EFFICIENT CODING OF DCT COEFFICIENTS BY JOINT POSITION-DEPENDENT ENCODING

Eric C. Reed and Jae S. Lim

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139 USA
ereed@image.mit.edu

ABSTRACT

In a typical MC-DCT encoding scheme, a large portion of the bit rate is used to encode the location and amplitude information of the nonzero quantized DCT coefficients. Therefore efficient encoding of the DCT coefficients is extremely important. In this paper we describe the Joint Position-Dependent Encoding (PDE) approach to encode the DCT coefficients. Joint PDE exploits the variations in statistical properties of the runlengths and amplitudes as a function of position by introducing a set of 2-D codebooks in which each quantized DCT coefficient is assigned to one codebook in the set based on its location. Utilizing an MPEG-2 codec, we compare the bit rates using the joint PDE variable length codes (VLC's) with the bit rates produced by the MPEG-2 VLC's. We also examine how performance is affected by the number of codebooks.

1. INTRODUCTION

For every nonzero quantized DCT coefficient within a block, both location and amplitude information need to be encoded. In typical video and image compression scenarios, most of the DCT coefficients in a block are quantized to zero. Runlength encoding [3] is one widely-used and effective method to represent the location information that takes advantage of the large number of coefficients quantized to zero. This approach orders the DCT coefficients into a 1-D vector (such as through a zigzag scan) and then encodes the runlengths which are defined as the number of zero coefficients between consecutive nonzero coefficients. When a nonzero coefficient is found, its amplitude along with the corresponding runlength is encoded. The runlengths and amplitudes can be entropy encoded in one of two ways. They can be encoded separately requiring two distinct 1-D codebooks or they can be encoded jointly with one 2-D codebook. In the joint encoding approach, each (runlength, amplitude) pair is treated as a single event and thus introduces a significantly larger number of possible outcomes compared to the separate encoding approach.

The MPEG-2 standard [2] uses the joint encoding approach and includes 2 Huffman codebooks in order to exploit the differences in statistics of the runlengths and amplitudes among intra (original image) and inter (MC - prediction residual) regions. However, the statistical properties of the runlengths and amplitudes also vary with position (frequency). Since each of the MPEG-2 codebooks are trained from a collection of all the intra or inter coefficients within a block, these codebooks do not exploit the statistics of the runlengths and amplitudes as they vary with position. The Position-Dependent Encoding (PDE) approach exploits these variations by introducing multiple Huffman codebooks where each codebook is trained only from the runlength and amplitude statistics of the coefficients assigned to that particular codebook. This allows the entropy coder to efficiently exploit the pertinent statistics of each DCT coefficient. Just like conventional runlength encoding, PDE can be applied in a separate or joint encoding environment.

Position-Dependent Encoding has been introduced [1, 4] where a separate encoding approach is taken in which two sets of multiple 1-D Huffman codebooks are designed. In their work, a non-standard MC-DCT coder is used and the results illustrate the potential of the PDE approach as an effective method for bit rate reduction. In this paper, we take a joint PDE approach in which one set of 2-D Huffman codebooks are designed [5]. Since joint PDE offers the benefit of exploiting the correlation between short runlengths and large amplitudes and also between long runlengths and small amplitudes, we should expect an increase in performance over the separate PDE approach. In addition, we use a full MPEG-2 codec to observe how joint PDE performs in this environment. Since MPEG-2 allows Intra (I), Predicted (P), and Bi-directionally predicted (B) frames, we observe how joint PDE performs for each frame type separately. The MC-DCT coder used in the previous work does not include B-frames. Furthermore, we consider how joint PDE performs in a practical setting. Thus, in order to emulate the training in an actual coding environment, we assume that the test sequence to be encoded is unknown and therefore the training set does not contain the statistics of the test sequence. In the previous work, the statistics of the test sequence are included in the training set. Also, since it is undesirable to have a large set of codebooks due to increased complexity and storage, we examine joint PDE performance with a reduced set of codebooks.

This research was supported by the MIT Advanced Television Research Program and by a DoD fellowship.

