EFFICIENT INTERFRAME TRANSFORM CODING USING TEMPORAL CONTEXT

Yui-Lam CHAN and Wan-Chi SIU

Department of Electronic Engineering
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
Tel: (852)27666229 Fax: (852)23628439 email: enwcsiu@hkpucc.polyu.edu.hk

ABSTRACT

Three dimensional (3D) transform coding can reduce the interframe redundancy among a number of consecutive frames, while the motion compensation technique can only reduce the redundancy of at most two frames. The former is very efficient when the correlation between interframe pixels is high. However, the performance will be degraded for complex scenes with a large amount of motions. This paper presents a three dimensional discrete cosine transform (3D-DCT) coding with variable temporal lengths, which is based on the local temporal activity. Two scene change detectors are used to detect the local temporal activity. Our idea is to let the motion activity in each block (with variable temporal length) be very low, while the efficiency of the 3D-DCT coding is increased. Through intensive computer simulations, the performance of the proposed 3D-DCT coding has been found to have substantial improvement over the conventional fixed length 3D-DCT coding. Furthermore, it is significant to point out that the performance of our proposed algorithm is better than that of the MPEG coding.

1. INTRODUCTION

Three dimensional (3D) transform coding has the advantage of reducing the interframe redundancy for more than one frame. Besides, it is prevented from an infinite propagation of transmission errors. The price to be paid is a longer delay and a large memory requirements. Thus, it is an alternative approach to hybrid coding concepts used in today's standards. Adaptive three dimensional discrete cosine transform(3D-DCT) coding[1-4] has been reported to be comparable to hybrid coding (motion compensation prediction along the temporal axis[5-6] and DCT frequency coding in the spatial domain). Furthermore, it has an asymmetric property with the decoding being much faster than the encoding and the computation complexity is even lower than that required for the MPEG like coder [7]. The 3D-DCT coding is very efficient when the amount of motions is small. In other words, the energy distribution in transform domain is more compact, therefore it is easier to code. The performance of the 3D-DCT coding will be affected in a complex scene with a large amount of motions.

In this paper, we use a variable temporal length 3D-DCT instead of a fixed length 3D-DCT. The temporal length of the 3D-DCT varies with local temporal activities. Thus, the motion activity in each block is still very low, while the coding efficiency becomes high. Our coding tests using computer simulation show that this technique is indeed very efficient. It has great improvement as compared with the fixed length 3D-DCT. Furthermore, the performance of our proposed algorithm is better than that of the MPEG coding.

Acknowledgment : This work was supported by The Croucher Foundation PolyU340/055.

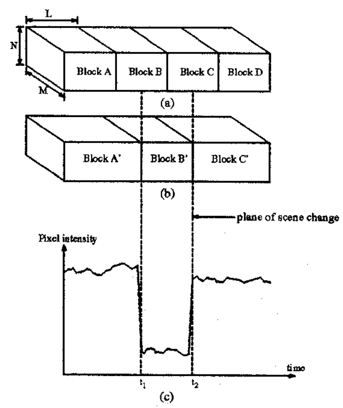

Figure 1: (a) Fixed length 3D-block (b) Variable temporal 3D-block (c) Pixel intensity which varies with time

2. THE PROPOSED VARIABLE TEMPORAL LENGTH 3D-DCT CODING

Theoretically, the most optimal transformation of the Markov-I type signals is the Karhunen-Loève transform (KLT) [8]. The KLT can completely decorrelate the signal in the transform domain, thus a high compression ratio can be achieved. It is well-known that the implementation of the KLT is difficult, while the DCT-II has been found asymptotically equivalent to the KLT for Markov-I signals as the correlation (ρ) between the pixels tends to 1 [8]. In conventional 3D-DCT coding, if a 3D-block has low frame-to-frame motion (the interframe pixels correlation is close to 1), then only the coefficients having low temporal frequency need to be transmitted. The coding is very efficient when the amount of motions in the block is small. If the $L \times M \times N$ pixel block is fixed, the coding performance will be degraded for complex scenes with a large amount of motions. Figure 1(a) illustrates the problem of a fixed length 3D-DCT coding. The temporal motion activity is shown

