MPEG Bit Rate Improvement Using Adaptive GOP

H. Farouk[1]
Computers and Systems Dept.
Electronic Research Institute
Heshamali68@hotmail.com
EGYPT

Abstract-- MPEG coding algorithm is a full motion compensated DCT. In MPEG coding, the video sequence first divided into groups of picture of frames (GOP), (I,P,B frames). The algorithm for adaptive GOP length as I frame position, depending on the indexing technique, In This paper an algorithm was developed to variable GOP length which was resulted in minimizing the bit rate with average 10% to 15% from the classical MPEG coding technique while keeping the same SNR with respect to the standard MPEG coding algorithm.

Keyword--: bit rate, GOP, motion compensation, video compression

I. Introduction

MPEG coding algorithm is a full motion compensated DCT. In MPEG coding, the video sequence first divided into groups of picture of frames (GOP). Each group may include three types of pictures or frames: intra code of (I) picture of frame, Predictive – coded (P) picture of frame and bidirectional productively coded (B) frame or picture as shown in Figure 1. I pictures are coded by intra frame technique only with no need for previous information. In other words, I pictures are self-sufficient. They are used as anchors for forward and/or backward prediction. P-pictures are coded using one-directional motion compensated prediction from a previous anchor frame, which could be either I- or P-picture. The distance between two nearest I-frames is denoted by N, which is the size of GOP.

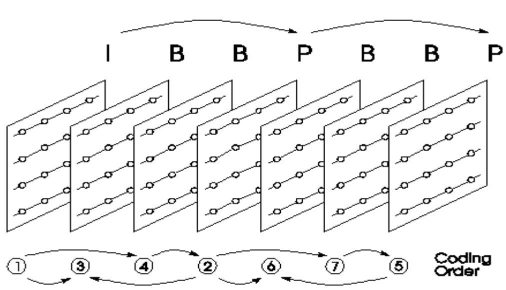

Figure 1: GOP Structure and coding seqence

The distance between two anchor frames is denoted by M. Parameters N and M both are user selectable parameters, which are selected by the user during the encoding. A large number of N and M will increase the coding performance but cause error propagation or drift. Usually, N is chosen from 12 to 15 and M from 1 to 3. If M is selected to be 1, this means no B-picture will be used. Last, P picture can be coded using prediction form either past or future anchor frames (I or P), or both. [1-7]

Section II will discuss the related work. Section III will discuss why we need adaptive I frame and then introducing the way to make I frame adaptive in section IV. The results will be given in section V.

II. Related Work

The best work in this area was done by presenting a new algorithm that adaptively selects the best possible reference frame for the predictive coding of generalized, or multi-view, video signals, based on estimated prediction similarity with the desired frame. They defined similarity between two frames as the absence of occlusion, and they estimated this quantity from the variance of composite displacement vector maps. The composite maps are obtained without requiring the computationally intensive process of motion estimation for each candidate reference frame. They provided prediction and compression performance results for generalized video signals using both this scheme and schemes where the reference frames were heuristically pre-selected. When the predicted frames were used in a modified MPEG encoder simulation, the signal compressed using the adaptively selected reference frames required, on average, more than 10% fewer bits to encode than the non-adaptive techniques. The disadvantage of this technique is the computational complexity in calculating the similarity between the frames.

III. Why to Make I Frame Adaptive?

The value of N and M is fixed for a session i.e. for one encoding session. But in our work we can change the value of N to fit the shot cut since we will program the encoder to

[1] Manuscript received February 5, 2007; Revised June 12, 2007 This work was supported Electronic Research Institute. Department Computers and Systems, Cairo, Egypt.

access data file with the frames number represent shot start which is the output of the indexing algorithm described in [10]. The fixing of N and M ignore the nature of shots with respect to its content by meaning that sometimes we have motion with short shots or fast motion, fast content changes internally on the frame level, or long shots with slow changes and so on.

