

NEW PIXEL-DCT DOMAIN CODING TECHNIQUE FOR OBJECT BASED AND FRAME BASED PREDICTION ERROR

Ko-Cheung Hui and Wan-Chi Siu

Centre for Multimedia Signal Processing
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Tel: (852) 27666229 Fax: (852) 23624741 email: enwcsiu@polyu.edu.hk

ABSTRACT

Discrete Cosine Transform (DCT) is widely used in modern video compression standards, such as the ITU-T H.263 and the ISO MPEG-4, to achieve high compression efficiency. A major merit of the DCT is its capability in high energy compaction for natural images. However, the motion prediction error frame is not a natural image but synthetically generated by the process of motion compensation. This process degrades the energy compaction efficiency of the DCT. We study the spatial distribution of the prediction errors resulting from either the full-search motion estimation or other fast search algorithms in order to improve the coding efficiency of the DCT. Subsequently, a Mixed Spatial-DCT-based Coding Scheme is proposed for coding the prediction errors in this paper. Our experimental results show that this coding scheme can successfully improve the compression performance of the traditional DCT-based video coder with block based motion compensation for arbitrary shaped video objects and, video sequences which contain moderate to high motion activities.

1. INTRODUCTION

In the past decade, various waveform coding techniques such as Discrete Cosine Transform (DCT), subband/wavelet and vector quantization have also been developed for the applications of video coding. It is known that DCT based coding is still popular and researchers are in favorite of it for video coding [1]. However, using DCT based coding to perform compression, the motion compensated prediction error is far from optimal. The statistical properties of the error are different from that of natural images. Since, the redundancies of the error which is synthetically generated by the process of motion

compensation cannot be exploited successfully by the DCT [2,3]. Moreover, the BMC assumes that the motion between successive frames is purely translational. This is not true for most of the real world video sequences. The prediction error is generally concentrated in a clustered portion of the image even though full search algorithm (FSA) is employed. It leads to a scattering of the DCT coefficients and makes the compression inefficient.

To resolve this inefficiency, in this paper, we introduce a new Mixed Spatial-DCT-based Coding technique (MSDC). The MSDC divides a prediction error MB into two components. Each component is characterized by its own spatial correlation. One component is then coded by using the binary bit plane coding and variable length coding techniques (VLC), and the second component is coded by using the traditional DCT-based method. Results of our experimental work show that the MSDC can achieve a better compression efficiency for the prediction error when compared with the traditional DCT-based coding technique.

2. CHARACTERISTICS OF THE MOTION COMPENSATED PREDICTION ERROR

Making use of motion compensation, modern video coding standards usually can attain efficient video compression. In most cases, it is also assumed that the DCT-based coding is applicable to process the prediction error after proper motion estimation in the encoder side. Examples of used motion estimation algorithms are partial distortion search (PDS) techniques [4], Adaptive Pixel Decimation for Block Motion Vector Estimation [5] and the Motion Vector Field Adaptive Fast Motion Estimation (MVFAST) [6]. However, as noted before, a possible failure of BMC comes from the fact that there is a special distribution of the prediction error. It destroys the energy compaction ability of the DCT.

MPEG-4 supports coding of arbitrary shaped objects with a set of tools. The MBs, which only partially filled with opaque pixel values, are called partial MBs in MPEG-4. MPEG-4 makes use of a repetitive padding technique [7] to pad a reference arbitrarily shaped VOP, such that block-based motion estimation/compensation of the partial MBs can be processed efficiently. This technique can effectively compensate the motion of video objects with low motion activity. However, for objects with moderate to high motion activities, it is unreliable to use padding to predict the pixel values of a partial MB. Moreover, inaccurate video object segmentation may also make the padding inappropriate. A segmented video object plane which contains part of a moving object and part of a still background is generally the result of inaccurate segmentation. Hence, replicating the wrong edge pixel values in the repetitive padding cannot compensate the motion of video objects efficiently.

The clustering effect of prediction errors occurs not only in partial MBs. We found that large clustered prediction errors also occur frequently in sequences with high motion activity, even though full search algorithm is employed. This undesirable result is due to poor prediction of BMC of irregular motion activities.

3. PROPOSED ALGORITHM

The clustered error decreases the efficiency of DCT based compression. Fortunately, errors join together in clusters. This fact reflects that some spatial redundancies remain in this prediction error.

