[13] T. R. Fischer, "A pyramid vector quantizer," *IEEE Trans. Inform. Theory*, pp. 568–583, July 1986.
[14] R. Rinaldo and G. Calvagno, "Coding by block prediction of multiresolution subimages," *IEEE Trans. Image Processing*, vol. 4, pp. 909–920, July 1995.
[15] J. D. Johnston, "A filter family designed for use in quadrature mirror filter banks," in *Proc. IEEE ICASSP*, Apr. 1980, pp. 291–294.

Variable Temporal-Length 3-D Discrete Cosine Transform Coding

Yui-Lam Chan and Wan-Chi Siu

***Abstract*— Three-dimensional discrete cosine transform (3-D DCT) coding has the advantage of reducing the interframe redundancy among a number of consecutive frames, while the motion compensation technique can only reduce the redundancy of at most two frames. However, the performance of the 3-D DCT coding will be degraded for complex scenes with a greater amount of motion. This paper presents a 3-D DCT coding with a variable temporal length that is determined by the scene change detector. Our idea is to let the motion activity in each block be very low, while the efficiency of the 3-D DCT coding could be increased. Experimental results show that this technique is indeed very efficient. The present approach has substantial improvement over the conventional fixed-length 3-D DCT coding and is also better than that of the Moving Picture Expert Group (MPEG) coding.**

I. Introduction

Three-dimensional (3-D) transform coding [1]–[3] is an alternative approach to the motion compensation transform coding (MCTC) technique used in today's video coding standards [4]. In video coding, the application of the discrete cosine transform (DCT) along the temporal axis is advantageous over motion compensation prediction schemes because the structure can be nonrecursive, which avoids infinite propagation of transmission errors. Besides, algorithms for adaptive 3-D DCT coding [5], [6] have been reported to be comparable to MCTC technique in certain kinds of image sequences. Furthermore, the 3-D DCT coding has an asymmetric property with decoding much faster than encoding and the computational complexity is even lower than that required for the Moving Picture Expert Group-like (MPEG-like) coder [7]. The price to be paid is a longer encoding delay and the requirement for a large memory size.

The 3-D DCT coding is very efficient when the amount of motion is low. This is a typical case that the amount of energy in the higher frequency components is low; hence, the energy compaction could be good. But, the performance of the 3-D DCT coding will be affected by complex scenes with a great amount of motion. In this paper, we use a variable temporal length instead of a fixed length for a 3-D block. The temporal length varies with local temporal activities. Thus, the motion activity in each block is still very low, while the 3-D DCT coding efficiency becomes high. Experimental results show that this technique is a great improvement over fixed-length 3-D DCT coding, while it could always achieve better quality comparing to that of the MCTC technique.

Manuscript received May 5, 1995; revised September 9, 1996. This work was supported by The Croucher Foundation under Grant PolyU340/055. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. A. Murat Tekalp.

The authors are with the Department of Electronic Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong (email: enwc-siu@hkpucc.polyu.edu.hk).

Publisher Item Identifier S 1057-7149(97)03088-1.

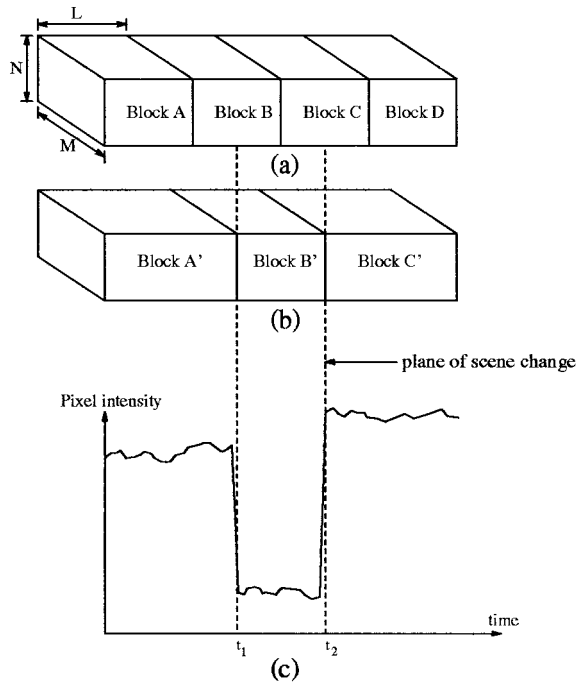

Fig. 1. (a) Fixed-length 3-D-block. (b) Variable-length 3-D block. (c) Pixel intensity, which varies with time.

