as quantization choices for one frame affect the achievable rate-distortion points for those frames that depend on it [3]. Note that different types of coding dependencies arise. In the SC case of Fig. 2(a), each view is coded independently, so only temporal dependency within each view can be observed. Instead, Figs. 2(b) and 2(c) represent cases where the set of anchor frames are encoded in IPPP or IBBP modes. This introduces additional dependencies across views. For example, when encoding frame V2T2, reconstructed frame V2T1 is used as a reference, and in turn V2T1 uses frame V1T1 as a reference (see Fig. 1).

Note that while the problem of dependent bit allocation has been considered in several contexts, including standard video [2–



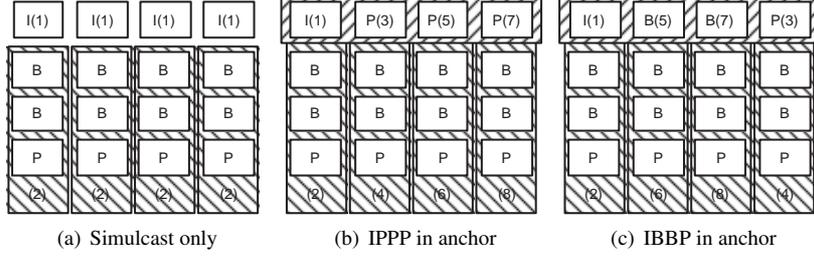
**Fig. 1.** Diagram for Multiview Video Coding.  $N$  is the number of view and  $M$  is the anchor frame interval.

4, 6] and stereo image coding [7], we believe its potential impact in multiview video coding has not been considered yet. Our work extends the frame-wise dependent bit allocation techniques of [3] (using a trellis representation and the VA) to a multiview video coding scenario where inter-view coding is used. This leads to a complex 2-D dependency problem, where the total number of video frames and candidate quantization choices involved can be very large. Moreover, a suboptimal choice of quantizer for a given frame may affect many other frames (if the frame in question is close to the root of the dependency tree). This suggests that a proper quantizer allocation may be more important in an MVC environment than for standard video. Indeed, our initial motivation to consider this problem came from the observation that in an H.264/AVC encoder we modified for MVC coding results were very sensitive to bit allocation (see results in this paper and in [1]).

In order to reduce the complexity of search in our MVC environment, we make use of the monotonicity property observed in [3]. To further reduce complexity, we show that the number of solutions to be searched can be reduced by considering only candidate solutions such that anchor and non-anchor frames are allocated similar quantizers. Our experimental results demonstrate that our proposed scheme outperforms both simulcast coding and MVC using a fixed quantization parameter (QP).

## 2. 2-D DEPENDENT BIT ALLOCATION

In what follows, distortion ( $D$ ) is measured as frame-wise mean square error (MSE). QP,  $q$ , rate,  $R$ , distortion,  $D$ , and Lagrangian cost  $J$  of the anchor frame in view  $i$ , are represented as  $q_i, R_i, D_i,$



**Fig. 2.** MVC example where the number of views is 4. The number in parenthesis is the order of encoding in the trellis expansion.

and  $J_i$ , respectively. We denote  $\bar{q}_i$  the quantization choice for the non-anchor frames in view  $i$  (in what follows we will assume the same quantizer is used for all non-anchor frames in a view).  $\bar{R}_i$ ,  $\bar{D}_i$ , and  $\bar{J}_i$  will denote the total rate, distortion and Lagrangian cost for all non-anchor frames in view  $i$ . In our notation  $q < q'$  means that quantizer  $q$  is finer, i.e., better quality, than  $q'$ .