In this work we take into consideration the nature of video content with respect to the speed of changes and the shot cut to change N within the same session. With variable N we can select the location to use I frame, this location will help in optimizing the bit rate since we can put I frame as far as possible during slow motion or long shots. Because with the classical MPEG coding algorithm we can consider I frame within the same shot, where it is preferred to code all relative frames as P or B frame as long as we are in the same shot. Coding Equal distance I frame, as classical MPEG does, will produce a large bit rate as table 1 tell us that the average bit rate for I frame (150 kbit/sec) is 3 times the average bit rate of frame P (50 kbit/sec) or 7.5 times the average bit rate of B frame (20 kbit/sec).[8]

Level	I	P	B
30 Hz SIF @ 1.15 Mbit/sec	150,000	50,000	20,000
30 Hz CCIR 601 @ 4 Mbit/sec	400,000	200,000	80,000
Level	**Average**		
30 Hz SIF @ 1.15 Mbit/sec	38,000		
30 Hz CCIR 601 @ 4 Mbit/sec	130,000		

Table 1: Pictures average bit rate

IV. How To Make I Frame Adaptive?

The I frame is made adaptive with variable position in the coding sequence since each frame type will be marked with its type through the coded video stream. We can use the output of the algorithm described in [9] for indexing video and detect the shot cut position either it is shot cut or gradual effect as shown in Figure 2, Figure 3 as frames difference calculation in both cases.

If the video shot (after indexing) is pan or tilt or zoom we can determine a relation between the numbers of pixels shift from one frame to the next one and if there is a need for a new frame. This will help in both encoding time and bit rate minimization as it could be applied online [10-12].

To classify the video stream into shots based on camera breaks, we use a technique that depends on histogram differences of the DC coefficient of the DCT blocks of coefficients in the compressed domain, or just compute the DC coefficient in the spatial domain. The DC coefficient only is effective because it contain the most entropy of the spatial block data

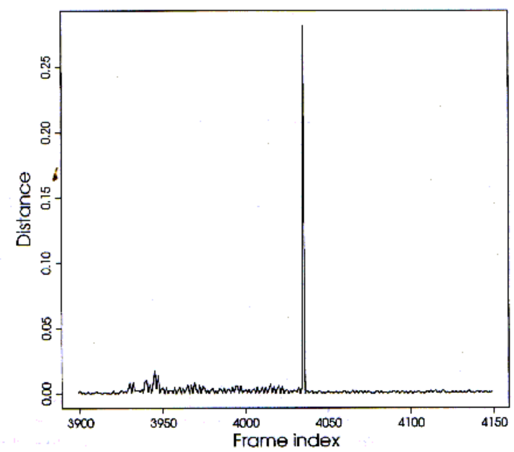

Figure 2: Sudden Change Shot Cut detection

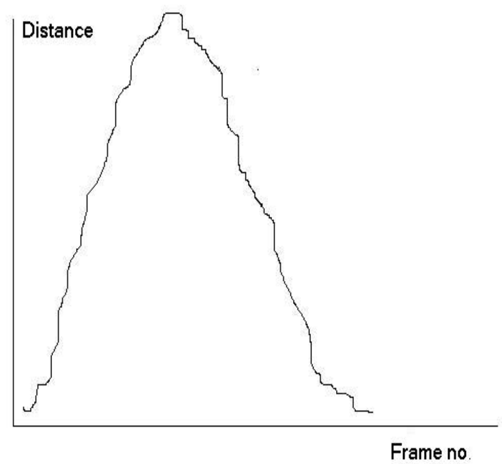

Figure 3: Gradual Change Shot Cut detection

Then the histogram will be available for any frame by distributing the calculated DC coefficient on the set range (from 0 to 4096 for uncompressed domain or from –1024 to +1024 for the compressed domain)

For each two successive frames we will calculate the histogram difference by subtracting each histogram value in a frame from the same one in the next frame, even if it is zero in any of them, and accumulate the magnitude of the differences (at the According to certain threshold we will process the histogram differences and any value exceed this threshold we will consider a cut between these 2 frames resulted in this difference value and consider the second frame of this pair of frames as a key frame which, could be saved in a library to express this shot information and be able to retrieve this shot. If we have a stream of 1000 frames we will have 999 histogram difference values for each 2 successive frames

We will consider allocating I frame in the position of shot cut mandatory as starting new frame information. This will help in bit rate minimization. We use this method in our technique.

V. Results

The tests given below keep the same SNR, as shown in Figure 1, because this algorithm may searches for a not found block and will code it as a new one.