II. The Proposed Variable Temporal-Length 3-D DCT Coding

It is well known that the theoretical coding performance of the DCT could be nearly equivalent to that obtained by the optimal Karhunen–Loève transform for highly correlated data [8]. Correspondingly, in conventional 3-D DCT coding, if a 3-D block has low frame-to-frame motion (the interframe pixels correlation is high), then only the coefficients having low temporal frequency need to be transmitted. However, the coding performance will be degraded for complex scenes with a large amount of motion. Fig. 1 illustrates the problem of fixed-length 3-D DCT coding. A possible pattern of the temporal motion activity is shown in Fig. 1(c). In this figure, there are scene changes at time t_1 and t_2 in the image sequence. So, 3-D-blocks B and C as shown in Fig. 1(a) could be considered to have high motion activities. This causes the high-frequency coefficients in these transform blocks having significant values. The distortion introduced by the coding process will probably spread over the whole 3-D 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 correspondence, we propose 3-D DCT coding for an adaptive adjustment of the length of the 3-D block in the temporal direction. It depends on the local activity in the image sequence. The temporal length is varied instead of a fixed one as shown in Fig. 1(b). The 3-D blocks, A′, B′ and C′, will remain to have low motion activity and high interframe pixels correlation. Thus, this variable temporal-length approach could take advantage of the 3-D DCT coding and achieve high coding efficiency.

Fig. 2(a) illustrates the block diagram of our proposed transform coder. First, the image sequence is divided into a number of time windows, W, which is a fixed number of image frames of the original

