A LOW BIT RATE HIERARCHICAL VIDEO CODEC

Kui Zhang, Mirosław Bober and Josef Kittler
Department of Electronic and Electrical Engineering,
University of Surrey, Guildford GU2 5XH, United Kingdom
Tel: +44-1483-300800 Ext. 3026; fax: +44-1483-34139
e-mail:K.Zhang@surrey.ac.uk

ABSTRACT

The performance of a very low bit rate video codec largely depends on the efficient use of motion compensated prediction technique and on a good coding control strategy. In our previous approach [6], we proposed a multiple layer video codec using affine motion compensation. In this paper, we further extend our affine compensated multi-layer codec by incorporating a new block level and designing a coding control strategy. A measure of coherent motion is used in the decision process which makes the codec perform efficiently at very low bit rate and for small size image sequences (QCIF and sub-QCIF format). The experimental results conduced on 15 MPEG test sequences in QCIF format show improvement in PSNR of 0.2 dB and reduction in bit rate of 0.9 kbits/second.

1 Introduction

In very low bit rate video coding, a successful application of motion prediction can significantly reduce the bit rate while maintaining high quality of the reconstructed image. The translational motion model, which is used in the majority of existing approaches, works well in fixed-block-size codec such as MPEG, H.261 and H.263. However, when larger regions are used in the coding process it is no longer capable of coping with complex motions occurring in real world sequences and results in relatively large prediction errors. This in turn increases the bit overhead of the spatial compensation stage in which the motion prediction error is encoded. Thus any attempt to reduce the overall bit rate must be primarily directed to improving motion compensated prediction.

The affine motion model which has been used by some researchers [4, 5] can accurately describe most of the 3D rigid motions. However, the problem of extracting motion parameters reliably and robustly is very complex. In our earlier approaches [8, 7], a robust Hough Transform based motion estimation technique [1] was used to estimate the translational motion parameters accurately and robustly. Recently, we have extended our earlier approaches by employing an affine motion compensated predictor in conjunction with a multi-layered processing structure as a basis of a very low bit rate video codec [6]. In this paper, we further extend the multi-layered structure to include yet another block level. This improves the performance of motion prediction for QCIF image sequences at very low bit rates. We also introduce a new coding control strategy which is specifically designed for the multiple layer codec, so that the coding efficiency is further improved.

2 Multiple Layer Video Codec

Video sequences can be classified into two categories depending on the movement of the background. They are: sequences with stationary background and sequences with moving background. The former can be thought to be a special case of the latter which has motion parameters corresponding to zero motion. Therefore, in general, we can treat all sequences as sequences with background motion. Because the region involved in the background motion compensation is large, a more complex motion model is required. As the region becomes smaller, the motion model can also be simplified so that less bits are needed to encode the motion parameters. In the following we describe a multiple layer video codec and corresponding coding control strategy to achieve high compression ratio and good image quality at bit rate below 24 kbits/sec.

2.1 Codec Structure

In [6], a three layer structure is used. It employs affine motion compensation for larger regions to reduce motion compensation (MC) errors and simple translational motion compensation for smaller regions to keep the bit rate low. However, the intermediate size microblock level (16×16) cannot always provide satisfactory motion compensation due to reduced image size and possible multiple motion within a block. To address this problem, an extra layer (called a basic block level) is introduced. The following layers are used:

1. Frame level: 6-parameter affine motion estimation and compensation is applied. This gives accurate and bit efficient compensation for global motion in scenes where camera is not stationary.

2. Group of microblocks level: the group of microblocks consists of 2×2 microblocks. A simplified 4 parameter affine model is applied for coding at the group of blocks level. We found that for 32×32 regions, this simplified model gives satisfactory results.

3. Microblock level: only translational model is used. Because the region involved is small, the translational model can reduce the bit rate and deliver reasonable performance.

4. Basic block level: when a motion boundary is present inside a microblock, translational motion parameters are estimated for each region (based on the luminance component). The motion parameters for chrominance blocks are estimated from luminance parameters.

