Proceedings of ICSP' 96

A MULTIRESOLUTION TWO-LAYER VIDEO CODEC FOR NETWORKING APPLICATIONS

K.T.Ng, S.C.Chan and T.S.Ng

Department of Electrical and Electronic Engineering
The University of Hong Kong, Pokfulam Road, Hong Kong
E-mail: scchan@hkueee.hku.hk

ABSTRACT

In this paper, we propose an improved multiresolution two-layer video codec that can operate at a base layer of 64 kbps and a variable bit rate second layer (VBR). The basic idea is to use multiresolution technique to reduce the original CIF format video to QCIF format. The low-resolution video is then encoded with an improved H.261 codec recently proposed in [4]. To reduce the bit rate of the second layer, the residuals are encoded with discrete cosine transform (DCT) using perceptually weighted quantization step sizes. Simulation results show that the Miss America sequence can be transmitted at 64 kbps in CIF format with reasonably good picture quality. For good picture quality, the second layer would typically require 190 kbps.

1. INTRODUCTION

The CCITT has recommended that future transmission of Broadband Integrated Services Digital Networks (B-ISDN) will be asynchronous, based on Asynchronous Transfer Mode (ATM). The performance of ATM networks in handling multimedia services very much depends on the way in which each service is coded and packetized. In particular video services due to their high bit rate requirement can influence the performance of ATM networks to a large extent.

Two-layer coding of video signals has proved to be a powerful technique in alleviating the problem of cell loss in ATM networks [1]. Since ATM networks support two levels of priority, in a congestion network, lower priority cells are discarded prior to the higher priority ones. In general, the base layer codes a low quality picture and is assigned a higher priority, while the residual coding distortions are coded by a second layer and transmitted with lower priority. In [2], the base layer is obtained by coding the input video using the H.261 video coder and the second layer is obtained by requantizing the residual quantization distortion. The step sizes of the requantizers determine the final picture quality. The base layer can be coded in either constant bit rate (CBR) or variable bit rate (VBR) and the second layer is coded in VBR. It was concluded that the CBR-VBR coder where the packetizer obtains its input from the input to the elastic buffer is the best candidate for two-layer coding technique.

In many applications, it is desirable to keep the bit rate for the base layer to a low value so that large number of channels can simultaneously be supported. For the H.261 codec, the operating bit rate is from 64 kbps to 384 kbps. At the lowest transmission bit rate (64 kbps), the picture quality will deteriorate rapidly especially for CIF pictures at a typical frame rate of 12.5 frames per second (fps). Another problem with the basic two-layer codec in [2] is that the bit rate required to transmit the second layer is very high. The main reason is that the step size of the requantizer in the second layer is 8, which is rather small.

In this paper, we propose an improved multiresolution two-layer codec that can operate at a base layer of 64 kbps and a VBR second layer with much lower bit rate. The basic idea is to use multiresolution technique [3] to reduce the original CIF format video to QCIF format. The low-resolution video is then encoded with an improved H.261 codec recently proposed in [4]. The latter is a constant bit rate (CBR) codec using a new buffer control algorithm which provides precise control of the buffer and leads to much better image quality. There are two ways to generate the second layer. One is to employ subband coding and transmit the quantized high frequency coefficients in the second layer. The second one is to encode the residuals as in pyramid coding. The advantage of the pyramid coding is that it is only needed to perform the low-low band analysis and synthesis filtering which leads to much less implementation complexity. The cost is the slight increase in the bit rate of the second layer. To reduce the bit rate of the second layer, the residual is encoded with discrete cosine transform (DCT) using perceptually weighted quantization step sizes.

In section 2, we shall briefly summarize the concept of multiresolution coding. The improved H.261 coder will

0-7803-2912-0

be discussed in section 3. Finally the architecture and results of the proposed two-layer coder will be given in Section 4 and 5.

2. PYRAMID CODING

In the multiresolution coding scheme [3], the original image is successively filtered and decimated by the filterbanks (horizontally and vertically by a factor of 2) to obtain a set of lower resolution images. In pyramid coding, the lower resolution image is then encoded and interpolated to the original domain followed by encoding of the residuals. The process repeats until the original resolution is reached. In subband coding, the subband coefficients are quantized. The encoded image is obtained by passing the quantized subband coefficients through the synthesis filter banks. Pyramid coding differs from subband coding in that the number of data samples after pyramid decomposition is slightly larger than the number of original samples. However, it has the advantages of lower implementation complexity and might be able to control the spatial domain error after reconstruction. There are a number of perfect reconstruction systems that can be used. The 7/9 biorthogonal wavelet [5] is chosen here due to its better perceptual quality and reasonable implementation complexity.