In our tests we apply many cases. We consider: GOP=40, GOP=60, GOP=240, GOP=300 and finally put GOP >= Frames number of the tested sequence. The tests given below are for GOP>= Frames number of the tested sequence

A. *Test 1:*

We apply test case depending on the indexing algorithm for a sequence of 300 frames (the test sequence of the indexing algorithm which is containing 12 keyframes) by taking F=300 and consider the keyframes resulting from the indexing algorithm. This gives a SNR difference around zero between the coding with variable I frame position and the classical MPEG coding algorithm as shown in Figure 4.

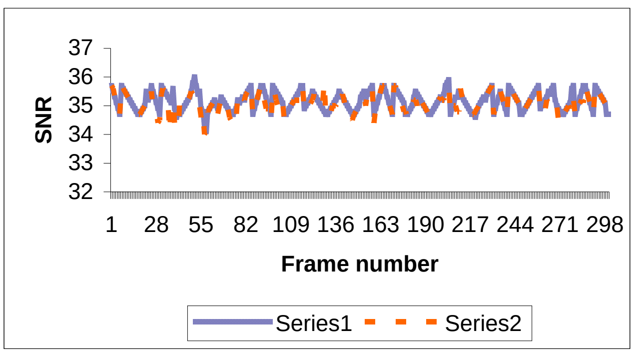

Figure 4: SNR for sequence Nike1.mpg coded with MPEG with respect to the original stream (series 1) is identical to sequence processed with adaptive I frame with respect to the original stream (series2).

Consider MPEG stream with GOP=15
IBBPBBPBBPBBPBP
with N=15, M=3 Where:
N is the length of GOP
M is the distance between I and P

With classical MPEG coding, the 300 frames will be divided into:

Number of GOP = 300 /15 =20
I frames = 1x20 = 20 (1 per GOP)
P frames =5x20 = 100 (5 per GOP)
B frames = 9x20 = 180 (9 per GOP)

According to table 1: this video sequence average bit rate is = (20x150) + (100x50) + (180x20)/ (no. of sec.)
= 11600/10 kbits/sec
= 1160 kbits/sec

With adaptive I frame, we will consider the I frame position is only the position of keyframes resulted from the indexing technique described in [6-7]

So, the 300 frames are divided into:
I frames = 12 frames (The number of keyframes)

P frames = (300 –12)/3 =96 frames
(Dividing over 3 because the predicted frames P and the bi-directional prediction B will compose groups of (BBP) so, P will be repeated each 3 frames)

B frames = 300 – (12 + 96)= 192 frames
According to table 1 : this video sequence bit rate average is
= (12x150)+(96x50)+(192x20)/no. of seconds = 10440/10 kbits/sec = 1044 kbits/sec

The bit rate is minimized by the factor:
(1-(1044/1160))x100 = 10 %

Test 2:

In this test we use a part of an actual movie (home alone 3):
From frame 7:9:00 to Frame 11:19:00
with frame rate 30 frame/sec.
(The format is Minutes:Second:frame number)
The number of frames in this test =
((11-7) x 60 + (19-9)) x 30 = 7500 frames
The number of keyframes in this test = 60 frames

For the same GOP as the previous test:
With classical MPEG coding, the 7500 frames are divided into:
Number of GOP = 7500 /15 =500
I frames = 1x500 = 500 (1 per GOP)
P frames =5x500 = 2500 (5 per GOP)
B frames = 9x500 = 4500 (9 per GOP)

According to table 1 : this video sequence bit rate average with classical MPEG coding is:
= (500x150) + (2500x50) + (4500x20)/ (no. of sec.)
= 290000/250 kbits/sec
= 1160 kbits/sec

I frames = 60 frames
P frames = (7500 – 60)/3 = 2480 frames
B frames = 7500 – (60 + 2480)
= 4960 frames

According to table 1 and by adding 10% of the average P type size to P type itself as the not found block will be coded as still block. This is not the same for test 1. Test 1 is composed of rather small shots. So, this video sequence average with adaptive I frame for test 2 is

= (60x150) + (2480x55) + (4960x20)/ (no. of sec.)
= 244600/250 kbits/sec
= 978.4 kbits/sec.