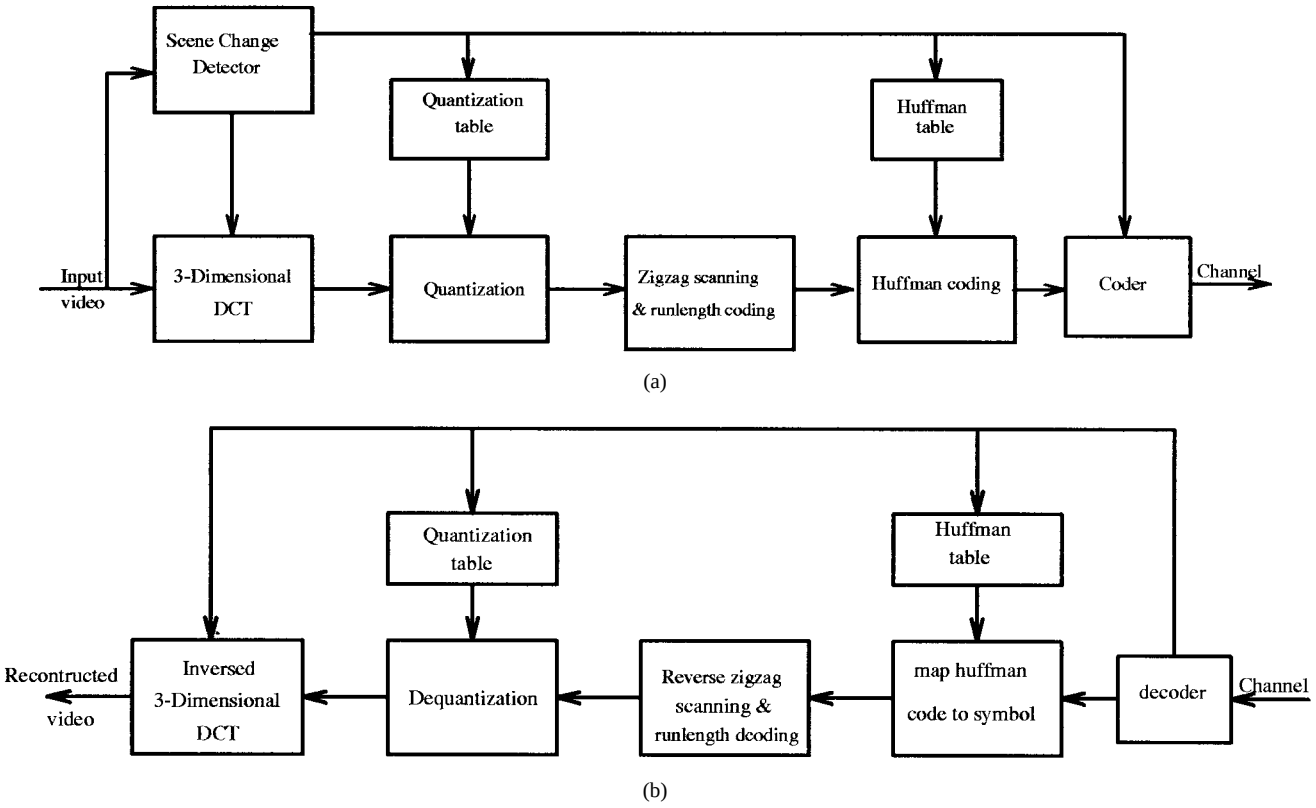

Fig. 2. Block diagram of our proposed variable temporal-length 3-D DCT (a) encoder and (b) decoder.

image sequence. The quality of the proposed variable temporal-length 3-D DCT coding is increased if a longer time window is used; however, the memory requirement is also unavoidably increased. Then, each time window is subdivided into a block sequence with a spatial dimension of $M \times N$ as depicted in Fig. 3(a). The pixel intensity in each block sequence can be mathematically represented as $I_t(x, y)$ with $x \in 0, 1, \cdots, M-1$, $y \in 0, 1, \cdots, N-1$, and $t \in 0, 1, \cdots, W-1$. To keep each 3-D block having only low motion activity, a scene change detector has to be used to identify the scene change of each block sequence independently, and this has to be done before the interframe transformation.

Now, let us define a set of P discontinuity planes, α_P, within a block sequence. The set α_P can then be written as a set with P indices such that $t_i : t_{-1}(=0) < t_0 < t_1 < \cdots < t_{P-1} < t_P(=W)$ as shown in Fig. 3(b). The scene change detector generates the subsequences, F_j, with temporal length $L_j = t_j - t_{j-1}$, where $j = 0, 1, \cdots, P$, and F_j is defined as

$$F_j = [I_{t_{j-1}}(x, y), I_{t_{j-1}+1}(x, y), \cdots, I_{t_j-1}(x, y)]. \quad (1)$$

An adaptive block filter (ABF) [9] can be developed for the minimization of the error function, which is defined as

$$E(\alpha_P) = \sum_{j=0}^{P} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} e[F_j(x, y)] \quad (2)$$

where $e[F_j(x, y)] = \sum_{i=0}^{L_j-1} [I_{t_{j-1}+i}(x, y) - \overline{I_j(x, y)}]^2$, and $\overline{I_j(x, y)}$ denotes the mean of $I_{t_{j-1}}(x, y)$, $I_{t_{j-1}+1}(x, y)$, $\cdots$, $I_{t_j-1}(x, y)$.

The global minimum of this error function, which relies on the possible optimal locations of the scene change planes, can be efficiently computed with dynamic programming [9]. The required number of scene change planes in each block sequence depends on its motion activity; therefore, it is reasonable to adaptively select the number of scene change planes, P, such that the minimum value of the error function, $\min_{\alpha_P} E(\alpha_P)$, could be less than a predefined threshold, E_0. In other words, the value of P in a block sequence increases when it contains many motions. On the other hand, smooth and minor temporal variations of the block sequence give small P. This method can obtain the optimal discontinuity planes of each block sequence, but a large demand of computational effort is unavoidable.

In order to resolve the problem of heavy computation, another fast scene change detector using the mean absolute difference (MAD) could also be used. The MAD of two successive 2-D blocks at $t+1$ and t, MAD_t, is defined as

$$\text{MAD}_t = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} |I_{t+1}(x, y) - I_t(x, y)|. \quad (3)$$

A block is called *scene change* if the value of MAD_t is greater than T_0, where T_0 is a predefined threshold. The MAD method is very simple. However, its accuracy is not as high as the former approach.

