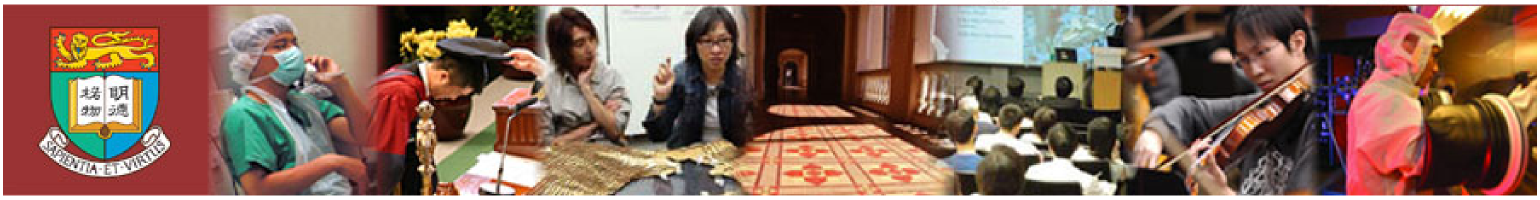

International Conference on
Information, Communications and Signal Processing
ICICS '97
Singapore, 9-12 September 1997

An Adaptive Multiresolution Modification Of The H.263 Video Coding Algorithm

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, an adaptive multiresolution approach for video coding is presented. The algorithm uses the information content to determine the resolution of the video to be encoded. An important advantage of the algorithm is that the codec can maintain a very stable frame rate with reasonable image quality during scene change and provide better quality video when the motion is less rapid. Simulation results show that the modified H.263 coder, using the proposed algorithm, can maintaining better image quality and a more steady frame rate than the TMN 5 algorithm at low bit-rate.

1 Introduction

Recent advances in digital technology enable multimedia objects like video and speech to be manipulated and transmitted over high speed network. Users can access multimedia information database, communicating with each other using video phones, and working together at different locations using various video services.

One problem with low bit-rate video coding is the limited number of bits available to encode the video. In case of scene change or video with rapid motion, a large amount of bits will be required to encode the picture. Though the encoder buffer can help to smooth out this fluctuation, the quality of the image can become very poor and significant blocking artifact will result. The effectiveness of the buffer also depends on the buffer control algorithms employed. In extreme case, it might not even able to encode the quantized coefficients (buffer overflow) after encoding the motion information and other overheads. A usual approach is to skip the current frame and let the buffer accumulate sufficient number of bits to encode the next image frame. As a result, the video might look jerky and the frame rate is not steady.

In this paper, instead of reducing the temporal resolution by frame skipping, we propose an adaptive multiresolution approach to deal with this problem. The basic idea is to switch the encoder to operate at lower resolution when the amount of bits available is severely limited. Reducing the spatial resolution is an effective approach to reduce the information to be encoded and the overheads associated with various video coding standards. Therefore, it is possible to make the best use of the available bits to capture the motion and produce reasonably image quality even at low bit-rate. In [1], we have found that it is possible to encode the CIF format Miss America sequence at 128 kbps with a frame rate of 12.5 f/s and 25 f/s using the H.261 coder together with subband decomposition. In this coder, the video is first decimated using the two dimensional separable 7/9 biorthogonal wavelet transform. The QCIF format video is then encoded using the H.261 algorithm with our proposed bit allocation buffer control algorithm [2]. At the decoder, the video is decoded and interpolated to the original resolution. The video quality is reasonably good because most of the available number of bits are used to represent the image data. A problem with this approach is that the signal to noise ratio is limited by the decimation process when the motion is less rapid. Here, we shall propose an adaptive strategy to switch the coder to operate at different resolutions to deal with this problem. The resulting codec is able to maintain very stable frame rate with reasonable quality during scene change and provide better quality video when the motion is less rapid.

The layout of the paper is as follows: In section 2, we shall briefly describe the structure of the modified H.263 algorithm. Section 3 is devoted to the proposed adaptive multiresolution codec and the adaptation strategies are discussed in Section 4. Computer simulation results and comparison with the TMN 5 codec is given in Section 5.

