MPEG-2 VIDEO CODING WITH IMAGE PARTITIONING

Ekaterina Barzykina, Panos Nasiopoulos and Rabab Ward
The University of British Columbia
Department of Electrical and Computer Engineering

Abstract - In this paper we present an original motion compensation strategy based on frame partitioning. The proposed method uses different temporal resolutions within a frame to improve compression. We present a new bit allocation and rate control algorithm complementing our motion compensation technique. This unique approach to bit allocation ensures the consistency of quality throughout a single frame and a GOP. For the same picture quality, frame partitioning alone yields an additional increase of up to 20 percent or more of the encoding efficiency.

1. INTRODUCTION

The MPEG compression standard has had an enormous impact on the communication and entertainment industries. The newest phase of the standard, MPEG-4, which is expected to reach its final stage in November 1998, combines the base features of MPEG-1 and MPEG-2 with a number of new ideas [1]. One of the unique features of MPEG-4 that is relevant to our work is the concept of a video object. A video object (VO) can be defined in a scene context, where a scene (video information of several consecutive frames) is composed of a number of objects. For example, a person moving across a background would represent a scene with the person classified as VO1 and the background as VO2. Each VO can be of arbitrary shape and is encoded separately. One of the advantages of video objects is that they can be encoded at different temporal/spatial scalability levels.

However, although the concept of video object is without a doubt attractive, the amount of overhead information needed to define objects and reconstruct scenes may become a significant drawback. This overhead can occupy 20 per cent or even more of the total stream bandwidth. Video objects require additional processing and overhead due to the contradiction between the arbitrary shape of video objects and their block based encoding.

In this paper we present a different and much simpler way to exploit the idea of video objects. We propose a new way of achieving different temporal/spatial scalability encoding for different parts of a frame. Our method works within the standardized MPEG-2 syntax and thus does not require any extra overhead information. This technique allows us to attain a significant improvement in the compression performance.

To maximize the advantages of our motion compensation approach, we develop an original bit allocation and rate control algorithm. Our bit allocation strategy results in a greater consistency of quality for reconstructed macroblocks within a frame and for frames within a GOP.

2. MOTION COMPENSATION WITH FRAME PARTITIONING

A. Frame partitioning

The two most challenging goals of video compression are, first, meeting the required bit rate and second, preserving picture quality. The objective is that the quality reaches its highest level for a given bit-rate/buffer size keeps this level consistent throughout the each frame and the entire video stream. Trying to attain better quality, different bit allocation techniques can go only so far in redistributing bits between frames. In order to achieve significant improvement in encoding efficiency and thus quality, an algorithm has to make better use of the redundancies in the input data.

Our proposed method significantly improves the use of temporal redundancies in video encoding. We introduce the idea of frame partitioning as part of a new two level motion compensation algorithm. This technique uses frame partitioning as an extra stage prior to MPEG motion compensation. The objective of frame partitioning is to separate each frame into significant and insignificant regions. The regions are found based on the degree of interdependency between parts of consecutive frames. Being designed to work within MPEG-2 syntax, frame partitioning is block based, and thus does not require any additional overhead.

In frame partitioning, the algorithm looks for the similarities between consecutive frames. After comparing the neighboring images, frame macroblocks are separated into two categories: significant and insignificant macroblocks. A macroblock is considered to be insignificant if it is not "visibly" different from its predicted macroblock. A significant macroblock has acquired new content in comparison with its predicted counterpart. Prediction error for significant macroblocks is calculated using the standard motion compensation procedure. The prediction error of insignificant macroblocks is made equal to zero. Motion information (motion vectors) continues to be transmitted for both types macroblocks.

It is evident that frame partitioning divides each frame into significant and insignificant regions. Drawing an analogy with MPEG-4, these regions are equivalent to two video objects, which are defined based on their temporal resolution. Significant regions will be encoded at a resolution equal to the frame rate, where as the resolution for insignificant regions is half of the frame rate or less.

Frame partitioning results in different temporal resolution for different parts of a single frame. The acquired flexibility can be very beneficial for encoding. The possibility of reducing the temporal resolution for some parts of the frame can prevent visible artifacts from appearing elsewhere and hence can result in bit improvement in picture quality.

B. Determination of a frame partitioning mask

The decision of whether or not a macroblock is labeled significant or insignificant is based on three factors: the dc error, the variance error and the absolute error.

