CODEBOOK ADAPTATION ALGORITHM FOR A SCENE ADAPTIVE VIDEO CODER

Dongmei Wang, and John Hartung

Signal Processing Research Department
AT&T Bell Labs
Murray Hill, NJ 07974

1. ABSTRACT

In this paper, we propose a codebook adaptation algorithm for very low bit rate, real-time video coding. Although adaptive codebook design has been studied in the past, its implementation at very low coding rates suitable for the MPEG4 standard remains significantly challenging. Our coder uses a standard motion compensated predictor with DCT quantization. It is unique in that it uses a hybrid scalar/vector quantizer to code predictor residuals. Bits are dynamically allocated to minimize distortion in the current frame, and scalar quantized blocks are used to adapt the VQ codebook. A codebook adaptation algorithm is described which uses an "equidistortion principle" and a competitive learning algorithm to continuously adapt the codewords. This training algorithm results in an increased use of the more efficient vector quantizer and improved video quality.

2. INTRODUCTION

The general purpose of video coding is to compress the video image sequence data for more efficient storage or transmission via a communication channel. From Shannon's rate-distortion theory, coding vectors, instead of scalars, gives better performance in theory when the vector quantizer is matched to the source statistics. A hybrid scalar/vector quantizer is proposed in our very low bit rate coder [1] to provide efficient quantization with changing signal statistics. The adaptive VQ schemes pursued before [2] explored the coding performance via parameters such as degree of adaptivity, vector dimensionality, and codebook size. The idea in our codebook adaptation algorithm concerns the tradeoff among the bits allocated to different types of information under fixed frame and bit rate constraints in our coder. As dynamic bit allocation allows us to optimize the coder performance for the current frame in the video sequence, codebook adaptation provides better coding quality for future frames.

In order to be implemented in real time, the complexity of the codebook adaptation algorithm must be constrained. At the same time, it must still provide optimal use of the codebook to maximize visual quality of the current frame, while adapting the codebook in order to improve the efficiency of the quantizer.

3. ALGORITHM

3.1. Analysis

The main idea in the adaptation algorithm is quite simple. To have a real time video coder, the traditional LBG codebook generation algorithm will not work. This is because we are unable to do extensive iterative learning in a large training set (e.g., all input image blocks) to find the best representative codewords for the current video sequence. Instead, we can only use scalar quantized blocks as training vectors for adaptation as new image frames arrive.

In order to stay within the bit rate limit of each frame, a dynamic bit allocation algorithm is used to achieve the best performance for the current frame by allocating bits between the scalar and vector quantizers. This results in the transmission of scalar quantized blocks, which are used to adapt the codebook for future frames. These scalar blocks, along with the quantizer statistics for previous frames, are used to adapt the codebook such that a larger proportion of bits are allocated to the more efficient vector quantizer for minimum coding distortion. This results in improved coding quality because of the increased use of vector quantization.

From Gersho's book [3], there are two Lloyd conditions that are necessary for an optimal MMSE quantizer: (1) the codebook vectors should be the centroids of the partitions of the vector space, and (2) the

0-7803-2431-5/95 $4.00

boundary between two adjacent partitions should be equidistant from the centroids of the partitions. These partitions are called Vonoroi regions when using the minimum distortion rule, and each Voronoi region is represented by a codeword. Applying these two conditions iteratively in codebook generation will result in "equiprobable" characteristics, that is, the probability of an input vector falling into a Voronoi region is the same for each region. A VQ codebook that satisfies these two necessary conditions will be locally optimal, which means that the performance obtained strongly depends on the initialization of the representative codewords to be optimized. Ueda and Nakano [4] derived an optimal codebook generation criterium, using an "equidistortion" principle, based on Gersho's [5] asymptotic performance theory of VQ. It says that for a large number of representative vectors, the expected distortion becomes minimum if and only if every partitioned region has the same minimum subdistortion. Our codebook adaptation algorithm attempts to satisfy this objective.

Thus, there are two basic objectives for our design. The first is to design an algorithm that approaches global optimality. The second is to develop a real time algorithm for the continuous adaptation of codewords based on new coding information.