In [3], a solution to the dependent bit allocation problem was proposed based on a trellis expansion and the VA. Our problem, which includes dependency across views, can be seen as an extension of this 1-D problem. A constrained 2-D dependent coding problem can then be formulated as follows (for the 2-view case):

$$\begin{aligned} \min_{q_1, q_2, \bar{q}_1, \bar{q}_2} [D_1(q_1) + \bar{D}_1(q_1, \bar{q}_1) + D_2(q_1, q_2) + \bar{D}_2(q_1, q_2, \bar{q}_2)] \\ \text{such that} \\ R_1(q_1) + \bar{R}_1(q_1, \bar{q}_1) + R_2(q_1, q_2) + \bar{R}_2(q_1, q_2, \bar{q}_2) \leq R_{budget}. \end{aligned} \quad (1)$$

This problem can be solved by considering an unconstrained problem with Lagrange multiplier  $\lambda \geq 0$  and cost  $J = D + \lambda R$  [5]:

$$\min_{q_1, q_2, \bar{q}_1, \bar{q}_2} [J_1(q_1) + \bar{J}_1(q_1, \bar{q}_1) + J_2(q_1, q_2) + \bar{J}_2(q_1, q_2, \bar{q}_2)] \quad (2)$$

where,

$$J_1(q_1) = D_1(q_1) + \lambda R_1(q_1) \quad (3a)$$

$$\bar{J}_1(q_1, \bar{q}_1) = \bar{D}_1(q_1, \bar{q}_1) + \lambda \bar{R}_1(q_1, \bar{q}_1) \quad (3b)$$

$$J_2(q_1, q_2) = D_2(q_1, q_2) + \lambda R_2(q_1, q_2) \quad (3c)$$

$$\bar{J}_2(q_1, q_2, \bar{q}_2) = \bar{D}_2(q_1, q_2, \bar{q}_2) + \lambda \bar{R}_2(q_1, q_2, \bar{q}_2) \quad (3d)$$

In a system with  $N$  views, assume that our bit allocation requires evaluating, on average,  $n_a$  coding choices for each anchor frame, and  $n_b$  for each set of non-anchor frames in a view. Then the bit allocation complexity will be  $O(n_b n_a^N)$ . We achieve a reduction in complexity based on two methods (to be explained in detail in the following sections). First, as in [3], we exploit the monotonicity property of dependent coding to help us reduce  $n_a$ . Second, we choose the non-anchor frame quantizers to be coarser than the quantizers chosen for the corresponding anchor frame, i.e.,  $\bar{q}_i \geq q_i$ , for the  $i$ -th view, so that a lower  $n_b$  can be used.

## 2.1. Monotonicity

The monotonicity property observed in [3] for a temporal dependency scenario states that, for two dependent frames (the second frame is motion predicted from the first one), we usually have:

$$J_2(q_1, q_2) \leq J_2(q'_1, q_2) \text{ for } q_1 \leq q'_1, \quad (4)$$

i.e., for a given quantizer,  $q_2$ , applied to the predicted frame, finer quantization of the predictor tends to lead to better RD characteristics for the predicted frame. This property usually holds when the frames in (4) are anchor frames. Similar properties can also be observed for the dependency within a view

$$\bar{J}_1(q_1, \bar{q}_1) \leq \bar{J}_1(q'_1, \bar{q}_1) \text{ for } q_1 \leq q'_1, \quad (5)$$

as well as when various levels of dependencies, across both views and time, are present, so that, for example:

$$\bar{J}_2(q_1, q_2, \bar{q}_2) \leq \bar{J}_2(q'_1, q_2, \bar{q}_2) \text{ for } q_1 \leq q'_1 \quad (6a)$$

$$\bar{J}_2(q_1, q_2, \bar{q}_2) \leq \bar{J}_2(q_1, q'_2, \bar{q}_2) \text{ for } q_2 \leq q'_2 \quad (6b)$$

From these monotonicity properties, the following lemma can be derived.