3.1. Separation of the prediction error MB into two components

Our algorithm starts by separating a prediction error into two components. Each component is then characterized by its own spatial correlation. Hence, we can apply different compression techniques to these two components based on their spatial characteristics. According to our discussion in Section 2, we know that large prediction errors tend to cluster together. This property motivated us to separate large prediction errors from an error block. When the magnitudes of the remaining error block are small and correlated, the traditional DCT coding can be applied directly. Moreover, the clustering property of error components enables us to treat them as an arbitrary shaped object plane. We suggest using the context-based arithmetic encoding (CAE) [8], which is a technique included in the MPEG-4 for binary alpha plane coding, to code separated error components.

A MB consists of four 8×8 pixels blocks. To separate the prediction errors, $E_i(x,y)$ of each block into two components, $Ec_i(x,y)$ and $Ed_i(x,y)$, we use a simple

approach which involves a thresholding accompanied with some modulus operations. Let us define,

$$E_i(x,y) = Ec_i(x,y) + Ed_i(x,y), \quad i = 0 \dots 3 \quad (1)$$

$$\text{where } Ed_i(x,y) = \begin{cases} E_i(x,y) & \text{if } |E_i(x,y)| < TS \\ \langle E_i(x,y) \rangle_{TS} & \text{otherwise} \end{cases} \quad (2)$$

$$Ec_i(x,y) = E_i(x,y) - Ed_i(x,y)$$

TS is a pre-defined threshold to detect peak errors and acts as a scaling factor for $Ec_i(x,y)$.

$\langle \cdot \rangle$ denotes modulus operator

It is obvious that the absolute values of error components in $Ed_i(x,y)$ are smaller than the pre-defined threshold value, TS . This separation cuts down abrupt peak errors from $E_i(x,y)$, and thus the traditional DCT becomes more efficient to code the resulting $Ed_i(x,y)$. We can then separate coding of the second error component, $Ec_i(x,y)$ into two parts, namely (i) shape and position of clustered errors, and (ii) magnitudes of the errors. The values of $Ec_i(x,y)$ are multiples of the threshold, TS , because of the modulus operation in eqn. (2). Hence, we only need to code $Ec'_i(x,y)$, that is the quotients of $Ec_i(x,y)$ divided by TS .

$$\text{or } Ec'_i(x,y) = Ec_i(x,y) / TS \quad (3)$$

3.2 Compression of the clustering regions with context-based arithmetic encoding (CAE) and variable length coding (VLC)

We use the intra mode of CAE [8] to code the shape of the separated error component, $Ec_i(x,y)$, in a MB, which is adopted in MPEG-4 for binary alpha plane coding. A block with 16×16 pixels, which contains the shape of four $Ec_i(x,y)$ blocks, is called a binary alpha block (BAB) in CAE. A template of ten pixels is used to define a context number, $C = \sum_{k=0}^9 c_k \times 2^k$. The context number is used to

index a probability table provided in MPEG-4 for a current pixel. Then, we use the indexed probability to drive an arithmetic encoder. When encoding a BAB, a border of width equal to 2 is extended from the current BAB for context number construction. The pixel positions in the extended border are padded with zero in our algorithm. The magnitudes of errors in $Ec_i(x,y)$ are firstly divided by TS . We encode the quotients with variable length codes, and raster scan is used to scan the MB from left to right and top to bottom.

3.3. Mode determination between Mixed Spatial-DCT-Based Coding and traditional DCT-based coding

The proposed Mixed Spatial-DCT-Based Coding technique is a way to exploit the redundancies remained in the error MBs. However, the traditional DCT-based coding is still useful for some MBs. In other words, the MSDC provides an additional mode to code the motion compensated prediction error for each MB. We add a mode bit to distinguish between the traditional DCT mode and the MSDC mode in the resulting bitstream.

In order to exploit the benefits from these two modes, one simple method is to test exhaustively the efficiency of the traditional DCT coding and that of the MSDC for each MB. It is obvious that we do not need to test a MB, when all pixel errors in the MB are smaller than the threshold, TS .

3.4. Mixed Spatial-DCT-Based Coding Scheme (MSDCS)

The procedure of the proposed Mixed Spatial-DCT-Based Coding Scheme (MSDCS) is summarized as shown below.