2. JOINT POSITION-DEPENDENT ENCODING

Based on our knowledge of the encoder and the source, it is possible to predict the behavior of the runlength and amplitude statistics. For example, in MPEG-2 quantization is usually performed after a weighting matrix is applied to the DCT block that favors low frequency coefficients over high frequency coefficients. It is also widely known that most video sequences contain predominately low frequency content. Based on these two observations, most of the nonzero quantized DCT coefficients are concentrated in the low frequency region with larger amplitudes compared to the amplitudes of the small number of nonzero quantized DCT coefficients in the high frequency region. This property is less evident for inter blocks where the amplitudes of the DCT coefficients are more uniformly spread throughout the block and depend on the characteristics of the MC-residual signal. In blocks where the nonzero coefficients are concentrated in the low frequency region, the runlengths beginning in the low frequency region are shorter than the runlengths beginning in the high frequency region. The joint encoding approach has the advantage of exploiting the correlation that exists between short runlengths and large amplitudes and also between long runlengths and small amplitudes.

In addition, the range of the possible runlengths varies with the starting position of the run. Clearly, the range of the runlengths decreases as the starting position of a nonzero quantized DCT coefficient increases within a block. Since MPEG-2 uses 8x8 block sizes, the runlength range starting at position (0,0) within a block (DC) is 0 to 63. Therefore, 6 bits are needed to represent all runlength possibilities. However, as we travel through the block starting at DC and follow some specified scanning pattern, the number of coefficients remaining in the block decrease thus decreasing the runlength range. Therefore, using 6 bits to represent the runlengths for nonzero mid and high frequency coefficients is unnecessary. This can lead to a considerable advantage if there are many nonzero mid to high frequency coefficients, which is the case for inter coded blocks.

Since joint PDE introduces multiple Huffman 2-D codebooks, the number of codebooks needed to achieve the benefits of this approach is an important issue for practical considerations. Fortunately, since neighboring coefficients tend to have similar runlength and amplitude statistics, coefficients can be grouped together and share a codebook. We can reduce the undesirable effect that grouping has on total performance by discriminating against less important coefficients. In other words, we can group many of the less important coefficients together while allowing more important coefficients to have their own codebooks. For example, since the luminance coefficients take up more of the bit rate than the chrominance coefficients, it makes sense to separate the statistics between the luminance and chrominance blocks and assign more codebooks to luminance blocks. Also, since there are less nonzero high frequency coefficients, we can group many of these coefficients together, especially for intra coefficients, without significantly affecting the performance of the joint PDE approach. An example assignment of four codebooks for any given block type is shown in Figure 1.

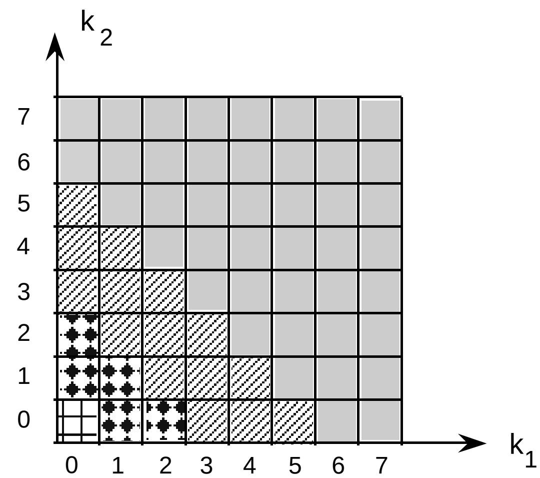

Figure 1: An example assignment of 4 codebooks within a 8x8 block. Coefficients that share codebooks have identical patterns.

3. COMPARISON WITH MPEG-2

In this section, we investigate the performance of the joint PDE approach and present the results. Our experiments use an MPEG-2 encoder operating within the high level and main profile specifications. The details of the specification can be found in [2]. I-frames are periodically inserted every 13 frames (N=12) and 2-B frames are inserted between I- and P-frames (M=3). MPEG-2 default matrices are used for coefficient quantization along with 4:2:0 chrominance sampling. Runlengths are defined based on the zigzag scan pattern and the DCT coefficients are uniformly quantized for an amplitude range of 0 to 1023 (i.e. mquant ranges from 2 to 62). Intra DC coefficients are not coded using PDE and are coded normally by MPEG-2.

