A LAYERED VIDEO CODING SCHEME WITH ITS OPTIMUM BIT ALLOCATION

R. Atta and M. Ghanbari

Audio & Video Networking Research Lab
Dept. of Electronic Systems Engineering
University of Essex, Colchester, UK
Emails: {reeatt, ghan} @essex.ac.uk

ABSTRACT

In this paper, we introduce the DCT pyramid as a layered video coding technique. For efficiency of coding, we propose an optimum bit allocation method for the layers of the DCT pyramid. The proposed method is based on minimization of the overall reconstructed error with the Lagrangian multiplier and distribution of the bit budget among the layers. Experimental results on the improvement due to optimum bit allocation are also presented.

1. INTRODUCTION

Layered video coding is a technique that can provide an interworking capability between different types of video services and the network constraints in a hierarchical structure. The structure consists of the base layer constituting the lowest resolution service with a set of enhancement data to provide higher resolution services.

In the past decade, several layered coding algorithms have been devised. Most well known methods are a combination of SNR, spatial and hybrid scalability in the MPEG-2, MPEG-4 and H.263 standard video codecs. Another set is based on the wavelet coding, that depending on the nature of coding of the wavelet coefficients, result in SNR or spatial scalability. Here we concentrate on a third method, known as the DCT pyramid [1] as explained in section 2. One of the fundamental aspects in layered coding which is particularly important for DCT pyramid is the allocation of the bit rate budget to various layers. Here the main aim of this paper is the derivation of an optimum bit allocation algorithm to the layers of the DCT pyramid.

Bit allocation is a classical problem in a source coding which has been extensively studied in the literature [2], [3]. It is more commonly related to the rate-distortion theory in minimizing the distortion for a given bit budget. Modeling of the rate and distortion characteristics has been frequently used within these schemes [4-6]. The rate control problem for layered video coding can be separated into three parts: 1) to allocate target bits for each frame according to the complexities of each frame and buffer fullness for a given constant channel rate, 2) to distribute the allocated target bits among the layers, 3) to select the quantization parameters for each macroblock in each layer.

In this paper, we use a three-layer DCT pyramid video codec, which generates three bitstreams one for the base layer and two for the reversed L-shapes in the enhancement layers [1]. We also derive a mathematical expression for the overall reconstructed error of the coding scheme based on the mean square error (MSE). The analytical solution of the bit allocation problem to distribute the target bits among the layers of the scheme is proposed. This analytical solution is based on minimizing the overall reconstructed error by using the Lagrange multiplier.

This paper is organized as follows: In section 2, a brief description of the DCT pyramid video coding scheme is presented. In section 3, a mathematical expression for the overall final reconstruction error of the coding scheme is derived, as well as the optimum bit allocation technique is presented. Experimental results are reported in section 4. Finally, section 5 concludes the paper.

2. DCT PYRAMID VIDEO CODING SCHEME

The first step towards achieving a Layered video coding is to design an efficient multilayer structural coding scheme. A brief description of the DCT pyramid video encoder proposed in [7] is given below:

Decimation process: In the DCT pyramid, the full spatial resolution future frame is decimated into a series of lower spatial resolution pictures as shown in Fig. 1. In the DCT decimation, the full spatial resolution input video is first transformed by a forward 8x8 DCT and that is followed by a 4x4 inverse DCT transform of the lower 4x4 frequency components of each transformed 8x8 resolution block. The result is a decimated version of the full resolution video input, which is reduced by a factor of 2 in both the horizontal and vertical directions. In case of three layers, the base layer receives input with a spatial resolution reduced by a factor of 16 (i.e., a factor

0-7803-7663-3/03/$17.00 ©2003 IEEE

of 4 in both horizontal and vertical directions) and is encoded like the conventional standard codecs.