0-7803-3073-0/96/$5.00 ©1996 IEEE

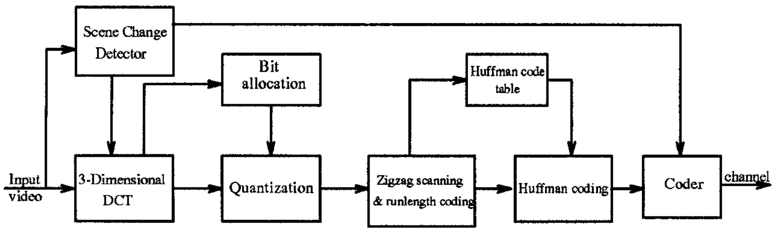

Figure 2: Block diagram of our proposed variable temporal length 3D-DCT coding system

in figure 1(c). In the figure, there is a scene change at the time t_1 and t_2 in the image sequence. So, inside the three dimensional blocks B and C, there are high motion activities and low interframe pixels correlations. This causes the high frequency coefficients in these transform blocks having significant values. Furthermore the distortion introduced by the coding process will probably spread over the whole 3D-block and be visible as they last for a long time on the decoded image sequence. In this case, the coding efficiency will be significantly decreased. In this paper, we propose a 3D-DCT coding with an adaptive adjustment of the length of the 3D-block in the time direction as shown in figure 1(b). It depends on the local activity in the image sequence. In other words, the temporal length, L, is varied instead of a fixed one. These blocks will remain in low motion activity and high interframe pixels correlation. Thus, this variable temporal length method could take the advantage of the 3D-DCT coding and achieve high coding efficiency.

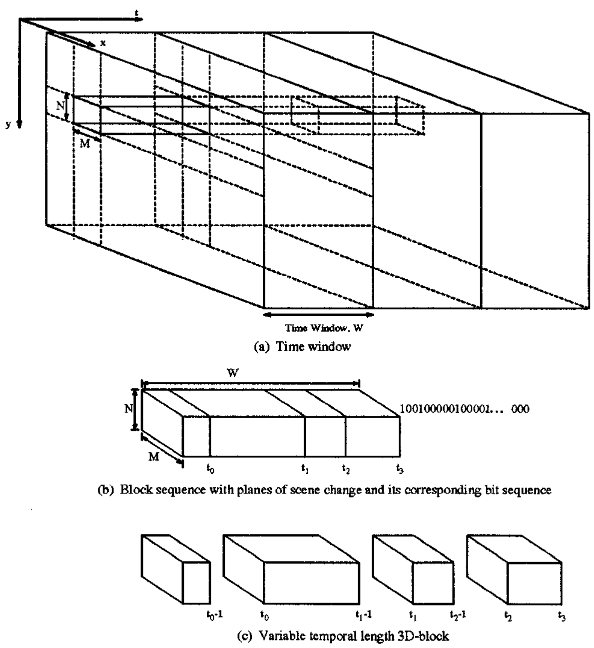

Figure 3: Details of variable temporal length 3D-DCT coding

Figure 2 illustrates the block diagram of our proposed variable temporal length 3D-DCT coding system. Each image sequence is divided into a number of time windows, W, which is a fixed number of image frames from the original image sequence as shown in figure 3(a). In order to determine the an appropriate cutting plane, a scene change detector has to be firstly used to detect the scene change before the interframe transformation can be done for the block. Two approaches are used in this paper. The first one is frame difference method. The frame difference (FD_k) of the block in successive frames $k+1$ and k is defined as :

$$FD_k = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} |I_{k+1}(x,y) - I_k(x,y)|$$

where $I_k(x,y)$ is the pixel value in the image sequence at time k and spatial location (x,y).

A block is called scene change if the value of FD_k in the block is greater than or equal to T_0, where T_0 is a predefined threshold. The scene change detection using frame difference is very simple. However, its accuracy is affected in the case of a number of consecutive frames having frame difference just a little bit smaller than T_0.