2 The Modified H.263 Algorithm

The basic configuration of the H.263 video coding algorithm [3] is similar to the H.261 Recommendation. There are five standardised picture formats: sub-QCIF, QCIF, CIF, 4CIF and 16CIF. Unlike H.261, half pixel precision is used for motion compensation. In addition to the basic video source coding algorithm, four negotiable coding options are included for improved performance: Unrestricted Motion Vectors, Syntax-based Arithmetic Coding, Advanced Prediction, and P-B Frames. These options can be used together or separately.

Fig. 1 shows the generalized form of the H.263 source coder. The main elements are prediction, block transformation and quantization. The video multiplex is arranged in a hierarchical structure with four layers. From top to bottom the layers are: Picture, Group of Block (GOB), Macroblock (MB), and Block Layers. The input frame is partitioned into macroblocks consisting of one luminance block of (16×16) pixels and two chrominance blocks of (8×8) pixels. The prediction is inter-picture and may be augmented by motion compensation (MC) (optional in the encoder) and a spatial filter.

A number of optimization has also been done in H.263 bit packing to make it more efficient as compared with H.261. For example, the End-Of-Block (EOB) is eliminated by specifying, in the VLC table, whether the transform coefficient is the last coefficient in the given block. Also, in H.263, the quantizer step size between consecutive macroblocks are constrained to reduce the number of bits needed to specify the quantizer scale. In fact, adjacent quantizer levels can only differ by ± 2 (DQUANT). This is quite different from H.261 where the quantizer scale for each macroblock can take on any of the 31 different values.

In a previous work [2], the authors had proposed a new buffer control algorithm for motion-compensated hybrid DPCM/DCT coding. The algorithm is based on the use of bit allocation algorithm to determine the quantization scale factors in such coder to meet a given target bit rate. The salient features of the scheme are that i) the quantization scale factors are determined using 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.

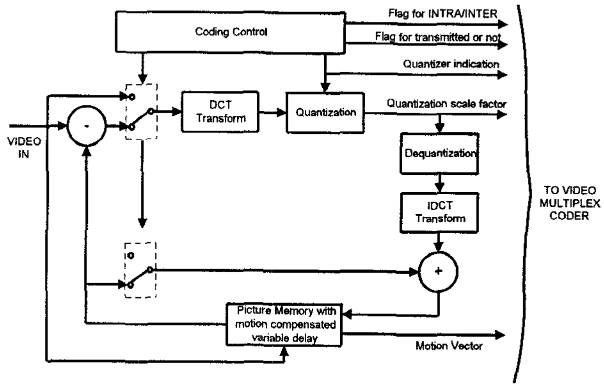

Figure 1: H.263 Source Coder.

The buffer control problem for H.263 is considerably more complicated than the H.261. This is due to the constraint in the quantizer, the highly coupled advanced prediction, and the use of PB-frames in the coder. We modified the H.263 algorithm so that the quantizer constraints are removed [4]. Also, we shall consider the default H.263 coder where the options are turned off. After performing the bit allocation, we allow the quantization scale factors to vary from 1 to 31 as in the H.261 algorithm. To save the number of bits to represent these scale factors, we modified the VLC codes in the macroblock type and coded block pattern to distinguish whether the difference in the quantizer scale factor is within the limit of ± 2. If it is true, then we send the appropriate code for the macroblock type and coded block pattern together with the differential value to the receiver. Otherwise, we send the corresponding code together with the five bits quantization scale factor to specify one of the 31 quantizer that is going to be used in that block.

3 Codec Description

The proposed adaptive multiresolution video encoder with two levels is shown in Fig. 2. The encoder employed is the modified H.263 codec that we have discussed in Section 2. It uses the bit allocation algorithm [2] to determine the quantization scale factor of each macroblock. Each frame is allocated the same number of bits. The input video at CIF format will pass through the H.263 encoder. The multiresolution decision logic will determine which source format or resolution the codec should operate.