In order to design the joint PDE Huffman codebooks, the runlength and amplitude statistics are collected separately for every coefficient within the following block types: I- luminance (Y), I- chrominance (UV), P-Y, P-UV, B-Y, and B-UV. This corresponds to a total of 382 different sets of statistics: 63 for I-Y coefficients, 63 for I- UV coefficients, 64 for P-Y coefficients, 64 for P-UV coefficients, 64 for B-Y coefficients, and 64 for B-UV coefficients. This segmentation supports a maximum of 382 codebooks and offers the freedom to reduce the number of codebooks to any desired level by appropriate grouping of the statistics.

The runlength and amplitude statistics are collected from 14 progressively scanned video sequences. Eight of these are 60 Hz camera video sequences (test sequences 1-4,7-10), another four are 24 Hz film sequences (test sequences 11-14), while two (test sequences 5-6) are sequences that were synthetically generated (zoom and pan) at MIT. The statistics are collected from a total of 13 frames of each video sequence. This corresponds to one full group-of-pictures plus one additional I-frame in the following group-of-pictures.

Each of the 14 video sequences is encoded with MPEG-2 at a target bit rate of 0.34 bits per pixel (bpp) using the rate control specified in Test Model 5. The most bits are allocated for I-frames and the least bits are allocated for B-frames. Then the MPEG-2 VLC's are replaced with our joint PDE VLC's and the bit rates are compared. In order to maintain the same quantized DCT coefficients for a fair comparison, the same mquant parameters derived from the rate control using the MPEG-2 VLC's are used for the joint PDE VLC's.

In the following experiments, each test sequence has a set of codebooks that is trained from the statistics of the remaining 13 test sequences. Therefore, there are a total of 14 different sets of codebooks, one for each test sequence. This scenario assumes the test sequence is unknown and thus its runlength and amplitude statistics are not included in the training set.

3.1. Upper Bound on Performance

In this experiment, each coefficient is given its own codebook. Even though this may lead to an impractical number of codebooks, this scenario provides an upper bound on joint PDE performance and can be used as a reference for a more practical scenario. Since MPEG-2 allows three separate frame types with each serving different purposes, we allocate separate codebooks for I-, P-, and B-blocks within which we also separate the luminance and chrominance blocks. As mentioned above, this corresponds to a total of 382 codebooks. Results are presented in Figures 2, 3 and 4 for the first 10 test sequences. The average decrease in bit rate for I-, P-, and B-blocks is 7.1 %, 14.4 %, and 22.8 %, respectively. The total bit rate using joint PDE is reduced by 9.5 % on average.

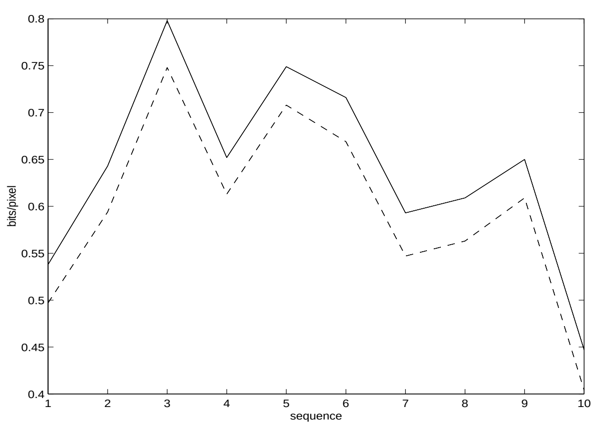

Figure 2: The required bpp using MPEG-2 (solid) and joint PDE (dashed) VLC's for I-blocks using 382 codebooks.