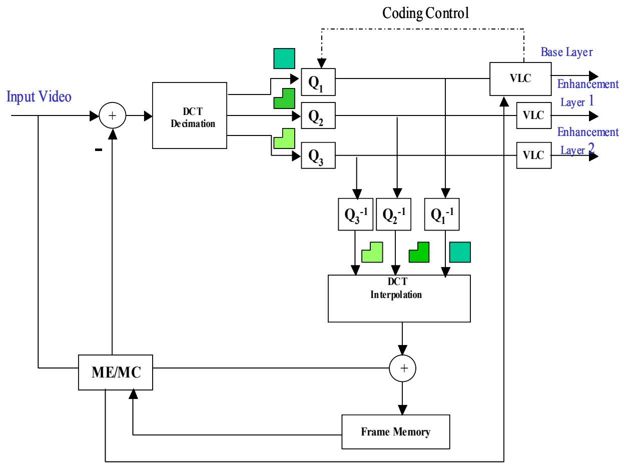

Fig. 1. A three-layer DCT pyramid video encoder.

Enhancement layers: These layers receive the higher spatial frequency components (reversed L-shapes) obtained from the decimation process rather than receiving the higher spatial resolutions. The base and enhancement layers are independently quantized and entropy coded before being transmitted.

Interpolation process: The base layer is first locally decoded to produce a lower spatial resolution of the original input. The reconstructed version of the base layer is then used to reconstruct the next higher resolution of the input video signal by using the DCT interpolation method [1]. In the same manner, the third-layer is reconstructed to generate the full resolution of the input video sequence. In DCT interpolation, the higher spatial resolution in the second layer is reconstructed by performing 4x4 forward DCT on the reconstructed base layer. Then, those coefficients are padded with the higher frequency coefficients (the reversed L-shape) to generate 8x8 coefficients in this layer. Finally, an inverse 8x8 DCT is performed. This procedure is continued to reconstruct the full resolution video sequence in the third layer.

3. MULTI-LAYER RATE CONTROL

In DCT pyramid video coding, the current frame to be encoded is decomposed into super macroblocks of 32x32 pixels, and each super macroblock is decimated into a 8x8 block in the base layer, four and sixteen reversed L-shape blocks in the first and second enhancement layers respectively, which are then variable-length coded.

Let B_{0ij}, B_{1ij}, and B_{2ij} be the number of bits produced in the ith block of the base and enhancement layers respectively in the jth frame and D_{0ij}, DL_{1ij}, and DL_{2ij} are their respective distortions. The distortion is measured in terms of the MSE between the original and the encoded macroblock. In the next section, we derive a mathematical expression for the overall reconstructed error of the coding scheme based on the MSE.

3.1. The MSE Function

The overall coding distortion is comprised of the distortions in the base layer blocks, and in the reversed L-shapes of the enhancement layers. The mean square error can be calculated either in the DCT or in the pixel domain. Let D_{0ij} be the MSE of the ith 8x8 block of the DCT coefficients in the base layer and D_{1ij} be the MSE of the reconstructed ith 16x16 larger block in the second layer, which is given by:

$$D_{1ij} = \frac{1}{16 \cdot 16}[64 \cdot D_{4x4} + 4 \cdot 48 \cdot DL_{1ij}] \qquad (1)$$

where DL_{1ij} is the mean square error of four reversed L-shapes in the second layer, each having 48 coefficients, and D_{4x4} is the MSE of four 4x4 blocks and is given by:

$$D_{4x4} = 4 . D_{0ij} \qquad (2)$$

In (2), the reconstructed pixel values (8x8 block) in the base layer are scaled by a factor of two before being forward-transformed by a 4x4 forward DCT. So D_{4x4} needs to be multiplied by a factor of two, and thus the mean square error is multiplied by a factor of four. By combining (1) and (2), we have

$$D_{1ij} = \frac{1}{16 \cdot 16}[4 \cdot 64 \cdot D_{0ij} + 4 \cdot 48 \cdot DL_{1ij}] \qquad (3)$$