When rapid motion is experienced in the video sequence, the decision logic will use the lower resolution mode for encoding. The previous reconstructed picture in the frame buffer and the current input picture will undergo the subband decomposition using the 7/9 biorthogonal wavelet transform in [5]. The low-low band with lower resolution will be passed to the modified H.263 encoder. A QCIF format bitstream will be generated. After the scene change or the motion have passed away, the encoder can operate at a higher resolution. The previous reconstructed QCIF format picture will pass through the synthesis filter for interpolation. The encoder will switch back to the higher resolution mode at CIF format. The "Source Format" field in each picture header will indicate the resolution the encoder is currently operating. A simple measure for mode switching is discussed in the next Section.

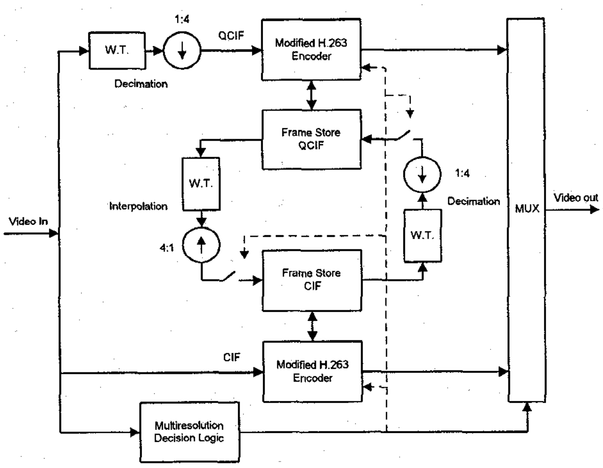

Fig. 2. Proposed two-level adaptive multiresolution video codec.

4 Adaptation or Mode Switching

When the number of bits used to encode the headers of the picture (B_H) (including motion vectors) exceed a factor α (say ½ or ¾) of the bit budget of that picture (B), it is very likely that the encoder will not have sufficient number of bits to encode the quantized coefficients at that resolution. This can serve as an effective measure for mode switching.

In the proposed approach, we shall perform the motion estimation in the current resolution j (where j = 1,...,L and 1 is the original resolution) to decide whether it is necessary to operate at a lower resolution mode. If this is the case ($B_H^{(j)} > \alpha^{(j)} \cdot B$), the input video and the previous reconstructed frame will be decimated for encoding. $\alpha^{(j)}$ is a constant which only depends on the resolution. The process is repeated until the image frame can be encoded efficiently with the allocated number of bits at resolution l ($l > j$).

If sufficient number of bits are available (i.e. $B_H^{(j)} < \alpha^{(j)} \cdot B$), then the encoder will try the next higher resolution mode. The reconstructed frame will be interpolated to the next higher resolution ($j-1$) for motion estimation. The process repeats until the best resolution is determined. In principle, the best performance can be obtained by selecting the resolution in which the PSNR is maximized. However, this can be time consuming due to the multiple encoding needed at each resolution tested. An ad hoc approach is to stop the searching when $B_H^{(k)}$ is just less than $\alpha^{(k)} \cdot B$. The complexity of the algorithm can be quite large without any fast algorithms. It is because we have to perform motion estimation at all the resolution tested. Fortunately, the resolution of the video for low-bit rate applications is usually limited to CIF or QCIF format with a frame rate less than or equal to 12.5 f/s. Therefore, the increased complexity is not excessive. Also, we can employ the hierarchical motion estimation algorithm like the one in [6] to perform the motion estimation. In the simplest two level codec, the modification to the H.263 is moderate. To inform the decoder about the image format, the encoder will send the format of the current image (QCIF or CIF) in the frame's header of each picture.

5 Experimental Results

Computer simulations were performed on several H.263 [3] based algorithms to evaluate the proposed adaptive multiresolution codec with two levels. The test video is the Miss America sequence at CIF format and 12.5 frames per second (f/s). The following encoding schemes are tested and compared:

i) TMN 5 Model (H.263 with buffer control) [7].

ii) Modified H.263 codec with $\alpha = 3/4$ (**Algorithm A**).

iii) Modified H.263 codec with $\alpha = 1/2$ (**Algorithm B**).

Subjective evaluation is performed using a Viewstore 6000 real time playback system with a 21 inch EIZO color monitor. The first 120 frames of the sequence was encoded by the various algorithms at 37.5 kbps.