The dc error is defined as the difference between the dc components of the original and the predicted macroblock after the DCT. The errors are found for the four 8x8 blocks of the luminance component of each macroblock. The maximum of these errors is chosen to be the dc error of the macroblock.

The absolute error is defined as the average pixel by pixel difference between the original and the predicted macroblocks. Similar to the dc error, the final value of the absolute error equals to the maximum of four absolute block errors.

The variance error is the difference between the average variances of the predicted macroblock and the current original macroblock.

A macroblock is classified as insignificant if all three errors do not exceed their corresponding threshold values.

3. BIT ALLOCATION AND RATE CONTROL

To maximize the effect of our frame partitioning technique, we developed a new bit allocation algorithm. Our algorithm presents a totally unique approach to bit allocation and rate control. The algorithm can potentially work in a real time scenario, but gives best results in two pass encoding with a delay of one GOP.

There are several distinctive features in our algorithm. First, we take the bit allocation process to a macroblock level without introducing an unreasonable increase in computational complexity. Then, we make the quality of each reconstructed macroblock the basis for all bit allocation decisions. Last, in order to ensure the validity of those decisions, we adopt an original strategy of "rate control first, bit allocation second" in determining the second pass encoding parameters (macroblock and frame targets as well as quantization factors).

A. Bit allocation based on macroblock quality

In a standard approach to bit allocation, a target number of bits is determined on a frame level. A target for each macroblock is derived as a fraction of the frame target. The exact relation between frame and macroblock targets depends on a preset bit distribution rule. For example, Test Model 5 [4] uses a uniform bit distribution, i.e., all macroblocks are targeted for the same number of bits. This distribution, as any other preset distribution, fails to recognize differences between the macroblocks themselves. As a result, the quality of the reconstructed image can vary considerably from one macroblock to the other.

The only way to ensure consistent quality throughout a frame is to take the bit allocation to the macroblock level. The question of consistent quality for all macroblocks of all frames in a GOP is even more important if we use the frame partitioning technique. The danger here is having a "perfect" quality "background" and an "object" encoded at a much lower quality or exactly the opposite. Frame partitioning can be successfully used only if we can achieve seamless quality borders between the background and the object, i.e., the quality of all reconstructed macroblocks must be very close.

We propose to base the decisions of macroblock bit allocation on a measurement of quality of the reconstructed blocks. There are a number of existing techniques for visual quality assessment. Most of them involve measuring different kinds of artifacts and then using a weighted sum of those measurements to come up with the final number. In order to limit the computational complexity, we use just one indicator of visual quality, signal to noise ratio (SNR):

SNR = 10 log (macroblock variance/macroblock mean square error)

Now we can state the objective of our bit allocation algorithm: to achieve a flat (same) SNR for all reconstructed macroblocks in all frames in a GOP. Since the variance of a macroblock in a frame is known, the bit allocation goal can be rephrased in terms of mean square error (MSE): we have to allocate such number of bits for each macroblock so that their resulting MSE would correspond to the same SNR value.

B. Rate control first, bit allocation second

The allocated target number of bits is accomplished through modifying the quantization factors (mquant) of each macroblock in a frame. In order to achieve the flat SNR, we would have to know the correspondence between MSE, mquant and the resulting number of bits (BITS) for all macroblocks. We will call those dependencies the rate control curves: MSE vs. mquant curve and BITS vs. mquant curve.

The task of determining the curves can be computationally complex. The exhaustive search for the desired MSE for each macroblock is not realistic, especially if the task involves processing data for all frames in a GOP. Encoding all macroblocks with a range of mquant and measuring their MSE and BITS for several frames is not practical either. We propose a way to reduce complexity and still determine the necessary rate control functions.

The resulting number of bits and MSE for macroblocks with similar variances is almost the same. We can use this fact in order to collect all of the rate control statistics in one preliminary encoding pass. We assemble all frame macroblocks, depending on their variances, into adjacent groups or variance bands. Now, we can estimate a single set of rate control curves for each band and use these statistics for bit allocation.