then

$$D_{1ij} = D_{0ij} + \frac{3}{4} \cdot DL_{1ij} \qquad (4)$$

using (4), the MSE of the reconstructed ith 32x32 super macroblock in the third layer D_{2ij} becomes

$$D_{2ij} = D_{0ij} + \frac{3}{4} \cdot DL_{1ij} + \frac{3}{4} \cdot DL_{2ij} \qquad (5)$$

where DL_{2ij} is the mean square error of sixteen reversed L-shape blocks in the third layer. Equation (5) is the mathematical expression of total MSE at the super macroblock-level. The overall distortion in the jth frame D_j is then given by:

$$D_j = \sum_{i=0}^{N-1} (D_{0i} + \frac{3}{4} \cdot DL_{1i} + \frac{3}{4} \cdot DL_{2i}) \qquad (6)$$

where N is the number of super macroblocks in the frame.

3.2. Bit-Distortion Model and Optimum Bit Allocation

An analytic model for the rate-distortion is highly desired. Usually, such a model is given in terms of the variance of the input sequence, the number of bits produced due to quantization, and some parameters that depend on the distribution of the coefficients. We define the relationship between *Bij* and D_{ij} for each layer of the DCT pyramid according to [5] and [6]:

$$D_{0ij} = \frac{A_0\ K_{0j}\ \alpha_{0ij}^2\ \sigma_{oij}^2}{12(B_{0ij} - A_0 C_{0j})}$$
$$DL_{1ij} = \frac{A_1\ K_{1j}\ \alpha_{1ij}^2\ \sigma_{1ij}^2}{12(B_{1ij} - A_1 C_{1j})} \qquad (7)$$
$$DL_{2ij} = \frac{A_2\ K_{2j}\ \alpha_{2ij}^2\ \sigma_{2ij}^2}{12(B_{2ij} - A_2 C_{2j})}$$

where A_0, A_1, and A_2 are the number of pixels in the blocks in the base and in the enhancement layers, B_{0ij}, $B_{1ij,}$ and B_{2ij} are the number of bits produced by encoding the ith block in the base and ith blocks in the enhancement layers in a jth frame and D_{0ij}, DL_{1ij}, and DL_{2ij} are their distortions induced by quantization of these blocks, σ_{0ij}^2, σ_{1ij}^2, and σ_{2ij}^2 are the variance of the ith block in the base and enhancement layers respectively, K is a parameter that depends on the encoder's coding efficiency and frame pixels' distribution, and C is the overhead rate (bits per pixel) of the motion and syntax in the frame.

Given the target number of bits for each frame T_j in the sequence, our objective is to find the optimum number of bits for the base and the enhancement layers T_{0j}^*, T_{1j}^*, and T_{2j}^*, that minimize the overall distortion subject to the bit budget constraint. The problem can be written in the form:

$$\text{minimize} \quad \sum_{i=0}^{N-1} D_{0ij} + \frac{3}{4} DL_{1ij} + \frac{3}{4} DL_{2ij} \qquad (8)$$
$$\text{Subject to} \quad T_{0j} + T_{1j} + T_{2j} = T_{\mathrm{j}}$$

We can find a unique solution using the Lagrange multiplier method to solve this optimization problem of (8) in its equivalent unconstrained form:

$$T_{0j}^*, T_{1j}^*, T_{2j}^* = \min \sum_{i=0}^{N-1} \frac{A_0\ K_{0j}\ \alpha_{0ij}^2\ \sigma_{oij}^2}{12(B_{0ij} - A_0 C_{0j})} + \frac{3}{4} \cdot \frac{A_1\ K_{1j}\ \alpha_{1ij}^2\ \sigma_{1ij}^2}{12(B_{1ij} - A_1 C_{1j})} + \frac{3}{4} \cdot \frac{A_2\ K_{2j}\ \alpha_{2ij}^2\ \sigma_{2ij}^2}{12(B_{2ij} - A_2 C_{2j})} + \lambda \left(T_j - \sum_{i=0}^{N-1} (B_{0ij} + B_{1ij} + B_{2ij}) \right) \qquad (9)$$