*Lemma 1:* If

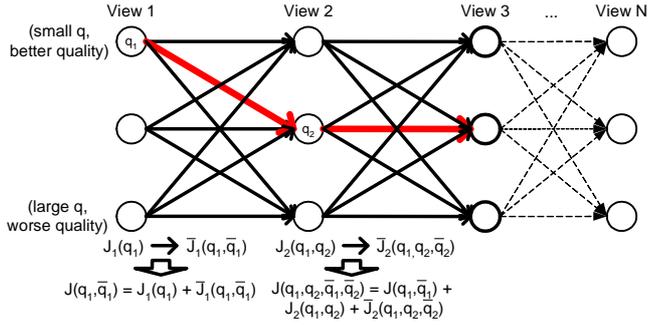
$$J_1(q_1) + \bar{J}_1(q_1, \bar{q}_1) < J_1(q'_1) + \bar{J}_1(q'_1, \bar{q}_1) \text{ for } q_1 < q'_1 \quad (7)$$

then  $q'_1$  is not in the optimal path set and can be pruned out.

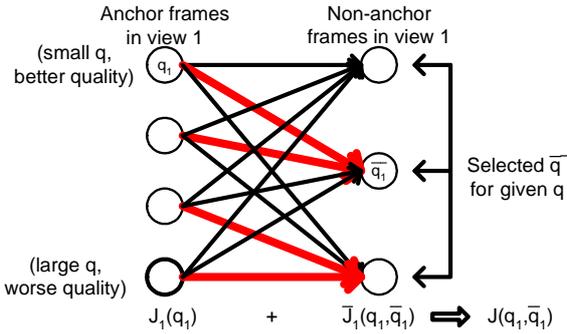
Lemma 1 above and the corresponding lemmas (1 and 2) in [3] are used in the pruning steps in our proposed algorithm. This algorithm is based on an IPPP anchor frame coding scheme as shown in Fig. 2(b). For the trellis expansion in anchor and non-anchor frames, refer to Figs. 3 and 4. In the following algorithm,  $q_1^i = \{q_1, q_2, \dots, q_i\}$  is an anchor frame quantizer allocation for views 1 through  $i$ .  $J_i(q_1^{i-1}, q_i)$  is the cost for a surviving anchor frame quantizer allocation  $q_1^{i-1}$  and anchor frame quantizer  $q_i$  in view  $i$ .  $\bar{J}_i(q_1^i, \bar{q}_i)$  is the cost for anchor frame quantizer allocation  $q_1^i$  and non-anchor frame quantizer  $\bar{q}_i$  in view  $i$ .  $J(q_1^i, \bar{q}_1^i)$  is the total cost with quantizer allocation  $q_1^i$  and  $\bar{q}_1^i$  for views 1 through  $i$ .

*Algorithm 1:*

1. For view  $i > 1$ , generate the Lagrangian cost of the anchor frame:  $J_i(q_1^{i-1}, q_i)$ , for all surviving quantizer allocations  $q_1^{i-1}$ , and for all choices of  $q_i$ . The anchor frame of view 1 is coded independently and has a cost  $J_1(q_1)$ .
2. For view  $i$ , generate the non-anchor frame cost:  $\bar{J}_i(q_1^i, \bar{q}_i)$  for  $\bar{q}_i$  for all surviving allocations  $q_1^{i-1}$ , and all anchor frame quantizers  $q_i$ , such that  $\bar{q}_i \geq q_i$ .
3. Find minimum non-anchor frame cost  $\bar{J}_i(q_1^i, \bar{q}_i)$  for each  $q_1^i$ .
4. Compute total cost  $J(q_1^i, \bar{q}_1^i) = J(q_1^{i-1}, \bar{q}_1^{i-1}) + J_i(q_1^{i-1}, q_i) + \bar{J}_i(q_1^i, \bar{q}_i)$  for each anchor frame quantizer  $q_i$ .
5. Use pruning condition of *Lemma 1* to eliminate suboptimal paths up to view  $i$ .



**Fig. 3.** Trellis expansion in anchor frame. Thick red line shows one of anchor frame quantizer allocations.



**Fig. 4.** Trellis expansion in view 1. For each anchor frame quantizer  $q$  a non-anchor frame quantizer  $\bar{q}$  with minimum cost can be chosen (thick red line) and the total cost for each quantizer allocation can be calculated.

- With every surviving path,  $q_1, q_2, \dots, q_i$ , proceed to view  $i + 1$  and go to Step 1.