The following section describes an adaptation algorithm which uses a split and merge algorithm to introduce new codewords into the codebook, and a competitive learning algorithm to continuously adapt the codebook.

3.2. Design

Based on the above analysis, we developed a real-time codebook adaptation algorithm as follows.

We define the codebook size as N, the codeword size as M, and assume that we start from an empty initial codebook. For the first several frames of the video sequence, while the number of different scalar (training) blocks is less than N, we split the Voronoi regions until the codebook has N codewords. We then adapt the codebook by splitting and merging regions based on their subdistortions, or apply competitive learning if a new codeword does not reduce the overall distortion.

We represent a codeword by $W_i(n)$, and Voronoi region distortion associated with $W_i(n)$ by $MSE_i(n)$, where i is the index of the codeword in the codebook, and n is the frame index in the video sequence. An input scalar block is represented by X. When a scalar block X is encoded using the vector quantizer, it will be associated with the Voronoi region whose centroid represents the block with minimum mse, and will contribute to the accumulated distortion of this region. The decision factor $d_i(n)$ indicates whether the input vector X belongs to ith Voronoi region.

$$d_i(n) = \begin{cases} 1, & if ||W_i(n) - X||^2 < ||W_j(n) - X||^2 \\ 0, & otherwise \end{cases} \quad (1)$$

$$MSE_i(n+1) = MSE_i(n) + ||W_i(n) - X||^2 \quad (2)$$

A Voronoi region is a candidate for splitting, i.e., a new scalar block associated with that region is added to the codebook, when the block coding distortion is large, and the region distortion compared to the distortion in other regions is large. In order to add an additional codeword, other Voronoi regions must be merged. A new codeword is added, and the two codewords having smallest distortion are merged, if the distortion of the merged region is not greater than the average region distortion. This will drive the Voronoi regions to have equal distortion.

If we do not meet the requirement to split and merge codewords using the scalar blocks, a competitive learning algorithm is used to update the current codewords in order to minimize their distortion.

The codewords are updated according to the following update rule.

$$W_i(n+1) = W_i(n) + l_i(n)[X - W_i(n)] \quad (3)$$

$l_i(n)$ is the learning factor [6], which determines the updating codeword's convergence rate. There are multiple ways to define this factor. One simple way is to use $\exp^{-\alpha n}$, $0 < \alpha < 1$. In order to consider the contribution of the neighbor to the Voronoi region being updated, other learning factors have also been studied [7]. For example, the learning factor is defined in [8] as:

$$l_i(n) = \begin{cases} 1, & if ||W_i(n) - X||^2 - B_i(n) < \\ & ||W_j(n) - X||^2 - B_j(n) \\ 0, & otherwise \end{cases} \quad (4)$$

where $B_i(n)$ is a bias form, and $R_i(n)$ is winning rate:

$$B_i(n) = C(1/N - R_i(n)), \quad C < 1 \quad (5)$$

$$R_i(n+1) = R_i(n) + A[d_i(n) - R_i(n)], \quad 0 < A < 1 (6)$$

This competitive learning changes the shape of the Voronoi region and thus changes the distortion of each region also. After adapting the codebook for each scalar encoded block the distortion is recalculated. To reduce the complexity in calculating the new distortion, input scalar blocks very close to one of the existing codewords are not used in training the codebook.