Another design of scene change detector is motivated by Olstad [9]. In this paper, we refer it to as adaptive temporal block filter. Assume that the block size in the spatial domain be $M \times N$ pixels. Let $F = (I_0(x,y), I_1(x,y), \ldots, I_{W-1}(x,y))$ be any consecutive pixel intensity at location (x,y) within the spatial block in a time window of W samples as shown in figure 3(a). Let also α_P represent a set of P planes of scene change within an image sequence with time window, W, and with the spatial dimensions of $M \times N$. The set α_P can then be written as a set of P indices, $t_i : 0 < t_0 < t_1 < \ldots < t_{P-1} < t_P = W$ as shown in figure 3(b). The scene change detector generates the following subsequences of the time window:

$$F_0 = (I_0(x,y), \ldots, I_{t_0-1}(x,y))$$

$$\vdots$$

$$F_{P-1} = (I_{t_{P-2}}(x,y), \ldots, I_{t_{P-1}-1}(x,y))$$
$$F_P = (I_{t_{P-1}}(x,y), \ldots, I_{W-1}(x,y))$$

In general, we can write

$$F_j = (I_{t_{j-1}}(x,y), \ldots, I_{t_j-1}(x,y))$$

In order to detect the scene change accurately, an error function (E) on the set α_P can be defined.

$$E(\alpha_P) = \sum_{j=0}^{P} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} e[F_j(x,y)]$$

where $e[(I_0, I_1, \ldots, I_{k-1})] = \sum_{i=1}^{k}(I_i - \overline{I})^2$, and $\overline{I}$ denotes the mean of $I_0, I_1, \ldots, I_{k-1}$

Hence a scene change detector can then be developed for the minimization of the error function as shown above. The global minimum for the error function and the associated localization of the scene change planes can be efficiently computed with dynamic programming [9]. The number of scene change planes, P, in each 3D-block is chosen such that the error function $E(\alpha_P) < E_0$, where E_0 is a predefined threshold. In other words, the number of scene change planes, P, in each 3D-block is increased when there is a large amount of motions of the 3D-block in the image sequence.

After the cutting planes are defined with the scene change detector in each block of sequence as shown in figure 3(b), a variable temporal length 3D-DCT is performed on each 3D-block. Suppose that the scene change detector produces a scene of P planes. This implies that the block sequence is segmented into $P+1$ 3D-blocks with different temporal lengths. Each of the 3D-block is transformed by the 3D-DCT with different temporal lengths, L, by cascading three 1D-DCTs, for $L \times M \times N$ block size. The forward transforms are defined as follows:

$$X(u,v,w) = \frac{8}{L \times M \times N} \sum_{k=0}^{L-1} \sum_{j=0}^{M-1} \sum_{i=0}^{N-1} C(u)C(v)C(w) I_k(j,i) \cos\pi\frac{(2k+1)u}{2L} \cos\pi\frac{(2j+1)v}{2M} \cos\pi\frac{(2i+1)w}{2N}$$

where $C(n) = \frac{1}{\sqrt{2}}$ for $n = 0$, otherwise $C(n) = 1$.

Except for the DC coefficient of each 3D-block, which is coded using 8 bits, the quantization of coefficients is done according to their variances and the temporal lengths of the 3D-block. Since bits allocated to different temporal length 3D-blocks may be unequal, a different quantizer is required. For instance, the number of bits allocated to the 3D-block with the shorter temporal length would be less than that of the longer one. Let B_l be the total number of bits assigned to the 3D-blocks with temporal length l, and B_l can be evaluated using the following equation:

$$B_l = B_o + lB_s$$

where B_o is the minimum number of bits for B_l and B_s is the number of bits for each frame of the 3D-block.

Then, the number of bits allocated to each individual transform coefficient is according to its variance [10]. After the bit allocation has been found, the coefficients in each 3D-block could then be quantized. The quantized coefficients in variable temporal length 3D-block are scanned in a zig-zag manner for the first frame, then followed by the successive frames. Then, the run-length coding and the huffman coding are employed to further reduce the bit rate.