Since the first image frame has to be encoded in intra mode, it will require a lot of bits. The TMN5 model will skip the first few frames until sufficient number of bits becomes available. The adaptive multiresolution codec will also be operated in intra mode. However, it will treat it as a scene change and operate at the lower resolution mode. Since we only have two levels here, if the number of bits is greater than that assigned to each image frame, it will also skip the current image frames. However, our codec will start encoding much earlier than the TMN5.

Fig. 3 shows the PSNR comparison of the encoding schemes tested. It can be seen that the proposed codec can maintain the frame rate at 12.5 f/s (60 frames) where the TMN 5 can only obtain an averaged frame rate of 9.28 f/s (37 frames). Algorithm A also achieves a higher averaged PSNR value than the TMN 5 at the same bit rate with more frames being encoded.

When comparing Algorithm B with TMN 5, it is found that the former is a little bit blurred around image edges due to the more frequent decimation and interpolation processes. But the blocking artifacts are greatly reduced. Both Algorithm A and Algorithm B can maintain a very steady frame rate justifying the proposed algorithm. The best tradeoff is achieved by algorithm A.

The effectiveness of the proposed algorithm in handling scene change is readily observed in the first reconstructed intra frame of the TMN 5 algorithm (37kbits) and Algorithm A (21kbits). The visual quality of both reconstructed images are comparable. However, about half of the bits can be saved by using the proposed algorithm as compared with the TMN5 algorithm.

6 Conclusion

We have presented an adaptive multiresolution approach for video coding. This algorithm adaptively adjusts the resolution of the video to be encoded and make better use of the bits available. An important advantage of the algorithm is that the codec can maintain a very stable frame rate with reasonable image quality during scene change and provide better quality video when the motion is less rapid. Simulation results showed that the modified H.263 coder, using the proposed algorithm, can maintain better image quality and a more steady frame rate than the TMN 5 algorithm at low bit-rate.

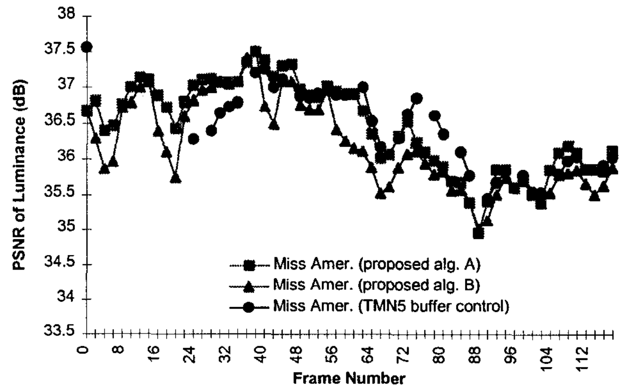

Fig. 3. PSNR comparison of different encoding schemes. (unlinked points mean that frames are being skipped)

Mean PSNR	TMN 5	Proposed Alg. A	Proposed Alg. B
Y Comp.	36.38	36.46	36.20
U Comp.	36.21	36.39	36.43
V Comp.	37.10	37.16	37.21

Table 1. Mean PSNR of Miss America sequence.

References

[1] K. T. Ng, S. C. Chan and T. S. Ng, "A multi-resolution Two-layer video codec for networking applications," in *Proc. 1996 Third Int. Conf. on Signal Processing,* Beijing, Oct. pp. 1071-1074.

[2] K. T. Ng, S. C. Chan and T. S. Ng, "Buffer control algorithm for low bit-rate video compression," *Proc. IEEE Int. Conf. on Image Proc., ICIP'96,* Lausanne, 1996, Vol I, pp. 685-688.

[3] Draft ITU-T Recommendation H.263, "Video coding for low bitrate communication", July,1995.

[4] K. T. Ng, S. C. Chan and T. S. Ng, "A modified H.263 algorithm using bit allocation buffer control algorithm," *Proc. IEEE ISCAS '97,* Hong Kong, Vol II, pp.1389-1392.

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

[6] K. W. Cheng and S. C. Chan, "Fast block matching algorithms for motion estimation" in *Proc. IEEE ICASSP '96,* Atlanta, Georgia, USA.

[7] ITU Telecommunication Standardization Sector LBC - 95. Study Group 15, Working Party 15/1. Expert's Group on Very Low Bitrate Visual Telephony. Video Codec Test Model, TMN5.