We consider 6 mquant samples spread through the possible range of mquant to be sufficient for estimation (for example 2, 4, 8, 16, 32 and 62). The range of variances in each band depends on the number of macroblocks that will be encoded with the same mquant. We found that we can obtain better estimation if we increase the number of macroblocks for each mquant rather than increasing the number on variance bands. A sufficient precision of estimation can be obtained with as little as 6 variance bands for a frame. By quantizing macroblocks with slightly different variances with the same mquant, we can obtain a good estimation of the value of MSE and BITS for a virtual macroblock with a variance equal to the average variance of a band.

An example of two BITS/mquant and MSE/mquant curves for two variance bands is shown in Figure 1.

Now, from rate control curves we can predict the number of bits needed to encode each macroblock at a certain SNR and its corresponding quantization factor.

If the encoding scenario allows a one GOP delay, then rate control curves are estimated for all frames in a GOP. For I frames, variance bands are assembled based on the variances in the original image. For P frames the variances of the predicted error image are used to group the macroblocks into bands. After gathering rate control statistics for all frames in a GOP, we consolidate the information into two sets of rate control curves: one for I frames and one for P frames. The separate approach to I and P frames corresponds to different quantization matrices in their processing. Flat quantization matrix is used for P frames, while the coefficients in I frames and Intra macroblocks in P frames are quantized with factors increasing towards higher frequencies.

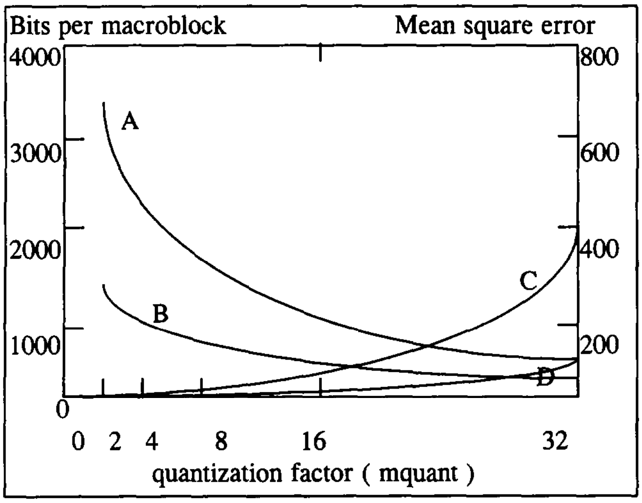

Fig.1. Rate control curves:
A, B - macroblock bits versus mquant
C, D - mean square error versus mquant
A, C – rate control curves for a variance band with avgvar=2300
B, D - rate control curves for a variance band with avgvar=200

I and P rate control curves allow us to clearly estimate the targets for the second pass encoding. For a constant bit rate, the algorithm can evaluate the highest macroblock SNR that is possible to achieve given the bit rate and VBV buffer limitations. In case of a variable bit rate, the algorithm can assess the number of bits required to encode the current GOP at a certain SNR level, with all macroblocks of all frames encoded with the same SNR.

The bit allocation/rate control estimation of pass 1 results in several pieces of information. One is the target mquants for all macroblocks of all frames in a GOP. Second is the target number of bits for each frame and the percentage bit distribution between the frames in a GOP. Third is the percentage distribution of macroblock bits for all frames in a GOP. During pass 2 encoding, the distribution of macroblock bits is used to adjust the macroblock quantization factors from their preset values.

C. Bit allocation and rate control algorithm

Our bit allocation algorithm consists of the following steps.
Encoding, first pass (can be over one frame or one GOP):

- Determine the place of each macroblock in a frame on a variance scale.
- Group the macroblocks into variance bands.
- Quantize macroblocks using sample mquant values (for example 2, 4, 8, 16, 32, and 62).
- After the quantization, get the number of bits for each macroblock and their mean square error (MSE).

Approximate two sets of rate control curves. One set will be used for I frames and Intra macroblocks of P frames, another will be used for Inter macroblock of P frames.
Bit allocation, after the first pass:

- Set an arbitrary signal to noise ratio (SNR) as the target quality for the second pass encoding.
- For each significant macroblock, determine MSE required to achieve the target SNR. For this step and the next two steps, use the piece wise approximation of the rate control curves. If possible, use two sets of rate control curves in current macroblock estimation: rate control curves of the band with its average variance above and below the variance of the current macroblock.
- For each significant macroblock, determine the MQUANT corresponding to the MSE found in the previous step and determine the number of bits that it will produce.
- Check if the total number of bits for all significant macroblocks exceeds the bit-rate/buffer size requirements or if it is too small. If either is correct, respectively decrease or increase the target SNR and repeat the bit allocation. Otherwise, go to the next step.
- Determine the local frame bit distribution model for a GOP and a local macroblock bit distribution model for a frame.