After P discontinuity planes have been detected in each block sequence, as shown in Fig. 3(b), the block sequence is then segmented into $P+1$ 3-D blocks with different temporal lengths. Each of the variable temporal-length 3-D block with the size $L_j \times M \times N$, as depicted in Fig. 3(c), is converted to the 3-D DCT domain with the following equation:

$$X(u, v, w) = \frac{8}{L_j \times M \times N} \sum_{t=0}^{L_j-1} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \cdot C(u)C(v)C(w)I_t(x, y) \cos \pi \frac{(2t+1)u}{2L_j} \cdot \cos \pi \frac{(2x+1)v}{2M} \cos \pi \frac{(2y+1)w}{2N} \quad (4)$$

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

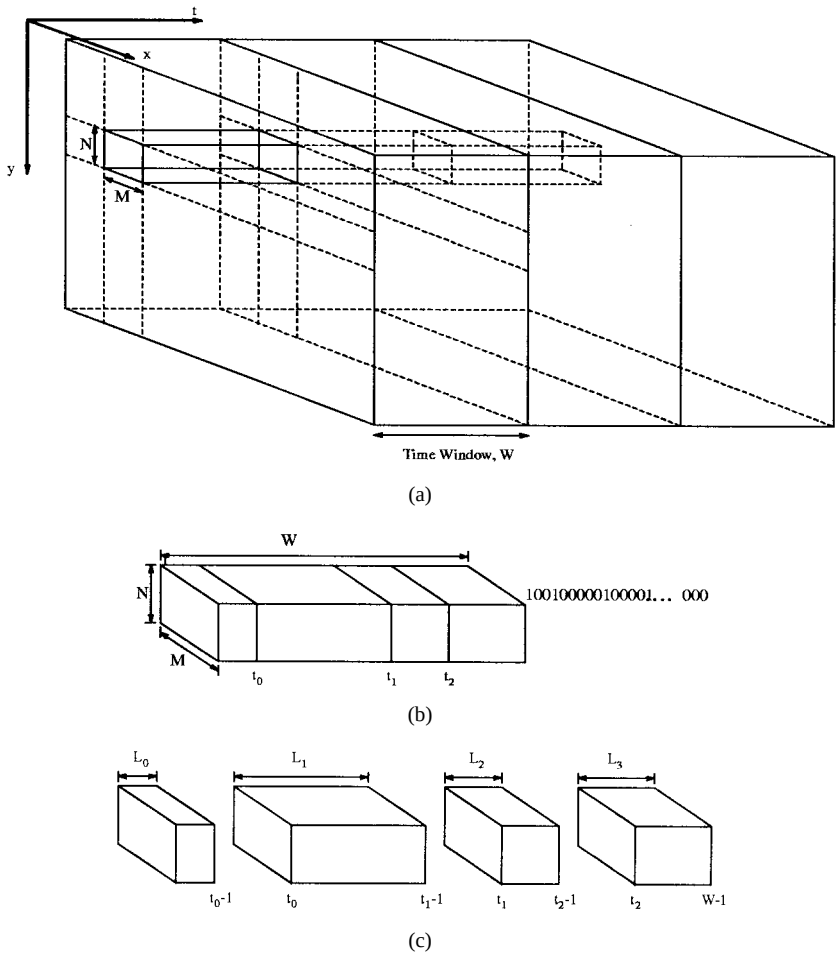

Fig. 3. Details of variable temporal-length 3-D DCT coding.

Then, each of the 3-D DCT coefficients is uniformly quantized to produce the quantized coefficients. Generally, it is well known that the dc coefficient of the DCT should be quantized accurately, otherwise the blocking effect may be noticeable. Thus, the dc coefficient is quantized using a 10-bit uniform quantizer. For the ac coefficients, the values of transformed coefficients are divided by the quantizer step size, and rounded to the nearest integer as follows:

$$\begin{aligned} &Q(u,\ v,\ w) \\ &\quad = \text{Integer Round}\left[\frac{4 \times X(u,\ v,\ w)}{\text{quantizer step size}}\right] \\ &\quad = \text{Integer Round}\left[\frac{4 \times X(u,\ v,\ w)}{(q+P+1) \times m_{L_j}(u,v,w)}\right] \end{aligned} \tag{5}$$

where $m_{L_j}(u,\ v,\ w)$ is the corresponding element of the quantization table with temporal length L_j and q is the quantizer scale.