First the frame level processing is applied and the image is divided into stationary and non-stationary regions. The motion compensated prediction is only applied to non-stationary regions. For stationary regions, a background memory and lossless updating are applied. Because of the low-pass filtering effect of sub-pixel accuracy motion compensation, the sharpness of the image is lost in the process of repetitive application of motion compensation. To remedy the problem, intra coding has to be applied after certain number of inter codings. In H.261, this is done by applying intra coding to the whole frame at regular intervals. The advantage of such an approach is that it not only restores the sharpness of the image but also recovers any block which has been erroneously motion compensated due to transmission errors or for some other reasons. The disadvantage is that the bit-rate overhead is large which may contribute to the buffer overflow. Consequently a larger buffer is required to accommodate the bursts in the bit stream. Since only regions which have been motion compensated certain number of times need to be intra coded, we use an MC-counter to record the numbers of the motion compensation applied. The counter is increased when motion compensation is applied to a block and reset when the block is intra coded. Whenever the value of the counter is greater than a given threshold, an intra coding is forced for that block. Because usually only a small percentage of blocks is motion compensated repeatedly many times, the overall overhead for extra intra coding is reduced. The only disadvantage is potential lost of robustness to transmission errors and therefore a channel coding may be required to increase error resilience.

2.2 Coding Control Strategy

For fixed block size codec, such as in H.261, the coding control only needs to decide when to use inter (motion prediction) or intra (DCT compression) method to achieve an efficient compression of the sequence with minimum distortion. Control strategies which are base on distortion measure (mean square error (MSE) or mean absolute error (MAE)) and image statistics can achieve good performance for such codecs [2, 3]. However, in a variable block size codec, an extra decision has to be taken to determine at which block level motion compensation should be performed. Once this decision is taken, the control strategy can determine whether the region should be motion compensated or spatially coded.

In our approach, a measure of coherent motion is used in the coding control procedure. Such measure, defined as a proportion of pixels moving coherently within a block (PC_i), is provided by the robust Hough based motion estimation:

$$PC_i = \frac{number\ of\ pixels\ moving\ coherently}{total\ number\ of\ pixels\ in\ the\ block} \qquad (1)$$

When multiple motion is present within a block, the PCi is low. This gives us a simple decision criterion for the selection of a suitable block level for MC. Such decision can be reached without the calculation of distortion measure.

We also use various levels of subpixel accuracy for different image sizes and types. For smaller image sizes (QCIF), a higher subpixel accuracy (1/4 pixel) and smaller motion vector range are used as default values so that a better motion prediction can be achieved while maintaining low bit rate. The maximum motion vector length and the coding overflow of the motion parameters are monitored so that the subpixel accuracy and motion range can adaptively be changed according to the motion presented in the sequence. The adopted distortion measure is based on mean absolute error (MAE) instead of mean square error (MSE). It seems that MAE gives better performance in low bit rate coding control than MSE.

3 Experimental Results

The experiments were carried out on 15 MPEG test sequences in colour QCIF format. Sequences from group A, B, and E were used. Initially, the tests were performed without the bit rate control and then the rate control module was switch on. The average PSNR results are shown in Table 1 and Table 2 respectively for all three components. From Table 1, it can be seen that the introduction of an additional layer improves PSNR for Y component and slightly reduces PSNR for both U and V components. This can be explained bu the fact that the extra layer introduced in proposed method only applies to Y component while the motion parameters in U and V components are the average of 4 motion parameters of Y blocks of motion compensation applied to the fourth level. On the other hand, under tight bit rate control, the proposed algorithm gives better PSNR for all three components due to improved motion compensation and coding control strategy. The average bit rate without and with bit rate control is shown in Table 3. On both occasions, the proposed algorithm gives on average a lower bit rate than that presented previously (with 3 Layers).

Average Results	proposed	3 Layer
PSNR (Y)	34.03	33.83
PSNR (U)	37.14	37.28
PSNR (V)	37.79	37.96

Table 1: **Average PSNR for all 15 sequences without bit rate control.**

Average Results	proposed	3 Layer
PSNR (Y)	33.66	33.61
PSNR (U)	37.39	36.58
PSNR (V)	37.99	36.95

Table 2: **Average PSNR for all 15 sequences with bit rate control (24 kbits/sec.).**