Note that for each anchor frame quantizer in each surviving allocation  $q_i^j$ , there is a corresponding non-anchor frame quantizer with minimum cost, which is shown as thick red line in Fig. 4.

The above algorithm can be easily modified for either IBBP or IBP coding of anchor frames. An additional step required to search for a solution under IBP coding of anchor frames would be to populate branches between I and P1 with costs  $J_{B1}(q_I, q_{P1}, q_{B1})$  and  $\bar{J}_{B1}(q_I, q_{P1}, q_{B1}, \bar{q}_{B1})$ .

## 2.2. Reduced Search Range

Even though complexity is reduced by taking advantage of the monotonicity property, further reductions are achievable by considering the relationship between  $q$  and  $\bar{q}$  in an optimal solution. According to our experiments, optimal bit allocations are such that there exists a strong correlation between  $q$  and  $\bar{q}$ . This is shown in Fig. 5, where we plot, for different values of  $\lambda$ , the pair of quantization values for anchor and non-anchor frames that minimize the Lagrangian cost for the given  $\lambda$ . The exact slope in Fig. 5 depends in general on the number of non-anchor frames and how the anchor frame is encoded. In what follows we provide an analysis that supports the type of relationship between quantizers that we observe in optimal solutions.

Let  $Q_1$  and  $Q_2$  be the quantization choices made for the anchor frame in a view and the non-anchor frames in the same view,

respectively, where smaller  $Q$  means finer quantization. The Lagrangian cost  $J$  for that view is then

$$J = D_1(Q_1) + D_2(Q_1, Q_2) + \lambda(R_1(Q_1) + R_2(Q_1, Q_2)). \quad (8)$$

In order to better understand the properties of the optimal solution we take derivatives of  $J$  with respect to  $Q_1$  and  $Q_2$ , and set them to zero:

$$\frac{\partial J}{\partial Q_1} = \frac{\partial D_1}{\partial Q_1} + \frac{\partial D_2}{\partial Q_1} + \lambda\left(\frac{\partial R_1}{\partial Q_1} + \frac{\partial R_2}{\partial Q_1}\right) = 0 \quad (9)$$

$$\frac{\partial J}{\partial Q_2} = 0 \Leftrightarrow \lambda = -\frac{\partial D_2}{\partial Q_2} / \frac{\partial R_2}{\partial Q_2} = -\frac{a_2}{b_2}, \quad (10)$$

where  $a_i = \partial D_i / \partial Q_i$  and  $b_i = \partial R_i / \partial Q_i$ . Then, from (9) and (10),

$$a_1 b_2 - a_2 b_1 = a_2 \frac{\partial R_2}{\partial Q_1} - \frac{\partial D_2}{\partial Q_1} b_2 \quad (11)$$

$$\frac{a_1}{b_1} - \frac{a_2}{b_2} = -\frac{1}{b_1} \left( \frac{\partial D_2}{\partial Q_1} + \lambda \frac{\partial R_2}{\partial Q_1} \right) \quad (12)$$

Note that, by the monotonicity property, if  $Q_1$  increases while  $Q_2$  remains constant then both  $D_2$  and  $R_2$  will tend to increase. Thus,  $\frac{\partial D_2}{\partial Q_1} \geq 0$  and  $\frac{\partial R_2}{\partial Q_1} \geq 0$ . Because  $a_i \geq 0$  and  $b_i \leq 0$ , from (12)

$$\frac{a_1}{b_1} \geq \frac{a_2}{b_2} \Leftrightarrow \left| \frac{\Delta D_1}{\Delta R_1} \right| \leq \left| \frac{\Delta D_2}{\Delta R_2} \right|, \quad (13)$$

in words, at optimality, the slope of operating point in the  $R_1 - D_1$  characteristic is smaller than the slope of the operating in the  $R_2 - D_2$  characteristics. Note that to derive (13) we only had to make one assumption, namely that the monotonicity property holds.