In (5), the quantizer step size is derived from various quantization tables, the quantizer scale, and the number of 3-D blocks in the block sequence. Theoretically, we have W possible quantization tables for W different temporal-length 3-D blocks. Five different image sequences were used to design these different temporal-length quantization tables. For instance, we have obtained the quantization table for a specific temporal length k by applying k temporal-length 3-D DCT to all possible $k \times M \times N$ 3-D blocks within all image sequences. Then, the variance of each transformed coefficient is estimated. Also, the transformed coefficients are assumed to have a Gaussian distribution [10] and the quantized coefficients are encoded using the Huffman coding. Then the quantization table with temporal length k is defined which is based on their coefficient variances using the least square minimization technique [10]. It is seen that the quantizer step size depends on the number of 3-D blocks in the block sequence. This is because the number of 3-D blocks in the block sequence is proportional to the total number of bits required to code the block sequence. Our strategy is to assign quantizer step sizes by considering the number of 3-D blocks in the block sequence such that the bit rate in each block sequence remains nearly constant. Thus, coarse quantization is applied to block sequences with a large number of 3-D blocks for the given bit rate. On the other hand, a fine quantization is allowed to block sequences with only a small number of 3-D blocks. Therefore, the number and the location of the discontinuity planes are directly related to the bit rate and the quantization error in the block sequence. Also, the overall bit rate could be controlled by the quantizer scale which is to be sent to the decoder.

After quantization, the quantized coefficients in each variable temporal-length 3-D block are scanned in a zig-zag manner for the first frame, followed by the zig-zag scanning of the successive frames. Then, the run-length coding and the Huffman coding are employed to reduce the bit rate further.

The decoder requires information about the scene changes from each block sequence such that the appropriate temporal-length inverse 3-D DCT can be performed. This side information can be recorded with a bit sequence where the number of bits is equal to the number of image frames in the current time window as shown in Fig. 3(b). A

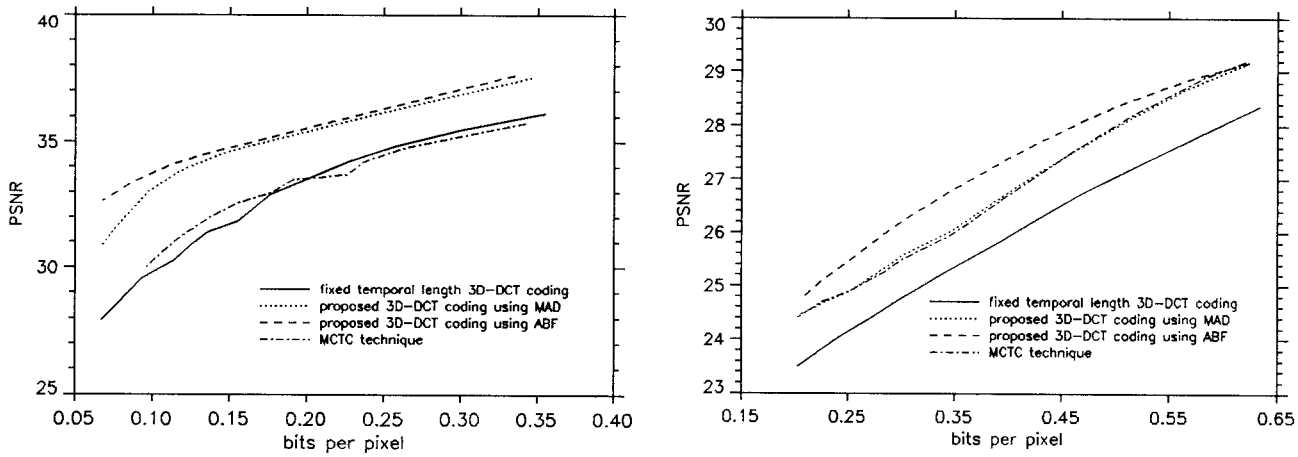

Fig. 4. Plot of PSNR against b/pixel for different algorithms on (a) Salesman and (b) Football sequences.

Fig. 5. Regions (black square ■) have inferior performance for our proposed algorithm using the ABF as compared with MCTC technique on (a) Salesman sequence at 0.1 b/pixel and (b) Football sequence at 0.25 b/pixel.

bit "1" indicates that this is the first frame in a new 3-D block. This bit sequence can be further compressed using the arithmetic coding [11]. Fig. 2(b) gives the block diagram for the decoding process.

III. Experimental Results and Discussions

In this section, results of computer simulations are provided to demonstrate the performance of our proposed transform coder on several well-known image sequences, including Salesman and Football, with 30 frames/s. The size of the frames in the Salesman and Football sequences are 352 × 288 and 352 × 240 pixels, respectively. Comparisons are to be given on the performance of the proposed transform coder and some conventional methods. For our approach, a time window of 32 frames has been employed. For the conventional methods, a fixed-length 3-D DCT coding with a 3-D block of size 8 × 8× 8 and the MCTC technique based on the skeleton of the MPEG standard [4] with *IPPPPPPP* group of pictures were used. The block size was 16 × 16 with the maximum displacement of 8 for full search motion estimation. For the measurement of image quality, the peak signal-to-noise ratios (PSNR's) were used.