The bit rate is minimized by the factor:
(1-(978.4/1160)) x100 = 15.6 %

So, from this tests, test1 and test2:

Test 1 represents short shots with many I frames required and the bit rate for frames number 300. The minimization average is 10%

Test 2 represents real movie with smooth, moderate width shot with relatively smaller number of I frames required for frames number 7500. The minimization average is 15%.

VI. Conclusion

The algorithm for adaptive I frame position, depending on the indexing technique, results in minimizing the bit rate with average 10% to 15% from the classical MPEG coding technique while keeping the same SNR with respect to the standard MPEG coding algorithm. The overhead is the video indexing calculations or any algorithm for keyframe extraction. But this will not consider a problem because the adaptive I frame position technique depends on historical calculations on the whole set of frame composes the sequence give an indexed video output which is another useful application. This means that we have no private overhead we consider this technique as an application to the indexing. This relies that this work will help in bit rate improvement for the offline applications.

VII. References

[1] ISO/IEC 11172: 'Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s'.

[2] ISO/IEC 13818: 'Generic coding of moving pictures and associated audio (MPEG-2)'.

[4] G.H. Park, M.W. Park, S Jeong, J. Cha, K. Kim, and J. Hong, "Adaptive GOP structure for SVC," ISO/IEC/JTC1/SC29/WG11/MPEG/ M11563, Hong Kong, Jan. 2005.

[3] J. Reichel, M. Wien, and H. Schwarz, "Scalable video model 3.0," ISO/IEC JTC1/SC29/WG11 N6716, Oct. 2004.

[5] Yokoyama, Y. “Adaptive GOP structure selection for real-time MPEG-2 video encoding”, International Conference on Image Processing, Page(s):832 - 835 vol.2, 2000.

[6] Yun Q. Shi Huifang Sun, “Image and Video Compression for Multimedia Engineering”, CRC press,1999.

[7] C. Stiller and J. Konrad, “Estimating motion in imaging sequence” IEEE Signal Processing Magazine, pp. 70-91, July, 1999.

[8] S. L. Regunathan, R. Zhang and K. Rose, "Scalable video coding with robust mode selection," Signal Processing: Image Communication, vol.16, (no.8), Elsevier,. p.725-32, May 2001

[9] R. Zhang, S. Regunathan and K. Rose, "Optimized Video Streaming Over Lossy Networks With Real-Time Estimation Of End-To-End Distortion," International Conference on Multimedia and Expo 2002, Lausanne, Switzerland, August 2002.

[10] H. Farouk, S. Mashali, M. Rashwan, A. Nassar, "Video Compression Improvements Based On Camera Motion" , IEEE ISSPIT Dec. 28-30, 2001.

[11] Ewerth, R.; Freisleben, B., “Improving cut detection in MPEG videos by GOP-oriented frame difference normalization”, Proceedings of the 17th International Conference on Pattern Recognition, ICPR, Page(s): 807 - 810 Vol.2, 23-26 Aug. 2004.

[12] W. Wolf, "Keyframe Selection By Motion Analysis", in Proc. IEEE, Conf. Acoust., Speech, and Signal Proc., 1996.

Dr. Farouk received his Ph.D. at 2001 from Electronics & Communications Dept., Faculty of Engineering, Cairo Univ. and his M.Sc. at 1996 from Electronics & Communications Dept., Faculty of Engineering, Cairo Univ.

Dr. Farouk is an Assistant Prof. at the Electronics Research Institute, Egypt, since 1993. His fields of research are signal processing, image compression, video processing, video compression, video indexing and retrieval, video on demand, pattern recognition and machine vision. He has many publication in this field in SPIE, IASTED, WSEAS conferences and WSEAS, ISA transaction. Dr. Farouk is a lecturer at American University in Cairo, AUC, in the field of networking administration, web mastering and developing and he also Cisco certified teacher since 2003. Dr. Farouk is an IT Projects Manager at Ministry of communications and Information Technology of Egypt since 2002. He is managing the Arabic eContent Initiative project which started in mid of 2004 and initiated formally in may 2005 with Egyptian publisher federation. He also supervise many governmental projects, portals development and information infrastructure building. Dr. Farouk supervises the international agreement of Microsoft with the Egyptian government. Dr. Farouk received a certificate in the project management from IESA Univ. SPAIN as attending course held with cooperation with Nile Univ. at 2004