A NEW ADAPTIVE INTERFRAME TRANSFORM CODING USING DIRECTIONAL CLASSIFICATION

Yui-Lam CHAN and Wan-Chi SIU

Department of Electronic Engineering
Hong Kong Polytechnic
Hung Hom, Kowloon, Hong Kong

ABSTRACT

Interframe transform coding is affected not only by the statistics of spatial details within a frame, but also by the variation of the amount of movement and other temporal activities in different regions of the image sequence. Therefore, adaptive techniques have to be used in order to achieve good image quality. In this paper, we propose a new version of the adaptive interframe coding method, namely directional classification, which is based on image sequence statistics. Blocks with different perceptual features such as edges and high motion activity are categorized to different classes. Then, a new adaptive quantization, associated with appropriate scanning and huffman coding, are employed based on the classification map. Coding tests using computer simulation show this technique is indeed very efficient.

1. INTRODUCTION

Intra/interframe 2-dimensional Discrete Cosine Transform (DCT) is most popularly used for digital video coding system. But, the intra/interframe 2D-DCT is not compatible with the "quick search mode" of current consumer VCR. Besides, periodical refreshing (periodical encoding a full frame of video using intra-mode) is needed to prevent from propagating errors occurring over the channel and initial turning on of the receiver. Whereas, interframe transform coding technique takes the advantage of frame-to-frame correlation as well as spatial image redundancies[1-3] without the drawbacks of intra/interframe 2-dimensional DCT. In [4], adaptive 3-dimensional DCT coding was reported to be comparable with intra/interframe 2-dimensional DCT in certain image sequences.

Among different 2-dimensional image data compression techniques, the adaptive transform has been generally recognized as an efficient means and has been considered to have good potentials. Chen and Smith [5] classified the blocks according to the summation of ac energy, four equally populated classes have been developed. Gimlett [6] proposed using the magnitude sum of the ac coefficient. These classification concepts can be extended to the three dimensional space.

Due to the image sequence property, it is not often that a 3D block will have simultaneously high spatial and temporal activities. Thus, the ac energy is not a good activity measure for coding purposes. But, in classified VQ[7-9], adaptive coding is introduced by classifying the block based on horizontal, vertical and diagonal energy. In this paper, based on the directional property, a newly selected set of frequency coefficients is used to indicate the spatial and temporal activity. Classification techniques may then be used to categorize each block according to the selected set of frequency coefficients. Then, bit assignment and quantization are employed as shown in figure 1. Based on the directional property, different classes have different coefficient distributions. Hence, a new adaptive scanning method is introduced to reduce the bit rate. At last, different huffman code tables which reflect the statistics of different classes are used for the sake of further compression. This adaptive coding technique utilizes the directional property of blocks and a significant improvement can be achieved as shown below.

This paper describes the coding algorithm and also presents the results obtained by our proposed adaptive interframe transform coding using directional classification. The contents of the following sections are as follows. Section 2 describes the three dimensional discrete cosine transform. The new proposed classification scheme is presented in section 3, while the new proposed scanning method is described in section 4. Experimental results obtained by computer simulations are discussed in section 5. Finally, a conclusion is given in section 6.