3. BUFFER CONTROL ALGORITHM

In the H.261 standard, the video multiplex is arranged in a hierarchical structure with four layers [6]. From top to bottom the layers are Picture, Group of Block (GOB), Macroblock (MB), Block. By using MQUANT in the structure of macroblock layer, each macroblock can use its own quantization scale factor. The output bit rate is controlled by varying the quantization scale factor, e.g., in RM8, it is updated at every 11th MB.

Consider encoding the residual image after motion compensation that consists of G GOB and N macroblocks at a target bit rate of R using the H.261 coder. We want to find a set of quantization scale factor $\{Q_n : n = 1,\ldots,N\}$ where $(Q_n \in \{1,2,\ldots 31\})$ for the N macroblocks that minimize the overall distortion

$$D = \sum_{n=1}^{N} D_n(Q_n) \qquad (1)$$

subject to the bit rate constraint:

$$B = H^{Pict} + G * H^{GOB} + \sum_{n=1}^{N} [H_n^{MB} + C_n(Q_n)] \le R \qquad (2)$$

Here, $D_k(Q_k)$ is the reconstructed distortion of the k-th MB if it is quantized with quantization scale factor Q_k, $C_k(Q_k)$ is the no. of bits generated in coding the DCT coefficients of the k-th MB with quantization scale factor Q_k, and H^{Pict}, H^{GOB} and H_k^{MB} are the numbers of bits generated for the picture header, the GOB headers and the k-th MB header, respectively. Information such as macroblock address, coded block pattern, quantizer step size, and motion vector is included in H_k^{MB}.

The minimization problem of (1) is very similar to the bit allocation problem. Unfortunately, the Macroblock Address and Motion Vector Data in H_n^{MB} can either be coded directly or coded differentially depending on whether the previous blocks are skipped or not. This in turn depends on the quantizer step sizes of the previous blocks. Hence, it is very difficult to obtain an optimal solution. The problem can be solved by using the mean value to estimate those parts of H_n^{MB} that will be affected by the nature of previous macroblocks. We can then calculate $\widetilde{H}_n^{MB}(Q_n)$ that only depend on Q_n. Eqn. (2) is then modified to :-

$$\widetilde{B} = \sum_{n=1}^{N} [\widetilde{H}_n^{MB}(Q_n) + C_n(Q_n)] \le \widetilde{R} \qquad (3)$$

where $\widetilde{B}$ is the number of bits used in coding the macroblocks and $\widetilde{R}$ is the target bit rate for coding the macroblocks excluding bits that used by H^{Pict} and H^{GOB} which are constant for all pictures.

Since the quantization of each macroblock is independent of each other, they can be viewed as allocating a given number of bits to N independent quantizers to minimize a total distortion measure. Define the efficiency λ_k as follows :-

$$\lambda_k = \max_q \frac{-\Delta D|_{Q_k \to q}}{\Delta \widetilde{B}|_{Q_k \to q}} \qquad (4)$$

with

$$\Delta D|_{Q_k \to q} = D_k(q) - D_k(Q_k) \qquad (5)$$

$$\Delta \widetilde{B}|_{Q_k \to q} = [\widetilde{H}_k^{MB}(q) + C_k(q)] - [\widetilde{H}_k^{MB}(Q_k) + C_k(Q_k)] \qquad (6)$$

where $\Delta D|_{Q_k \to q}$ and $\Delta \widetilde{B}|_{Q_k \to q}$ are respectively the change in distortion and the change in overall bit rate used for all macroblocks when the quantization scale factor of k-th macroblock Q_k is replaced by q.