Fig. 4 provides a comparison of the PSNR's among different algorithms and for various b/pixel on the Salesman and Football sequences. The proposed 3-D DCT coding technique provides a significant performance improvement over the fixed-length 3-D DCT coding and the MCTC technique on the Salesman sequence as shown in Fig. 4(a). It is seen that the proposed 3-D DCT coding technique using the MAD method as scene change detector has a 3–3.5 dB and 1–1.5 dB PSNR improvement as compared to that of the fixed-length 3-D DCT coding and the MCTC technique at low bit rate and high bit rate respectively. The PSNR performance can be further improved by using the ABF. As shown in Fig. 4(a), the proposed 3-D DCT coding using the ABF even achieves about 4–5 dB improvement as compared with the fixed-length 3-D DCT coding and the MCTC technique at low bit rate. This observation is expected since the Salesman sequence contains a person with small size who is moving quickly against a complicated background. In this type of complicated background, our proposed transform coder outperforms the MCTC technique in terms of both the PSNR and the subjective inspection, as depicted in Figs. 4(a) and 5(a), respectively. It is because our proposed transform coder can simultaneously reduce the interframe redundancy among a number of consecutive frames in this kind of complicated background, and a fine quantization is applied to a small number of transformed coefficients for a given bit rate. On the other hand, the degraded performance of the MCTC technique at a low bit rate is mainly due to poorly reconstructed previous frames, and thus the error has been propagated to consecutive frames. However, in the region of the moving person in the Salesman sequence, the scene change detectors produce a number of discontinuity planes in the block sequence. Then, a coarse quantization is required in order

(a)

(b)

(c)

(d)

(e)

Fig. 6. Reconstructed frame 16 of the sequence Salesman at 0.1 b/pixel. (a) Original. (b) Fixed temporal-length 3-D DCT coding. (c) Proposed 3-D DCT coding using ABF. (d) Proposed 3-D DCT coding using MAD. (e) MCTC technique.

to maintain the same bit rate. Therefore, the quantization error has been slightly increased. For each 3-D block, it contains only low motions, hence the temporal high-frequency coefficients seldom have significant values such that the 3-D DCT coding is still very efficient. Comparing with the MCTC technique, our approach is favorable even for frames containing moving persons, as shown in Fig. 5(a). Our proposed transform coder appears better due to the fact that the MCTC technique becomes inferior when rotational movements (the box in the Salesman sequence) and uncovered regions (the lower end of the tie in the Salesman sequence) are involved. Again, the poorly reconstructed previous frames at the low bit rate seriously affect the performance of motion compensation in the moving region. Thus, our proposed algorithm can obtain better performance as compared to the MCTC technique especially at the low bit rate. It is also obvious that the proposed transform coder with different scene change detectors performs better than that of the fixed-length 3-D DCT coding. As shown in Fig. 6(b), the fixed-length 3-D DCT coding suffers from serious "blocking" artifacts due to a large number of nonzero temporal high-frequency coefficients. The distortion introduced by the coarse quantization in the blocks with substantial motions is great at low

TABLE I
COMPARISON OF THE TIME USED OF OUR PROPOSED CODER AND THE MCTC TECHNIQUE

sequence	Proposed coder using ABF		Proposed coder using MAD		MCTC technique	
	encoding(sec)	decoding(sec)	encoding(sec)	decoding(sec)	encoding(sec)	decoding(sec)
salesman	920	331	343	321	980	120
football	2212	189	226	215	800	90

bit rates. In our proposed algorithm, the adaptive adjustment of the temporal-length 3-D DCT makes the 3-D blocks having high interframe pixels correlation. Thus, the algorithm can achieve high subjective quality at low bit rate, as depicted in Fig. 6(c) and (d).