0-8186-6950-0/94 $4.00 © 1994 IEEE

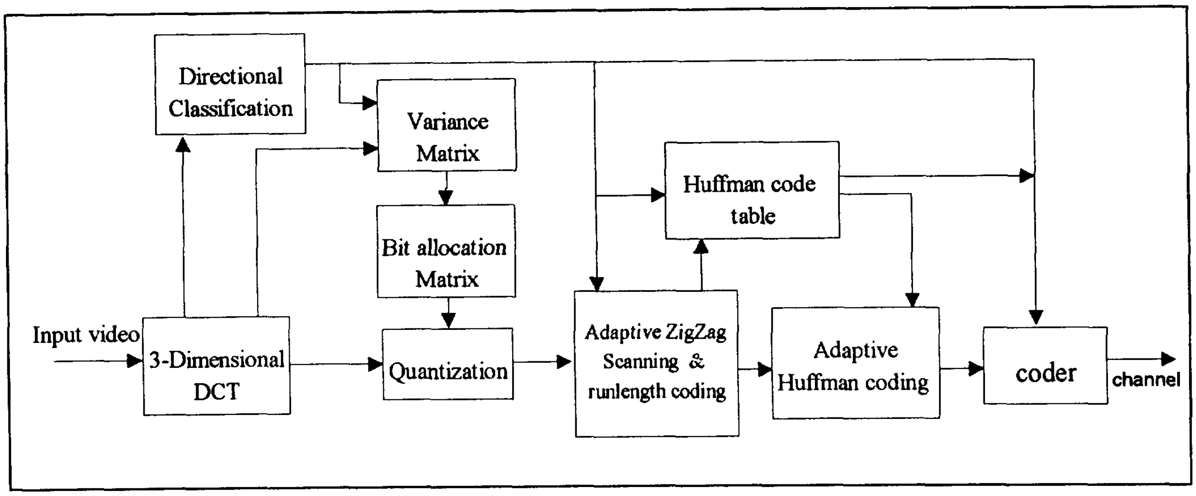


Figure 1: Block diagram of adaptive interframe transforms coding system

2. THREE DIMENSIONAL DISCRETE COSINE TRANSFORM

The two-dimensional discrete cosine transform is considered as an effective way to decorrelate still images in the transform domain. The following paragraphs present a summary of the extension of transform coding concepts to three dimensional space.

Since the picture is presented at high frame rate, there is considerable correlation between successive frames. Hence a three-dimensional discrete cosine transformation of a block is employed to reduce temporal correlation as well as spatial correlation.

Each image sequence is divided into blocks of 8x8x8 pixels and a three-dimensional discrete cosine transform (3D-DCT) is carried out on each block. The forward transform is defined as follows:

$$X(u,v,w) = \frac{1}{64}\sum_{k=0}^{7}\sum_{j=0}^{7}\sum_{i=0}^{7} C(u)C(v)C(w)x(k,j,i) \cos\pi\frac{(2k+1)u}{16}\cos\pi\frac{(2j+1)v}{16} \cos\pi\frac{(2i+1)w}{16}$$

where $x(k,j,i)$ is the pixel value in the image sequence for

$$C(n) = \begin{cases} \frac{1}{\sqrt{2}} & \text{for n=0,} \\ 1 & \text{otherwise} \end{cases}$$

3. DIRECTIONAL CLASSIFICATION

The transformed blocks are assigned to suitable classes according to their activity level. Within each class, bits are allocated to individual transform coefficients according to the variance matrix of the class [10]. The conventional classification is based on the summation of the ac energy(E), defined as follows:

$$E = \sum_{u=0}^{7}\sum_{v=0}^{7}\sum_{w=0}^{7}[X(u,v,w)]^2 - [X(0,0,0)]^2$$

Based on the summation of the ac energy(E), four equally populated classes have been developed. But large distortion is observed in high activity blocks(i.e. blocks with large values of E). It is due to this fact that the ac energy cannot discriminate whether the block contains high motion activity, detail vertical edge or detail horizontal edge in high activity block, even it shows that the block contains high frequency coefficients.

In the following scheme, we define some new parameters which can reflect the directionality of the block such as high motion activity, detail vertical edge and horizontal edge. A newly selected set of transform coefficients should be used to identify the type of the blocks. Let us define, as shown below, C_v, C_h and C_t to reflect the extent of the horizontal, vertical and temporal energy respectively. And also, eight classes are defined as shown in table 1 whereas T_v, T_h, T_t are the

Table 1: Directional Classification Table

Class	$C_h > T_h$	$C_v > T_v$	$C_t > T_t$
0			
1	$\surd$		
2		$\surd$	
3			$\surd$
4	$\surd$	$\surd$	
5	$\surd$		$\surd$
6		$\surd$	$\surd$
7	$\surd$	$\surd$	$\surd$

threshold energy of C_v, C_h and C_t respectively.

$$C_v = \sum_{w=0}^{7}[X(0,0,w)]^2 - [X(0,0,0)]^2$$

$$C_h = \sum_{v=0}^{7}[X(0,v,0)]^2 - [X(0,0,0)]^2$$

$$C_v = \sum_{u=0}^{7}[X(u,0,0)]^2 - [X(0,0,0)]^2$$

Class 0 is the low activity block which is in the still background. Class 1, 2 and 3 are the blocks containing detail vertical edge, horizontal edge and high motion activity respectively. Class 4 represents high activity in spatial domain. Class 5 reflects the block which has both vertical edge and high motion activity. Class 6 reflects horizontal edge and high motion activity in same block. Class 7 is the block which has simultaneously high spatial details and high temporal activity.

4. ADAPTIVE SCANNING AND HUFFMAN CODING

Since the distributions of the quantized coefficients are different in different classes. For example, in class 1 and 2, most coefficients are concentrated in the first frame, whereas, in class 3, most coefficients are concentrated in the temporal direction. For efficient use of the run length coding, adaptive scanning is used [4]. The 3-D block is realized by coding each block with the two scanning stages and the scanning chosen is the one with the lowest bit rate. The first spatial, second temporal and first temporal , second spatial scanning are described as follows:

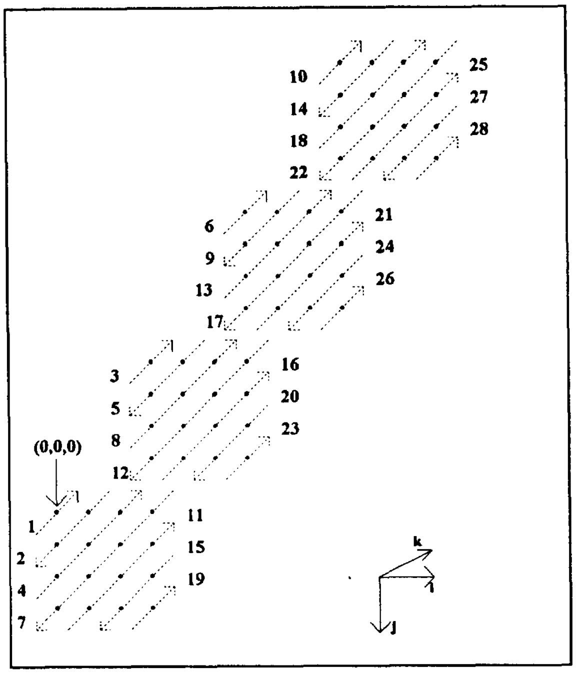


Figure 2: New scanning sequence

1. *First spatial, second temporal* - the zig-zag scanning of the first frame, followed by the zig-zag scanning of the successive frames. This scanning method is suitably used in still area block. However, in video coding without classification, this scanning method is more suitable.
2. *First temporal, second spatial* - scanning the corresponding coefficients of successive frames from low to high frequencies. For spatial scanning, it follows the zig-zag scanning. This scanning method is suitably used in moving area block.

In our proposed coding scheme, the sequences of operations:first spatial, second temporal and first temporal, second spatial, are suitably used in classes 0, 1, 2, 4 and 3, 5, 6 respectively. So, the scanning method is based on the classification map without any extra computation. But, in class 7, since it contains both horizontal, vertical and temporal energy, the above two scanning methods are not suitable. A new scanning sequence is proposed as shown in figure 2. The number in figure 2 is the scanning order of quantized coefficients. For simplicity of illustration, a 4x4x4 block is used. The scanning sequence is:

(0,0,0)⟶(0,0,1)⟶(0,1,0)
⟶(1,0,0)⟶(0,2,0)⟶(0,1,1)
⟶(0,0,2)⟶(1,0,1)⟶(1,1,0)
⟶(2,0,0)⟶(0,0,3)⟶(0,1,2)
⟶(0,2,3)⟶(0,3,0)⟶(1,2,0)
⟶(1,1,1)⟶(1,0,2)⟶(2,0,1)
⟶(2,1,0)⟶(3,0,0)⟶(0,3,1)
⟶(0,2,0)⟶(0,1,3)⟶(1,0,3)
⟶(1,1,2)⟶(1,2,3)⟶ ......

According to the data in the scanning sequence, two kinds of symbols can be generated for the sake of further compression. These two symbols are coded in zero-runlength method. Based on these symbols in different classes, a different huffman code table is generated. Each symbol is then replaced by its corresponding code.

5. RESULT

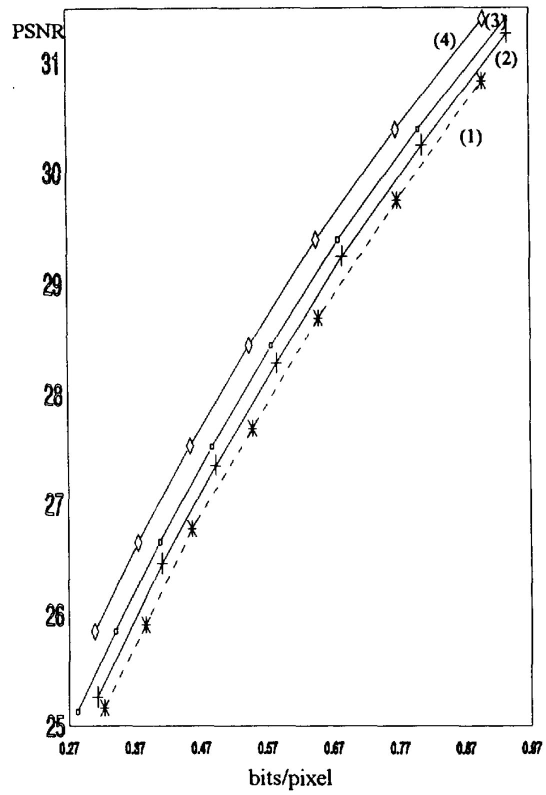
A series of computer simulations have been conducted to evaluate the performance of the adaptive interframe transform coding. These include the 32 frames image sequence "football". The sequence contains translation, zooming-out and both slow and fast paning. A peak signal-to-noise ratio(PSNR) measure of image fidelity is used to quantitatively evaluate the performance of the interframe coding systems. The PSNR of an original image $x(k,j,i)$ and its reconstruction image $x'(k,j,i)$ over spatial coordinates j,i and frame coordinate k is defined as :

$$PSNR = 10 \log \frac{255^2}{MSE}$$

where

$$MSE = \frac{1}{512} \sum_{k=0}^{7} \sum_{j=0}^{7} \sum_{i=0}^{7} [x(k,j,i) - x'(k,j,i)]^2$$