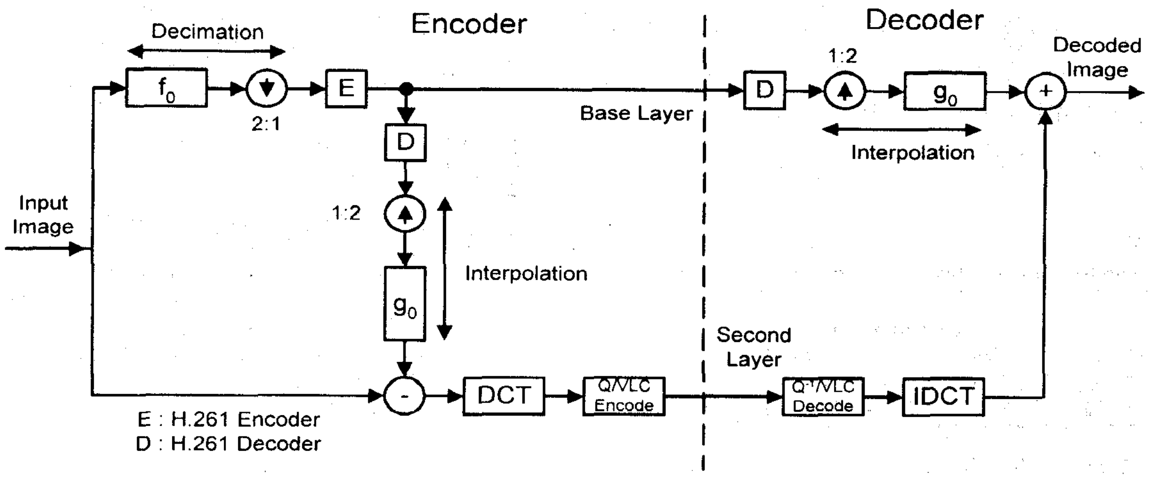

Fig. 1. Proposed Two-Layer Video Codec

The algorithm is depicted as follows :-

1. Calculate function $D_n(q)$ and $[\widetilde{H}_n^{MB}(q)+C_n(q)]$ for n = 1,....N and q = 1,....31.
2. Initialize all quantization scale factors to maximum value of Q_n.
3. Calculate λ_n and put them into a list in descending order of their values.
4. Update the quantization scale factor Q_p that the ratio of decrease in distortion to increase in bit rate is maximized (i.e., at the beginning of the list).
5. Calculate the new value of λ_p and insert it into the list.
6. Repeat Steps 4 and 5 while $\widetilde{B} \leq \widetilde{R}$.

It can be observed that our algorithm has precise control of the bit rate for each image frame. Therefore, both the buffer requirement and the delay can be kept to minimal by using a constant bit rate for each picture.

In traditional buffer control algorithms, the quantizer step size for the current block depends only on the current buffer status and sometimes on the activities of the block. Therefore, the quantization error of adjacent blocks can vary significantly especially at the boundary of moving objects. This is undesirable as it will generate significant blocking artifacts in the encoded images. The salient features of our scheme are that i) the quantization scale factors are determined from the information of the whole picture; ii) it has precise control of the buffer; and iii) it tries to allocate the given number of bits as efficient as possible in a rate-distortion sense.

4. CODEC DESCRIPTION

The proposed two-layer video codec is shown in Fig. 1. The input video will first undergo the subband decomposition using 7/9 biorthogonal wavelet in [5]. The low-low band is encoded using the improved H.261 coder [4]. The residual image is then partitioned into square blocks of size (8×8). Each (8×8) block will undergo the DCT and quantization as in H.261. The main difference here is that the quantizer step sizes for each transform coefficient (q_{ij}) are different and is generated from the psychovisual threshold matrix T as follows :-

$$\mathbf{Q}_M = \mathbf{Q}\cdot\mathbf{T}$$

or equivalently

$$q_{ij} = \mathrm{Q}\cdot t_{ij} \qquad i,j=0,\ldots.,7$$
$$\text{with } [\mathbf{Q}]_{ij}=q_{ij} \text{ and } [\mathbf{T}]_{ij}=t_{ij}$$

Here t_{ij} are the psychovisual thresholds found from JPEG standard [7]. Q is fixed for each image. By varying Q, the bit rate for the second layer can be controlled. The results in Section 5 shown that the use of psychovisually weighted quantization step sizes can greatly reduce the bits required to transmit the second layer while preserving the perceptual quality of the encoded image.

The second layer retains the syntax of the video multiplex coder of H.261 [6], except removing Type Information, Motion Vector Data and Quantizer Information from the headers that are not needed in the second layer.

5. EXPERIMENTAL RESULTS

Computer simulations of two-layer codec on CIF image sequence Miss America was used to evaluate the proposed paradigm. The base layer (after subband decomposition)

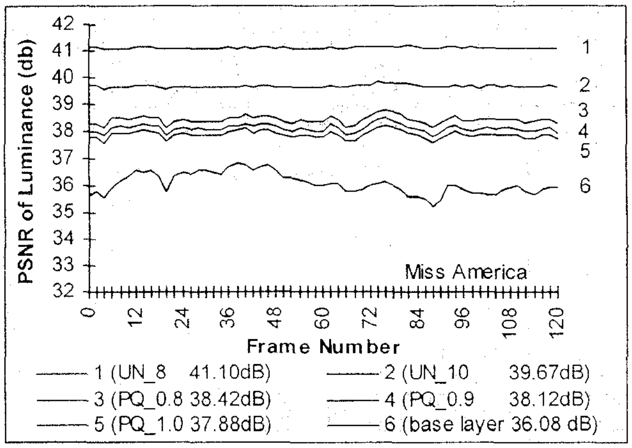

Fig. 2(a). PSNR comparison of various methods.