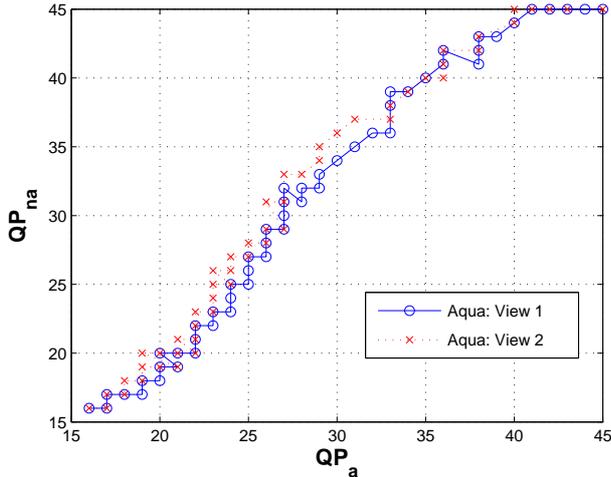
Given that the slope  $\left| \frac{\Delta D}{\Delta R} \right|$  of a convex  $RD$  decreases as  $Q$  decreases (i.e., as the coding quality improves), we can conclude that if  $R_1 - D_1$  and  $R_2 - D_2$  have similar shape, then from (13), at optimality  $Q_2 > Q_1$ . In our case of interest,  $Q_2$  is the quantizer used to encode several non-anchor frames. In this case  $\left| \frac{\Delta D_2}{\Delta R_2} \right|$  would be the slope of the aggregate  $RD$  characteristic. While the absolute values of  $R_2$  and  $D_2$  are likely to be larger than those for  $R_1$  and  $D_1$  at a given  $Q$ , the shapes of the curves and corresponding slopes can still be assumed to be similar. This approximation agrees well with our observed experimental behavior and provides a tool for complexity reductions.

## 2.3. Search Algorithm for Non-anchor Frames

Up to now, for simplicity, we have assumed that the same quantizer is used for all non-anchor frames. We now propose a non-anchor frame quantizer search algorithm, which operates for a given anchor frame quantizer, and to reduce complexity, uses the following property (based on the discussion of the previous section): a frame close to root of the dependency tree has more influence on cost and therefore a better quantizer should be applied to it. Thus we begin the search with the frame which is close to the root. In the following algorithm,  $\bar{q} = \{Q_2, Q_3, \dots, Q_M\}$ , is now the vector of quantizers allocated to the non-anchor frames in a given view.

*Algorithm 2: Dependent coding in each view*

- Given  $\lambda$  and the  $QP$  of anchor frame  $q$ , initialize  $\bar{q} = \{Q_2, Q_3, \dots, Q_M\} = \{q_0, q_0, \dots, q_0\}$ , where  $q_0 \leq q$ .
- For frames  $i = 2, 3, \dots, M$ , find  $\alpha_i = \frac{\partial J}{\partial Q_i} = \left( \sum_{j=i}^n \frac{\partial D_j}{\partial Q_i} \right) + \lambda \left( \sum_{j=i}^n \frac{\partial R_j}{\partial Q_i} \right)$ .



**Fig. 5.** Relationship between  $q$  ( $QP_a$ ) and  $\bar{q}$  ( $QP_{na}$ ) for Aquarium sequence.

3. - If  $\alpha_i < 0$ ,  $Q_i = Q_i + 1$ . Increase  $Q_j$  which is less than  $Q_i$  for  $j = \{i + 1, \dots, M\}$ .  
- If  $\bar{J}$  decreases update  $\bar{q}$ . Proceed to the next frame.
4. Repeat step 2 - 3 until there is no update in  $Q_i$

In this algorithm  $\alpha_i$  is calculated for the current  $\bar{q}$ . Then in order to make  $\alpha_i$  closer to 0, we increase by 1  $Q_i$  if  $\alpha_i < 0$ . Then, using the property motivated in the previous section, we also increase  $Q_j$  such that  $Q_j < Q_i$  for  $j > i$ .