by setting partial derivatives to zero in (9), we obtain the following expression for the optimised target number of bits for each layer:

$$T_{0j}^* = \frac{\sqrt{A_0 K_{0j}}\ \ S_{0j}\ \left[T_j - N \sum_{m=0}^{2} A_m C_{mj} \right]}{\sqrt{A_0 K_{0j}}\ \ S_{0j} + \sqrt{\frac{3}{4}} \left(\sqrt{A_1 K_{1j}}\ \ S_{1j} + \sqrt{A_2 K_{2j}}\ \ S_{2j} \right)} + A_0 N C_0$$

$$T_{1j}^* = \frac{\sqrt{\frac{3}{4}} \sqrt{A_1 K_{1j}}\ \ S_{1j}\ \left[T_j - N \sum_{m=0}^{2} A_m C_{mj} \right]}{\sqrt{A_0 K_{0j}}\ \ S_{0j} + \sqrt{\frac{3}{4}} \left(\sqrt{A_1 K_{1j}}\ \ S_{1j} + \sqrt{A_2 K_{2j}}\ \ S_{2j} \right)} + A_1 N C_1$$

$$T_{2j}^* = T_j - T_{0j}^* - T_{1j}^* \qquad (10)$$

where $S_{0\mathrm{j}}$, $\mathrm{S}_{1\mathrm{j}}$, and $\mathrm{S}_{2\mathrm{j}}$ are the sum of the variance of the block's DCT prediction error in the base layer and the variance of the reversed L-shape blocks in the enhancement layers.

The formula in (10) is the key for allocating a target number of bits for the base and enhancement layers. Once the target number of bits for each frame in the sequence is known, the bit allocation for each layer can be done independently and hence, the TMN8 rate control [5] can be used for each layer to select the quantization step size for all blocks in the base and the reversed L-shapes in the enhancement layers.

4. EXPERIMENTAL RESULTS

Several experiments have been carried out to investigate the performance of the DCT pyramid video codec with the proposed optimum bit allocation algorithm. First, three-layer DCT pyramid has been implemented for CIF (352x288) resolution video input. The DCT pyramid video encoder generates three coded bitstreams, one at the base layer of the quarter-QCIF (88x72) and two at the enhancement layers for the reversed L-shapes. Second, the rate control method (TMN8) described in [5] has been used separately in the encoder to select the quantization parameters.

In our simulations, the first frame is intraframe coded with fixed quantization parameters, which are 5, 10, and 20 at the base and enhancement layers respectively. The frame layer rate control in TMN8 assigns a target number of bits for the current frame to be encoded after the first intracoded frame. Then, the bit allocation technique, as defined by equation (10), is implemented to allocate the target bits for the base and the reversed L-shapes in the enhancement layers and hence, the macroblock layer rate control is used to select the quantization parameters. Finally, the proposed optimum bit allocation technique is compared with a non-optimum bit allocation one. In the non-optimum bit allocation algorithm, we assign a constant target number of bits per frame in the base and the first reversed L-shapes layers, equal to the average number of bits obtained from the optimum bit allocation of these layers. That is T_{0j} and T_{1j} for non-optimum are constant for all frames. Since the encoded bit is different from the target bit rate, then the remaining bit rate budget $T_{2j} = T_j - T_{0j} - T_{1j}$, which is assigned to the second reversed L-shapes layer, can vary from frame to frame. It seems this is the best possible method of non-optimum coding. This is because the assigned number of bits to the base and the first enhancement layers are derived by an optimum algorithm. The only difference from an optimum method of coding of these layers is that they

TABLE I: Comparison between optimum and non-optimum bit allocation techniques for various sequences, frame rates and target bit rates.