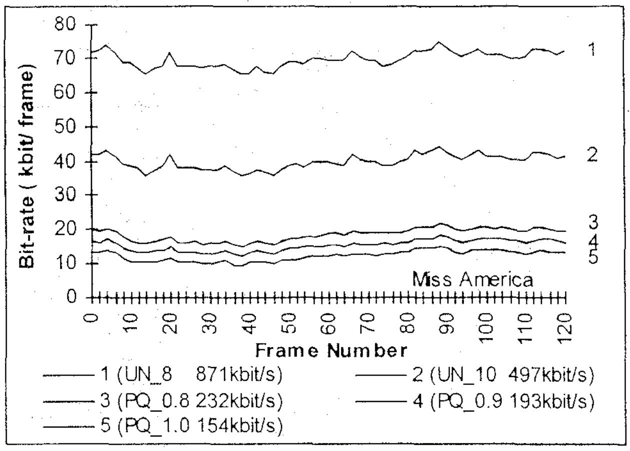

Fig. 2(b). Bit rate comparison of various methods.

was encoded by H.261 codec operating at 64 kbps (12.5 frame per second). The improved H.261 base layer codec achieves a higher PSNR value (40.70 dB) than conventional coder at the same bit rate [4]. If only the base layer is used to reconstruct the original image (i.e., using only Low-Low band in the synthesis filtering), a PSNR of 36.08 dB can be achieved. The perceptual quality of the picture is rather good. This is very difficult to achieve if the base layer is generated by directly coding the CIF image at 64 kbps. Fig. 2 shows the effects of using different quantizers for encoding the second layer :

i) uniform quantizers with step sizes equal to 8 and 10 (UN_8, UN_10).
ii) psychovisually weighted quantizers with Q equals to 0.8, 0.9 and 1.0 (PQ_0.8, PQ_0.9 and PQ_1.0).
iii) Decoding base-layer only.

It can be observed that our proposed methods require much fewer bits for the second layer than using uniform quantizers (UN_8, UN_10). On the other hand, the PSNR is about 2 dB & 4 dB lower than UN_10 and UN_8, respectively. When looking at the decoded images, it is found that PQ_0.8 has almost identical visual quality to UN_10. The differences among Q = 0.8, 0.9 and 1 are noticeable but are very close to each other. Therefore, we concluded that a value of Q = 0.8 to 1 is sufficient for most applications.

6. CONCLUSION

In this paper, we have proposed an improved multiresolution two-layer video codec that can operate at a base layer of 64 kbps and a variable bit rate second layer. The basic idea is to use multiresolution technique to reduce the original CIF format video to QCIF format. The low-resolution video is then encoded with an improved H.261 codec recently proposed in [4]. To reduce the bit rate of the second layer, the residuals are encoded with discrete cosine transform (DCT) using perceptually weighted quantization step sizes. For the Miss America sequence, simulation results show that the proposed perceptually weighted quantization can achieve similar visual quality to traditional method using uniform step size with a much lower bit rate. A value of Q = 0.8 to 1 is also found to be sufficient for most applications.

ACKNOWLEDGMENT

This work is support by the Hong Kong Research Grants Council and the CRCG of the University of Hong Kong.

REFERENCES

[1] M. Ghanbari, "Two-layer coding of video signals for VBR networks," IEEE J. Select. Areas Commun., vol. SAC-7, no. 5, pp. 771-781, June 1989.

[2] M. Ghanbari and J. Azari, "Effect of bit rate variation of the base layer on the performance of two-layer video codecs," IEEE Trans. Circuit and Systems for Video Technology, Vol. 4, No. 1, pp. 8-17, Feb. 1994.

[3] K. M. Uz et al, "Interpolative multiresolution coding of advanced television with compatible subchannels," IEEE Trans. Circuit and Systems for Video Technology, Vol. 1, No. 1, pp. 86-99, March 1991.

[4] K. T. Ng, S. C. Chan and T. S. Ng, "Buffer control algorithm for low bit-rate video compression," to appear in Proc. IEEE ICIP'96, Lausanne, Switzerland.

[5] M. Antonini, et al, "Image coding using wavelet transform," IEEE Trans. Image Processing, vol. 1. no. 2, April 1992.

[6] CCITT Recommendation H.261, "Video Coding for Audiovisual Services at px64 kbits/s," Genera, Aug. 1990.

[7] ISO DIS 10918-1, " Digital Compression and Coding of Continuous-Tone Still Images".