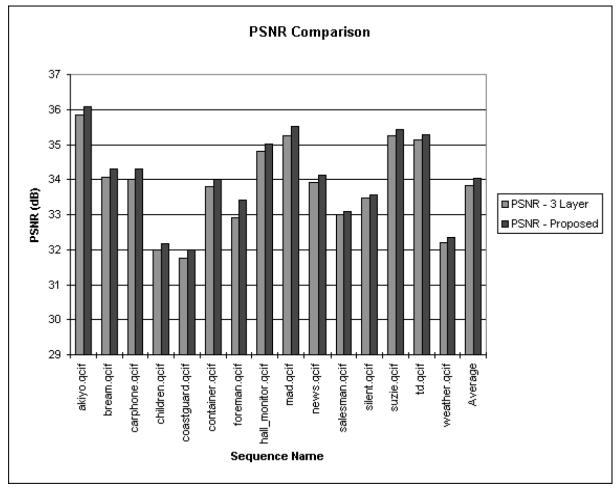
Test Condition	Bit Rate (bit/pixel)	
	proposed	3 Layer
Without Rate Control	39.61	40.51
With Rate Control	24.73	25.49

Table 3: **Average bit rate for all 15 sequences.**

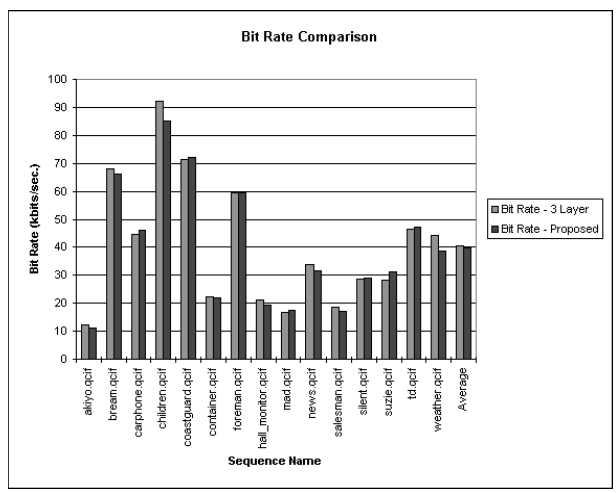
The individual results are shown in Figure 1 and Figure 2. The PSNR shown are for luminance component (Y) only. In both figures, the left bars (lighter) represent the results for 3 Layer structure [6]; the right bars (darker) represent the results for the proposed method.

Figure 1 gives test results under no bit rate control, with the quantization step fixed. From Figure 1 (a) it can be seen that the proposed algorithm gives higher PSNR (Y) for all 15 sequences in the test. In Figure 1 (b), we see that 9 out of 15 sequences has reduced bit rate as compared with our previous approach and 6 sequences have a slightly higher bit rate. The 5 out of 6 sequences where higher bit rate occurred are the ones with large and complex motion such as "carphone". This can be attributed to the extra bit used to encode motion parameters for the fourth level. The signal to noise ratio are improved though. Without bit rate control, the proposed algorithm gives an improvement in bit rate of 0.9 kbits/sec. whereas the average PSNR(Y) increased by 0.2 dB.

The test results with bit rate control are shown in Figure 2. From Figure 2 (a), we see that only 7 out of the 15 sequences maintain higher PSNR (Y). However, the proposed methods still performs better on average than the 3 layer method. In Figure 2 (b), it can be seen that 10 out of the 15 sequences has a lower bit rate for the proposed algorithm. With rate control under target bit rate of 24 kbits/sec. and 10 Hz frame rate, the proposed codec has a slightly higher PSNR (Y) (0.05dB) and a lower bit rate of 0.76 kbits/sec. on average. It can be noted that the bit rate is closer to target 24 kbits/sec for the 4 layer structure, which suggest that such structure is better suited for application where the bit rate has to be strictly maintained.

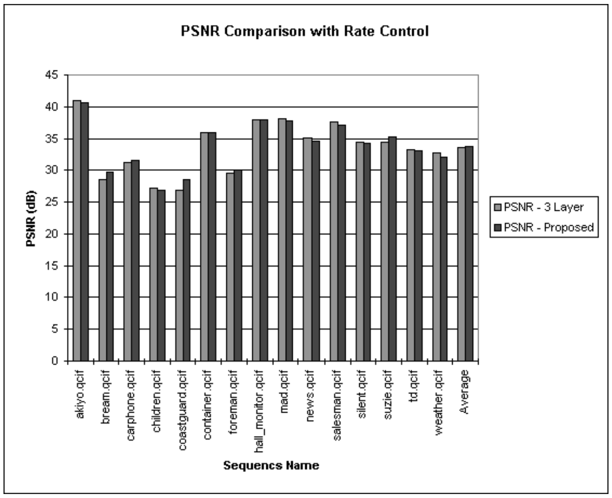


(a)