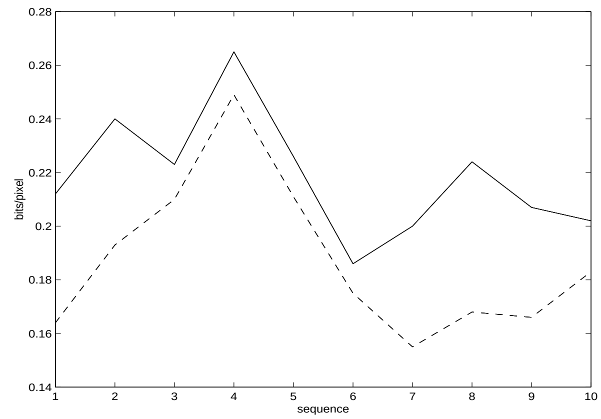

Figure 3: The required bpp using MPEG-2 (solid) and joint PDE (dashed) VLC's for P-blocks using 382 codebooks.

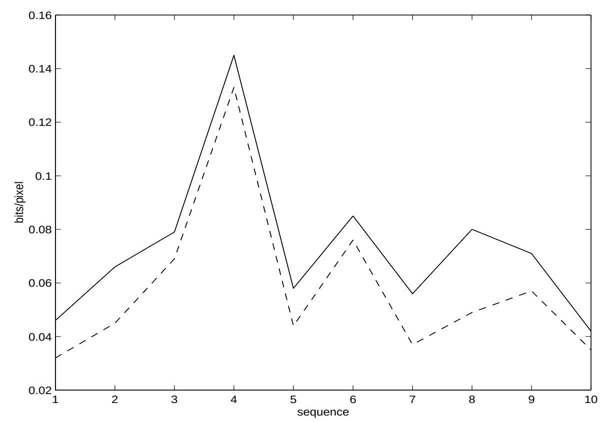

Figure 4: The required bpp using MPEG-2 (solid) and joint PDE (dashed) VLC's for B-blocks using 382 codebooks.

3.2. A Practical Example

The results presented here are what would be observed in a more practical setting. The total number of codebooks are reduced from 382 to 31: 8 Intra Y, 3 Intra UV, 13 Inter Y, and 7 Inter UV. In this particular segmentation, the statistics of the P- and B- blocks are combined to form one set of inter blocks. In addition, many of the high frequency coefficients are grouped together to share a codebook. Furthermore, escape codes are introduced by limiting the maximum allowable codeword length to 13 in each of the 31 codebooks. This reduced the average size of the set of codebooks to approximately 114 codewords. For comparison, MPEG-2 uses 2 codebooks with 114 entries each where we introduce 31 codebooks with an average number of 114 entries. Escape codes were not used in the experiments presented in section 3.1.

More codebooks have been assigned to inter blocks since they occupy more of the bit rate and also since joint PDE performs significantly better for inter blocks compared to intra blocks (section 3.1). Also, each intra block has one less codebook since intra DC coefficients are not coded using PDE while inter DC coefficients are given their own codebooks. For similar reasons, more codebooks are assigned to luminance coefficients. The results are summarized in Fig-

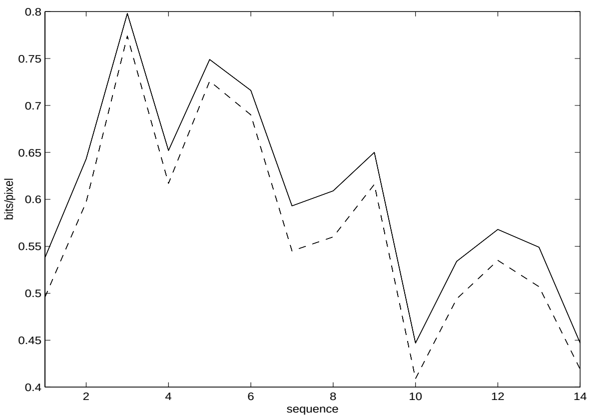

Figure 5: The required bpp using MPEG-2 (solid) and joint PDE (dashed) VLC's for Intra blocks using 31 codebooks.