				Optimum Bit Allocation		Non Optimum Bit Allocation	
Video sequence	Total Frames	Target R Kbps	Frame Rate	Obtained Rate	PSNR (dB)	Obtained Rate	PSNR (dB)
“Claire”	97	300	30	300.83	40.22	302.52	39.52
“Claire”	97	600	30	600.66	41.87	603.4	41.56
“Claire”	97	100	10	100.32	38.28	101.02	37.76
“Claire”	97	64	10	64.19	36.17	64.2	36.09
“Salesman”	75	300	30	279.93	29.67	296.76	27.97
“Salesman”	75	600	30	504.92	32.80	572.53	32.89
“Salesman”	75	100	10	95.89	29.69	100.22	25.76

are fixed. Hence any improvement that our optimum bit allocation algorithm may give over this non-optimum one is expected to be even larger for other sub-optimum methods.

Table I describes the video sequences, the total number of frames, target bit rate, and frame rate and also shows the bit rates and the average luminance PSNR obtained by the optimum and non- optimum bit allocation techniques. Fig. 2 shows the PSNR/frame of the CIF size “Salesman” sequence encoded at 30 frame/s and at the bit rate of 300 Kb/s. The improvement in picture quality due to optimum bit allocation is very evident.

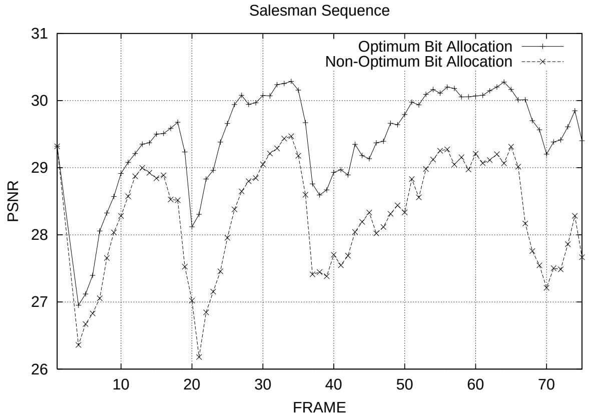

Fig. 2. PSNR/frame for the sequence “Salesman” coded at rate of 300 Kb/s.

5. CONCLUSION

We have proposed an optimum bit allocation technique for the DCT pyramid to allocate the target bits among the layers. Our technique is based on an analysis of the rate and distortion within the Lagrangian optimization method that minimizes the overall coding for a given bit budget. This analysis provides analytical solution of the bit allocation for the DCT pyramidal video coding. We implemented an optimum bit allocation technique of the DCT pyramid and compared its performance to the non-optimum case. The experimental results indicate that the optimum bit allocation can improve the picture quality significantly.

6. REFERENCES

[1] K. H. Tan and M. Ghanbari, “Layered image coding using the DCT pyramid,” *IEEE Trans. On Image Processing*, Vol. 4, No. 4, April 1995.

[2] Y. Shoham and A. Gersho, “Efficient bit allocation for an arrbitary set of quantizers,” *IEEE* Trans*. Acoustic, Speech, Signal Processing*, Vol. 36, pp. 1445-1453, 1988.

[3] K.Ramchandran, A. Ortega, and M. Vetterli, “Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders,” *IEEE Trans. On Image Processing*, Vol. 3, No. 5, Sept. 1994.

[4] B. Tao, W. Dickinson, and H. A. Peterson, “Adaptive model-driven bit allocation for MPEG video coding,” *IEEE Trans. On Circuits and Systems for Video Technology*, Vol. 10, No. 1, Feb. 2000.

[5] J. Corbera and S. Lei, “Rate control in DCT video coding for low-delay communications,” *IEEE Trans. On Circuits and Systems for Video Technology*, Vol. 9, No. 1, Oct. 1999.

[6] J. Corbera and S. Lei, “A frame layer bit allocation for H.263+,” *IEEE Trans. On Circuits and Systems for Video Technology*, Vol. 10, No. 7, Oct. 2000.

[7] R. Atta and M. Ghanbari, “An efficient layered video codec based on DCT pyramid,” *IEEE International Conference on ICASP2002,* Orlando, May 2002.