Encoding, second pass:

- Quantize the first significant macroblock using the target MQUANT determined in the bit allocation.
- Depending on the relationship between the resulting and the expected number of bits, modify the next MQUANT. The MQUANT is modified percentage wise from its target MQUANT using the standard adaptive quantization techniques ([4]) with the local macroblock distribution model instead of the uniform model.

4. RESULTS

Standard MPEG verification sequences were used to test the proposed ideas. Each sequence was composed of 352 x 240 frames, 330 macroblocks per frame. The testing was performed using one GOP (10 frames in our tests) for all four sequences.

The tests show that the size of the frame partitioning mask can be very significant. Table 1 presents the number of macroblocks that were found to be insignificant in encoding.

TABLE 1
Size of the frame partitioning mask in macroblocks

Frame	Sequences Tennis	Garden	Football	Mobile
1 , I	0	0	0	0
2, P	260	185	303	220
3, P	224	198	290	131
4, P	226	146	296	184
5, P	204	173	287	135
6, P	167	172	290	163
7, P	218	165	279	142
8, P	205	167	274	168
9, P	214	138	256	173
10, P	234	127	266	161

To illustrate the bit rate improvement of frame partitioning over standard motion estimation we use encoding with fixed mquant. We compare the performance of our algorithm with the standard MPEG-2 solution. In the case of the same quantization for all macroblocks, the quality of both streams would be the same. Table 2 shows the number of bits required to encode 10 frames with mquant fixed at 12.

TABLE 2
Size of the encoded files

	Bits to encode one GOP	
Sequences	No frame partitioning	With frame partitioning
Tennis	741008	334048
Football	1060344	868680
Garden	1318016	1037120
Mobile	1926656	1525488

The results of the tests show that frame partitioning improves the encoding efficiency by 21 per cent for "garden" and "mobile" sequences, by 19 per cent for "football" and by 55 percent for "tennis" sequence.

To test our bit allocation algorithm, we compare the SNR statistics for a single I frame encoded with our method and with Test Model 5 algorithm. In both cases the resulting number of bits for a frame was the same (88 k bits). The comparison is illustrated in Table 3. The Table shows minimum, maximum and average signal to noise ratios for a frame, as well as variance of the SNR as a measure of consistency of quality.

TABLE 3
SNR statistics comparison

	SNR statistics			
Encoding algorithm	min SNR	max SNR	avg SNR	var SNR
Our bit allocation algorithm	4.0	10.1	6.1	2.9
Model Test 5 algorithm	-0.3	18.1	6.0	9.8

In Table 1 through 3 we illustrated how selected parts of our algorithm improve encoding. The combination of the frame partitioning technique with our bit allocation/rate control method allows us to attain an even higher level of efficiency and quality of encoding.

5. CONCLUSIONS

The idea of frame partitioning embeds the concept of a video object within the syntax of MPEG-2 standard. Frame partitioning allows us to use different temporal resolutions for more and less significant parts of frames in a video stream. By giving up high temporal precision in the "background" areas, we can attain higher than otherwise quality in the visually important areas of the picture. The macroblock based bit allocation strategy allows us to achieve consistent picture quality throughout a frame. The proposed rate control algorithm carries out compression that is consistent with allocated bit targets. The above scheme improves the compression bit rate by an average 20 percent and eliminates fluctuations in visual quality.

References

[1] Leonardo Chiariglione, "MPEG and multimedia communications", IEEE Trans. Circuits Syst. Video Technol., vol. 7, no. 1, pp. 12-19, Feb. 1997.
[2] Wei Ding and Bede Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling", IEEE Trans. Circuits Syst. Video Technol., vol. 6, no. 1, pp. 12-19, Feb. 1996.
[3] Atul Puri and R. Aravind, "Motion-compensated video coding with adaptive perceptual quantization", IEEE Trans. Circuits Syst. Video Technol., vol. 4, no. 4, pp. 351-361, Dec. 1991.
[4] ISO-IEC/JTC1/SC29/WG11, "Test Model 5", Draft, Apr. 1993.