The decoding requires information about the scene changes from each 3D-block sequence such that the appropriate temporal length inverse 3D-DCT can be performed. This side information can be recorded with a bit sequence 10010...010 where the number of bits is equal to the number of images in the current time window as shown in figure 3(b). A bit '1' indicates that this is the first frame in a new 3D-block. This bit sequence can be further compressed using the Lempel-Ziv coding [11] or the arithmetic coding [12]. For example, the temporal length is 3 for the first 3D-block in the above bit sequence.

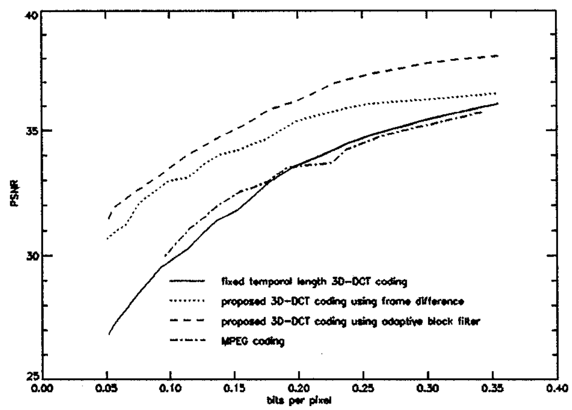

Figure 4: Plot of PSNR against bits per pixel for different algorithms

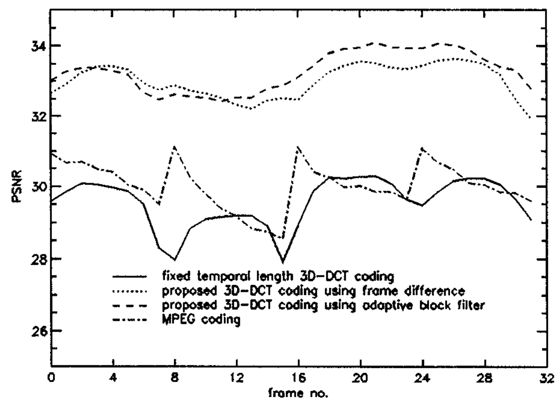

Figure 5: Result for "salesman" sequence at 0.1 bits per pixel

3. RESULT

A series of computer simulations have been conducted to evaluate the performance of the adaptive interframe transform coding. The 32 frames image sequence "salesman" with the size of 352×288 pixels has been used. The peak signal-to-noise ratio(PSNR) measure of image fidelity has been used to quantitatively evaluate the performance of the interframe coding systems.

Figure 4 provides a comparison on the peak signal-to-noise ratios (PSNR) for different algorithms and for various bits per pixel. The PSNR performance of the proposed variable temporal length 3D-DCT coding using frame difference method as scene detector has a great improvement (about

3.5dB) over the fixed temporal length ($8 \times 8 \times 8$ for each 3D-block) 3D-DCT coding at low bit rate. But, there is an improvement of 1-2dB for cases using above 0.2 bits per pixel. The PSNR performance can be further improved by using an adaptive temporal block filter. As shown in figure 4, the proposed 3D-DCT coding using adaptive temporal block filter can achieve about 4dB and 3dB improvement as compared with the fixed temporal length 3D-DCT coding at low and high bit rates respectively. The results indicate that the proposed algorithms have a significant improvement as compared with the conventional 3D-DCT coding. To further evaluate our algorithm, we have compared it with a typical scheme according to the MPEG coding [13]. The comparisons presented in this paper are based on the MPEG coding with *IPPPPPPP* group of picture. The block size is 16×16 with the maximum displacement 8 for motion estimation. The performance of the MPEG coding is better than that of the conventional fixed temporal length 3D-DCT coding for cases below 0.2 bit per pixel. But the PSNR performance of our proposed algorithm is better than that of MPEG coding, as depicted in figure 4. The MPEG coder lags behind by approximately 3-3.5dB at different bit rates.

Figure 5 shows the performance of different algorithms at 0.1 bits per pixel against frame numbers. As expected, the values of the PSNR of MPEG coding for *P*-frames are worse than that of the *I*-frames. Thus, the PSNR performance of MPEG coding is very frustrated. The same is also true for the conventional 3D-DCT coding. But, the PSNR performance of our proposed algorithms is better, and the values of the PSNR of the reconstructed frames are very smooth. It implies that our proposed algorithms are less suffering from the "jerkiness" of the motion. Besides, The algorithms also have the advantage over the conventional 3D-DCT coding that they are free of annoying block distortion. These results clearly verify the effectiveness of the proposed scheme.