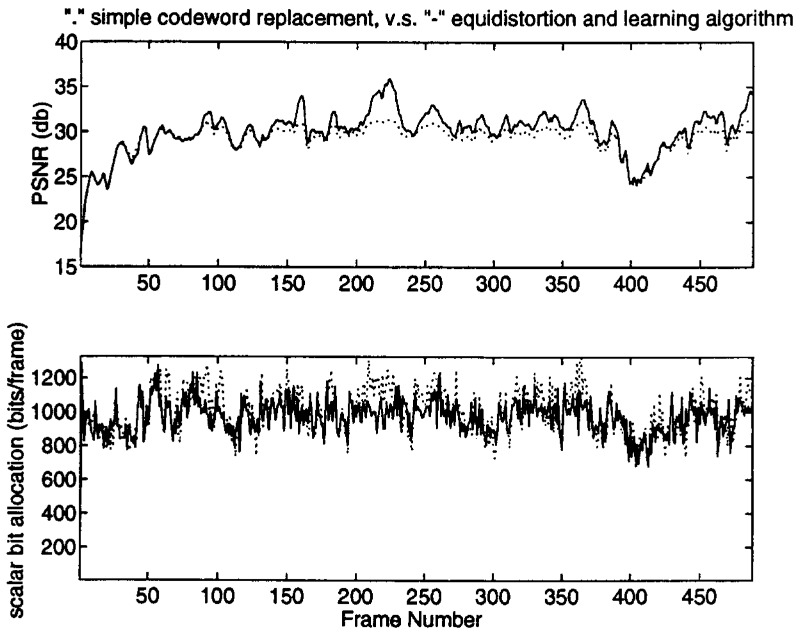

Figure 1: Results

4. RESULTS

Codebook adaptation increases use of the vector quantizer and therefore the decoded image quality. A constant bitrate is maintained, while more bits are allocated to scalar blocks, which are used to train the VQ codebook, when it is not well adapted to the sequence. This is illustrated in Figure 1, which shows results for a sequence coded at a constant 10 frames per second and 16 kbps. Here we compare a simple codeword replacement algorithm with the adaptation algorithm based on the equidistortion principle and learning. As shown in the plots, the PSNR improves during training, while the rate allocated to scalar blocks decreases. This is due to codebook adaptation and subsequently greater allocation of bits to the vector quantizer.

Similarly, as shown for representative frames in Table1, the PSNR increased in the new scheme and the bits allocated to the scalar quantized image blocks dropped overall.

Frame	PSNR		Scalar Bit Allocate	
Number	old(dB)	new(dB)	old(bits)	new(bits)
1	17.161	17.161	4	4
95	30.572	31.664	1040	1001
220	31.307	35.195	1059	902
345	30.338	31.766	1157	1045
485	31.167	34.324	1164	1035

Table 1: Comparison of Performace

5. CONCLUSION

Our hybrid scalar/vector quantizer based very low bit rate coder is designed for real time implementation on current video DSP chips. It is unique in that it uses simple dynamic bit allocation and codebook adaptation to gain the optimal individual frame and overall video sequence perceptual visual quality. A formal subjective study was done to compare this algorithm to more standard MCDCT algorithms. Stronger preference was found for the new algorithm.

6. REFERENCES

[1] J. Hartung, and T. Michel, "Constant Frame Rate Scene Adaptive Video Coder for Very Low Bit Rate Applications," *Submitted to Image Communication - Special Issue on Coding Techniques for Very Low Bit Rate Video*, Aug. 1994.

[2] M. Goldberg, P. R. Boucher, and S. Shlien, "Image Compression Using Adaptive Vector Quantization," *IEEE Trans. Commun.*, vol. 34, pp. 180–187, Feb. 1986.

[3] A. Gersho, and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Boston, 1992.

[4] N. Ueda, and R. Nakano, "A New Learning Approach Based on Equidistortion Principle for Optimal Vector Quantizer Design," *1993 IEEE Workshop on Neural Networks for Signal Processing*, pp. 362-371.

[5] A. Gersho, "Asymtotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. 25, pp. 373–380, 1979.

[6] T. Kohonen, *Self-Organization and Associative Memory*, Springer-Verlag, Berlin, 1984.

[7] S. C. Ahalt, A. K. Krishnamurthy, P. Chen, and D. E. Melton, "Competitive Learning Algorithms for Vector Quantization," *Neural Networks*, vol. 3, pp. 277–290, 1990.

[8] D. DeSieno, "Adding a Conscience to Competitive Learning, " *IEEE International Conference on Neural Networks*, pp. 1117–1124, 1988.

[9] N. Jayant, J. Johnston, and R. Safranek, "Signal Compression Based on Models of Human Perception," *Proceedings of The IEEE*, vol. 81, no. 10, Oct. 1993.