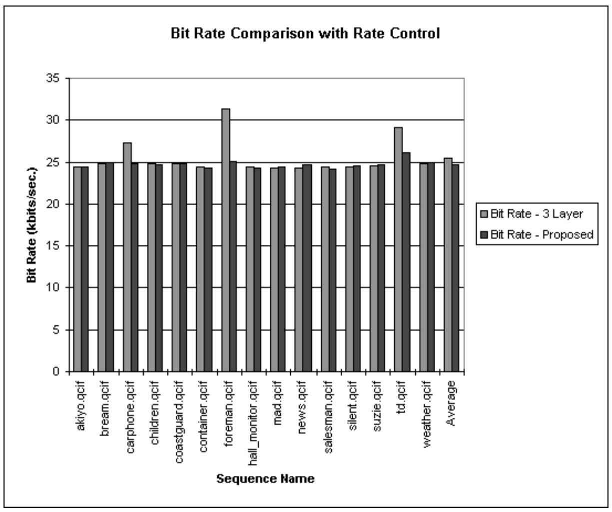


(b)

Figure 1: Average results for 15 tested sequences without bit rate control (a) PSNR (b) Bit rate



(a)



(b)

Figure 2: Average results for tested sequences with bit rate control (a) PSNR (b) Bit rate

4 Conclusion

In this paper, we extend affine compensated multi-layer codec [6] by adding a new coding layer and developing a new coding control strategy. Experiments with large numbers of sequences show improvement in the performance, particularly for small size image format (like QCIF or sub QCIF). We also show that our bit rate control strategy works efficiently even for complex sequences including background motion. This is achieved by affine motion compensation applied at frame level.

The experimental results show that for the head and shoulder sequences with homogeneous background, the proposed algorithm performs comparably with the existing techniques. For sequences with large and complex motion, the proposed method is capable of delivering a higher image quality both in terms of PSNR and subjective tests. For some of the sequences, a slightly higher bit rate may result due to the extra bit overhead when the fourth level motion compensation is used.

Under the bit rate control, the proposed method appears to be more responsive to the rate control mechanism and gives an evenly distributed bit rate for all tested sequences. It also gives on average higher PSNR and lower bit rate. It is clear that for very low bit rate applications and for small size image format, the new components (the fourth layer and new coding control strategy) benefit the performance of the codec.

References

[1] M. Bober and J. Kittler. Robust motion analysis. In *Proceedings, CVPR '94 (IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, June 20-24, 1994*, pages 947–952. IEEE Computer Society Press, 1994.

[2] ITU-T. *Recommendation H.261: Video Codec for Audiovisual Services at p 64 kbits.* International Telecommunication Union, 1994.

[3] ITU-T. *Draft ITU-T Recommendation H.263: Video Coding for Low Bitrate Communication.* International Telecommunication Union, 5 July, 1995.

[4] Hirohisa Jozawa. Segment-based video coding using an affine motion model. In *SPIE Proceedings, Visual Communications and Image Processing'94, Chicago, USA*, volume 2308, pages 1605–1614, 25-29 September, 1994.

[5] H. G. Musmann, P. Pirsch, and H. Grallert. Advances in picture coding. *Proceedings of the IEEE*, 73(4):523–548, 1985.

[6] K. Zhang, M. Bober, and J. Kittler. Video coding using affine motion compensated prediction. In *ICASSP-96,Atlanta, USA*, 7-10 May, 1996.

[7] K. Zhang, M. Bober, and J. Kittler. Variable block size video coding with motion prediction and motion segmentation. In *SPIE Proceedings, Digital Video Compression: Algorithms and Technologie 1995, San Jose, Califorlia*, volume 2419, pages 62–70, February, 1995.

[8] K. Zhang, M. Bober, and J. Kittler. Robust motion estimation and multistage VQ for sequence compression. In *ICIP-94, International Conference on Image Processing, Austin, USA*, volume II, pages 452–456, November, 1994.