4. CONCLUSIONS

We propose a variable temporal length 3D-DCT coding which adaptively adjusts the temporal length of the 3D-block. The interframe correlation within each 3D-block is high such that the advantage of 3D-DCT can be fully utilized. The adjustment of the temporal length of each 3D-block depends on local activities in the image sequence. Two scene change detectors have been used to determine the local activity of the block sequence. A series of computer simulations show that the proposed variable temporal length 3D-DCT coding can achieve reconstructed image sequences with good quality at different bit rates. It has a significant improvement as compared with the fixed temporal length 3D-DCT coding using different scene change detectors. Besides, the PSNR performance of our proposed algorithm has 3-3.5dB improvement as compared with that of the MPEG coding.

Results of our study indicate that significant gain over the standard hybrid coding techniques used nowadays is possible through the 3D-DCT coding with variable temporal length. However, many problems still remain to be resolved, including the high memory requirement and the design of perfect optimal scene change detector. Nevertheless, three dimensional frequency coding is a fruitful direction that should not be forgotten for studying techniques for video compression.

5. REFERENCES

[1] J.A.Roese and W.K.Pratt, "Interframe Cosine Transform Image Coding", IEEE Transactions on Communications, Vol-25, Nov. 1977, pp.1329-1338.

[2] Olivier Chantelou, Christian Remus, "Adaptive Transform Coding of HDTV Picture", International Workshop on Signal Processing of HDTV, L'Aguila, Italy, Elsevier Science Publishers, Feb-Mar.,1988, pp.231-8.

[3] Jong Won Kim, Shin Ho Lee and Sang Uk Lee, "On the adaptive three-dimensional transform coding techniques for moving images" Signal Processing: Image Communication, April, 1995, pp. 525-543.

[4] Yui-Lam Chan and Wan-Chi Siu, "A New Adaptive Interframe Transform Coding using Directional Classification" Proceedings, IEEE Int. Conf. on Image Processing, No.2, pp.977-981, Nov. 1994, U.S.A.

[5] Yui-Lam Chan and Wan-Chi Siu, "An Adaptive Multiple-Candidate Hierarchical search for Block Matching Algorithm", Electronics Letters, vol. 31, no.9, pp. 1637-1639, Sept. 1995, U.K.

[6] Yui-Lam Chan and Wan-Chi Siu, "A New Block Motion Vector Estimation using Adaptive Pixel Decimation", IEEE Transactions on Circuits and Systems for Video Technology, Vol.6, No.1, Feb.,1996

[7] Yui-Lam Chan and Wan-Chi Siu, "Fast Interframe Transform Coding Based on Characteristic of Transform Coefficients and Frame Difference" Proceedings, IEEE Int. Sym. on Circuits and System, pp. 449-52, Vol. 1, April-May 1995, Seattle, U.S.A.

[8] K.R.Rao and R.Yip, Discrete Cosine Transform-Algorithms, Advantages and Applications. New York: Academic Press, 1990.

[9] B.Olstad, "Adaptive Temporal Decimation for Video Compression Algorithms" Journal of Electronic Imaging, Vol.2, Jan., 1993, pp.5-18.

[10] A.Segall, "Bit Allocation and Encoding for Vector Sources" IEEE Transactions on Information Theory, Vol.22, No.2, Mar., 1976, pp.162-169.

[11] Jacob Ziv and Abraham Lempel, "A Universal Algorithm for Sequential Data Compression" IEEE Trans. on Inf. Theory, Vol.23, No.3, May, 1977, pp.337-43.

[12] Paul G. Howard and Jeffrey Scott Vitter "Arithmetic Coding for Data Compression" Proceedings of the IEEE, Vol. 82, No.6, June 1994, pp.857-865

[13] D.J. LeGall, "MPEG: A Video Compression Standard for Multimedia Applications", Communications of the ACM, Vol-34, April 1991, pp.47-58.