Figure 3 provides a comparison on peak signal-to-noise ratio(PSNR) between the 3D-DCT without classification and different adaptive 3D-DCT algorithms. The PSNR performance of adaptive coding based on [5] is only slightly better than that of 3D-DCT without classification (about 0.19dB). This is due to the fact that this classification only bases on the ac energy, but no consideration of the directional energy of the block. To resolve this problem, C_h, C_v and C_t have been previously defined to identify the directional property. Graph (3) makes use of this property and show that it can achieve the improvement on the PSNR (about 0.2dB over graph (2)). Another advantage of the new classification scheme is that it can achieve the improvement in compression ratio with adaptive scanning, but



(1) - 3D-DCT without classification
(2) - classification based on [5]
(3) - proposed classification with adaptive quantization
(4) - proposed classification with adaptive quantization, scanning and huffman coding

Figure 3: Plot of PSNR against bits/pixel

no overhead is required. Based on the classification map, when $C_t > T_t$, (class 3, 5, 6) we can use first temporal, second spatial scanning method and in class 7, the new scanning method(see Fig. 2) is employed. In classification based on the ac energy, only the magnitude of ac energy is calculated, it has no advantage for using adaptive scanning in high activity class. This is due to the fact that the large ac energy blocks do not know the actual distribution of coefficients. Graph (4) shows the performance of classification with adaptive quantization, scanning and huffman coding. The PSNR performance of this adaptive coding is 0.65dB better than that of 3D-DCT without classification, and 0.45dB than that of classification based on the ac energy.

6. CONCLUSION

A new classification scheme for adaptive 3D-DCT coding which takes the advantage of directionality for different styles of blocks has been described. This classification scheme is based on the horizontal, vertical and temporal energy. Adaptive quantization, scanning and huffman coding method is used based on the classification map.

A conventional adaptive coding for image can extend into three dimensional space for video coding. However, the bit allocation map for high activity cannot correctly reflect the energy distribution of horizontal, vertical or temporal direction. So, the PSNR performance of the new classification scheme is better than that of the conventional classification scheme with only a little more overhead information ($\frac{1}{512}$ bits/pixel). Besides, the computation can also save for calculating C_h, C_v and C_t comparing with calculating the total ac energy.

In summary, the use of the new simple classification minimizes the inner class differences and improves the objective image quality. Obviously, the classification is simple and matched with the more important block characteristics.

7. REFERENCES

[1] J.A.Roese and W.K.Pratt, "Theoretical performance methods for interframe transform and hybrid transform/DPCM coders", Proceedings SPIE, Vol-87, August 1976,pp.172-179.

[2] J.A.Roese and W.K.Pratt, "Interframe cosine Transform image codin", IEEE Transactions on communications, Vol-25, November 1977, pp.1329-1338.

[3] T.RAJNATARAJAN, "On interframe transform coding", IEEE Transactions on communications, Vol-25, November 1977, pp.1329-1338.

[4] Olivier Chantelou, Christian Remus, "Adaptive transform coding of HDTV picture", International workshop on signal processing of HDTV, L'Aguila, Italy, Elsevier Science Publishers, North Holland, pp.231-238, Feb-March,1988.

[5] W.H. Chen and C.H.Smith, "Adaptive coding of monochrome and color images " , IEEE transactions on communications, Vol-25, November,1977,pp.1285-1292.

[6] J.I.Gimlett, "Use of 'activity' classes in adaptive transform image coding", IEEE Transactions on communications, Vol-23, July 1975, pp.785-786.

[7] B.RAMAMURTHI and A.Gersho, "Classified Vector quantization of images", IEEE Transactions on communications, Vol-34, November 1986, pp.1105-1115.

[8] K. Aizawa, H.Harashima and H. Miyakawa, "Adaptive discrete cosine transform coding with vector quantization for color images", ICASSP 86, intl. Conf. on Acoust. Speech, Signal process, April,1986,pp. 985-988

[9] H. Holzlwimmer, A. v. Brandt, and W. Tengler, "A 64 Kbit/s motion compensated transform coder using vector quantization with scene adaptive codebook", ICC 87, Intl. Conf. on communications, June 1987,pp.151-156.

[10] A.Segall,"Bit allocation and encoding for vector sources", IEEE transactions on information theory, Vol-22, Mar,1976, pp.162-169.