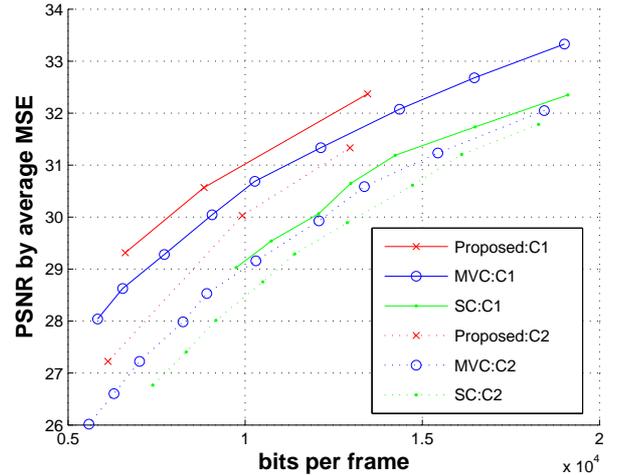
### 3. RESULTS AND CONCLUSIONS

Using H.264/AVC reference codec, JM7.6, we encoded the Aquarium multiview sequences from Tanimoto Lab using three different coding schemes, i.e., SC in 2(a), MVC in 2(b) with fixed QP and optimized QP by proposed Algorithm 1. In the experiment all non-anchor frames in a view were assigned the same quantizer. Two different coding conditions are used: First, all possible block sizes can be used and intra coding is enabled (C1). Second, only 8x8 block size is used and intra coding is disabled except I frame (C2). The first 7 frames of Views 1, 2, and 3 are used in the experiment. As can be seen in Fig. 6, the proposed algorithm provides a gain of 0.5 - 1dB as compared to MVC with fixed QP. In trellis expansion, six quantizers are selected as candidates for anchor and only three quantizers are selected for non-anchor frames using correlation between anchor and non-anchor quantizer.

2-D bit allocation scheme with reduced complexity is proposed. Complexity of data generation in trellis expansion is significant due to the dimensionality. We extend the monotonicity property from [3] and use the fact that optimal solutions tend to show correlation between quantizers of anchor and non-anchor frames. Also non-anchor frame quantizer search algorithm is introduced.

### 4. REFERENCES

[1] G. Chen, J.H. Kim, J. Lopez, and A. Ortega. Response to call for evidence on multi-view video coding. *M11731, ISO/IEC JTC1/SC29/WG11*, Jan. 2005.



**Fig. 6.** SC vs. MVC with fixed QP vs. MVC by proposed Algorithm 1. Average number of bits for 21 frames is used and PSNR is calculated as  $10\log_{10}(255^2/(average\ MSE\ for\ 21\ frames))$ .  $\lambda$  is 200, 500, and 900 with C1 and 300, 700, and 2500 with C2 in proposed algorithm.

[2] S. Liu and C.-C. J. Kuo. Joint temporal-spatial bit allocation for video coding with dependency. *IEEE Trans. Circuits and Systems for Video Technology*, 15(1), Jan. 2005.

[3] K. Ramchandran, A. Ortega, and M. Vetterli. Bit allocation for dependent quantization with applications to multiresolution and mpeg video coders. *IEEE Trans. Image Processing*, 3(5), Sep. 1994.

[4] Y. Sermadevi and S. S. Hemami. Efficient bit allocation for dependent video coding. In *Proc. of Data Compression Conference*, pages 232 – 241, Mar. 2004.

[5] Y. Shoham and A. Gersho. Efficient bit allocation for an arbitrary set of quantizers. *IEEE Trans. Acoustics, Speech, and Signal Processing*, 36(9), Sep. 1988.

[6] K. M. Uz, J. M. Shapiro, and M. Czigler. Optimal bit allocation in the presence of quantizer feedback. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 385 – 388, Apr. 1993.

[7] W. Woo and A. Ortega. Optimal blockwise dependent quantization for stereo image coding. *IEEE Trans. Circuits and Systems for Video Technology*, 9(6), Sep. 1999.