On the other hand, the Football sequence contains complicated motions for all its image frames. Thus, bad performance for the MCTC technique at low bit rate is expected, mainly due to poorly reconstructed previous frames. Also, the fixed-length 3-D DCT coding techniques yields poor performance because the temporal high-frequency coefficients have significant values. However, our proposed transform coder using the ABF has 2-dB PSNR improvement as compared with that of the fixed-length 3-D .DCT coding for various bit rates, as shown in Fig. 4(b). Compared with the MCTC technique, our proposed coder using the ABF still has 1-dB PSNR improvement at low bit rate. However, as the bit rate increases, the MCTC technique no longer suffers from the poorly reconstructed previous frames. Thus, the MCTC technique becomes comparable to our proposed transform coder.

In order to analyze the computational complexity of the proposed transform coder, different algorithms are realized on a Sun Sparc10 workstation and the system call, *getrusage()*, has also been used to analyze the timing for operations, not including the system overheads. Table I compares the times used for our proposed transform coder and the MCTC technique. It shows that our transform coder using the MAD as scene change detector is much faster than that of the MCTC technique in encoding. However, in the case of using the ABF, the time required for encoding is comparable with the MCTC technique for the Salesman sequence and shows very time consuming for the Football sequence. It is due to the fact that the time requirement of the adaptive block filter depends on the number of discontinuity planes in the block sequence. Thus, the complexity of our proposed transform coder using the ABF is significantly increased in the sequence that contains many motions throughout all the image frames.

IV. CONCLUSIONS

A novel variable temporal-length 3-D DCT coding is proposed. This proposed 3-D DCT coding adaptively adjusts the temporal length of the 3-D block. The interframe correlation within each 3-D block is high such that the advantage of the 3-D DCT can be fully utilized. The adjustment of the temporal length of each 3-D block depends on the local activities in the image sequence. Two scene change detectors have been used to determine the local activity of the block sequence. The scene change detectors using the MAD method is simple and fast; however, it does not give the optimal discontinuity planes. The adaptive block filter method can obtain more accurate discontinuity planes, but a relatively high computational complexity is required. A series of computer simulations shows that the proposed variable temporal-length 3-D 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 3-D DCT and the motion compensation transform coding technique. These results support the subjective view and clearly verify the effectiveness of the proposed scheme. Results of our study also indicate that significant gain over the standard MCTC techniques used nowadays is possible through the 3-D DCT coding with variable temporal lengths. However, many problems still remain to be resolved, including the high memory requirement and the design of fast optimal scene change detector. Nevertheless, the 3-D frequency coding is a fruitful direction that should not be forgotten for the study on the techniques for video compression.

REFERENCES

[1] J. A. Roese and W. K. Pratt, "Interframe cosine transform image coding," *IEEE Trans. Commun.*, vol. 25, pp. 1329–1338, Nov. 1977.
[2] G. Karlsson and M. Vetterli, "Three dimensional subband coding of video," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 1988, pp. 1100–1103.
[3] C. I. Podilchuk, N. S. Jayant, and N. Farvardin, "Three-dimensional subband coding of video," *IEEE Trans. Image Processing*, vol. 4, pp. 125–139, Feb. 1995.
[4] D. J. LeGall, "MPEG: A video compression standard for multimedia applications," *Commun. ACM*, vol. 34, pp. 47–58, Apr. 1991.
[5] O. Chantelou and C. Remus, "Adaptive transform coding of HDTV picture," in *Int. Workshop Signal Processing in HDTV*, L'Aguila, Italy, Feb.–Mar. 1988, pp. 231–238.
[6] Y. L. Chan and W. C. Siu, "A new adaptive interframe transform coding using directional classification," in *Proc. IEEE Int. Conf. Image Processing*, Nov. 1994, no. 2, pp. 977–981.
[7] ——, "Fast interframe transform coding based on characteristic of transform coefficients and frame difference," in *Proc. IEEE Int. Symp. Circuits and Systems*, Apr. 1995, pp. 449–452.
[8] K. R. Rao and R. Yip, *Discrete Cosine Transform-Algorithms, Advantages and Applications.* New York: Academic, 1990.
[9] B. Olstad, "Noise reduction in ultrasound images using multiple linear regression in a temporal context," in *Proc. SPIE*, vol. 1451, pp. 269–281 1992.
[10] A. Segall, "Bit allocation and encoding for vector sources," *IEEE Trans. Inform. Theory*, vol. 22, pp. 162–169, Mar. 1976.
[11] P. G. Howard and J. S. Vitter, "Arithmetic coding for data compression," *Proc. IEEE*, vol. 82, pp. 857–865, June 1994.