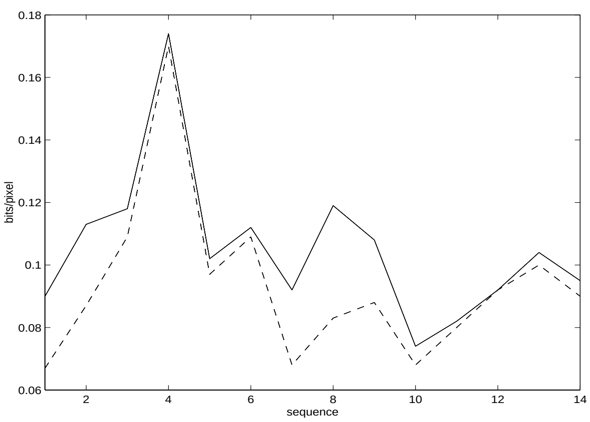

Figure 6: The required bpp using MPEG-2 (solid) and joint PDE (dashed) VLC's for Inter blocks using 31 codebooks.

ures 5, 6 and 7. The overall percentage decrease in the bit rate is reduced from 9.5 % (using 382 codebooks) to only 7.6 %, a difference of less than 2 %. The maximum decrease in total bit rate is 16.5 % and occurs with test sequence 8. Similarly, the minimum decrease is 2.4 % and occurs with test sequence 12. The average decrease in bit rate for intra blocks is 6.7 % and the average decrease for inter blocks is 12.4 %.

4. COMMENTS

The results show that the joint PDE VLC's outperform the MPEG-2 VLC's in all cases and that by using 31 2-D Huffman codebooks, we can achieve an average decrease of 7.6 % in total bit rate. In addition, our results also suggest that the joint PDE approach performs better than the separate PDE approach where a 5.8% average decrease in bit rate is reported using 94 total runlength codebooks and 14 total amplitude codebooks [1, 4].

The joint PDE approach can be used as an effective

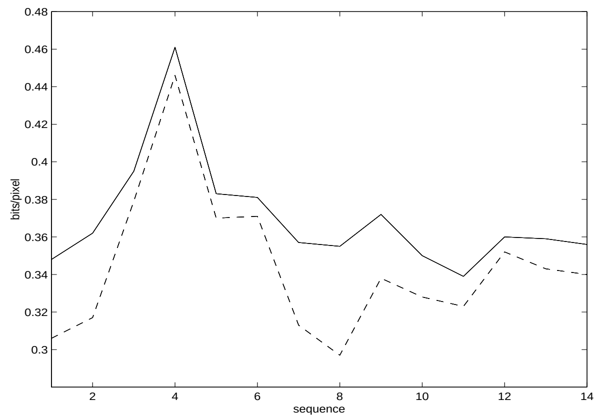

Figure 7: The total bpp using MPEG-2 (solid) and joint PDE (dashed) VLC's using 31 codebooks.

method to reduce bit rates of video sequences transmitted over limited bandwidth channels and also as an efficient method to store video sequences. In a different setting, the bits saved by the PDE approach can be used to increase the quality of the coded video. Even though we use an MPEG-2 encoder in this work, PDE can be applied to other image and video compression standards using any transform/subband filtering scheme.

In addition, the PDE approach can be tested over a wide range of bit rates. From section 3.1 and the fact that B-frames are allocated the least number of bits, the results suggest that a negative correlation exists between allocated bits and effectiveness of the joint PDE scheme. It would be interesting to investigate how joint PDE performs at very low bit rates (10 - 64 kb/s) using the H.263 standard. However, at low bit rates, overhead (such as the motion vectors) occupies a larger percentage of the bit rate and therefore may diminish the overall performance of the PDE approach.

Acknowledgements: The authors would like to thank members in the Advanced Television Research Program at MIT and especially John Apostolopoulos for many insightful discussions.

5. REFERENCES

[1] J. G. Apostolopoulos, A. Pfajfer, H. M. Jung, and J. S. Lim. Position-Dependent Encoding. In *ICASSP Proceedings*, volume 5, pages 573–576, 1994.

[2] ITU-T Rec. H.262. *Generic Coding of Moving Pictures and Associated Audio*, March 1996. Standard Draft.

[3] Jae S. Lim. *Two-Dimensional Signal and Image Processing*. Prentice Hall, Englewood Cliffs, New Jersey, 1990.

[4] A. Pfajfer. Position-Dependent Encoding. Master's thesis, M.I.T, May 1994.

[5] E. Reed. Improvement of MPEG-2 Compression by Position-Dependent Encoding. Master's thesis, M.I.T, February 1996.