1. Code the input motion compensated prediction error, $E_i(x,y)$ in a MB with traditional DCT-based coding, and count the required number of bits as B_{DCT} .
2. Calculate $Ed_i(x,y)$ from $E_i(x,y)$ according to eqn. (2). Set threshold value TS to 16 in our simulation which has been determined experimentally.
3. Calculate $Ec_i(x,y)$ using eqn. (3).
4. Create the shape of clustered errors with the four $Ec_i(x,y)$ s in the MB.
5. Code the shape of the clustered errors with CAE, and count the required number of bits as B_{CAE} .
6. Code the magnitudes of $Ec_i(x,y)$ s with VLC, and count the required number of bits for the MB as B_{VLC} .
7. Code the $Ed_i(x,y)$ with the DCT-based coding, and count the required number of bits as B_{m-DCT} .
8. Calculate the total required number of bits for the MSDC, B_{MSDC} , where $B_{MSDC} = B_{CAE} + B_{VLC} + B_{m-DCT}$.
9. If $B_{MSDC} \geq B_{DCT}$, go to step 11.
10. Set the mode bit to represent MSDC mode, including the mode bit, and use the results of step 5, 6 and 7 to form the resulting bitstream.
11. Set the mode bit to represent the DCT mode, and make use of the result of step 1 for bitstream formation.

4. EXPERIMENTS

In order to evaluate the coding efficiency of the MSDCS, we used a video sequence and a video object for the evaluation. Sequences "Table Tennis", and "Segmented Stefan", were used to test coding performance for frame based and arbitrary shaped video objects coding.

The optimization of a coding system is essentially a multi-dimensional problem. The key issues concerned in

this problem are: bit-rate, quality (PSNR), speed-up (or computational gain), algorithmic complexity, memory size and bandwidth. There is always a trade-off among all these five key factors. It is the reason that fast motion estimation algorithms have attracted a lot of attention in the past decade. In order to evaluate the compatibility of the MSDCS, we have compared the coding results of our proposed scheme and the traditional DCT-based prediction results of the PDS which is a fast full search algorithm and the MVFAST which is a lossy fast algorithm.

We compared the rate-distortion performance of the DCT-based coding to that of the MSDCS by varying the quantization parameter, Qp . The results show that the MSDCS can outperform the traditional DCT-based coding especially in the high video quality region. This advantage vanishes gradually with increasing value of Qp . The reason is that the bitstream of MSDC consists of three components, bits of CAE, VLC and quantized DCT coefficients. Only the last component is a function of Qp . The bits required for coding CAE and VLC remain fairly constant; hence for low bitrate situation they dominate the bit rate requirement. However, experimental result shown in Figure 1 still confirm that for the tested video object, the MSDCS can achieve better rate-distortion performance comparing to the traditional method for Qp ranging from 1 to 6. On average, with the same PSNR, the MSDCS saves about 12% of the bitrate. The improvement given by the MSDCS in the tested sequence is not as significant as the situation in the tested video objects. The corresponding bitrate saving is about 6% on average.

Figure 2 show that the MSDCS can still obtain better rate-distortion performance in high video quality situation, when Qp ranges from 1 to 3. From Figure 2, sequence "Table Tennis" gives the significant improvement when comparing to the traditional DCT-based coding. The "Table Tennis" contains high motion activity and complex background with rapid camera motion. These factors are the causes of clustered prediction error and thus the DCT-based coding for these error MBs can be improved by the MSDC. From our simulation results, it is verified that the MSDC successfully improves the coding efficiency of MBs with clustered prediction errors. Hence, the proposed MSDCS is very suitable for coding arbitrary shaped objects or sequences with complex motion activity in high video quality application.

Moreover, it is worth to point out that the MSDCS accompanied with the MVFAST provides similar, even much better rate-distortion performance when comparing to the results of PDS with traditional DCT-based coding in all of our experiments. In addition, the required computational time is reduced significantly. We have found that about 80% of computational time and 12% of bitrate saving are achievable for the tested video objects on average.

5. CONCLUSIONS

In this paper, we have proposed a new algorithm, the Mixed Spatial-DCT-Based Coding Scheme (MSDCS), for coding motion compensated prediction errors. The MSDCS makes use of the phenomenon that pixel matching errors in some MBs tend to appear together in a cluster form. The reasons of this phenomenon include inefficient block based motion estimation, inaccurate segmented video objects, results of repetitive padding in MPEG-4, and high and complex motion activities. The algorithm exploits the redundancies remained and caused by these phenomena in an error MB. Our experimental results show that the MSDCS successfully improves the coding efficiency of the traditional DCT-based coding for MBs with clustered prediction errors. On average, with reasonable additional computation, the MSDCS saves up to 12% of the bit rates comparing to the results of the DCT-based coding when testing video objects and sequences with high video quality are coded with the same PSNR. The average required time for additional computation is about 1.9 to 2.4 times of the execution time for MVFAST. When comparing to the fast full search algorithm, using the PDS, only 13% to 19% of the execution time of the PDS is required. Hence, the MSDCS can save much of the computational time comparing to the traditional system if the optimal PSNR performance is used. The major advantages of the MSDCS include conceptual simplicity, and well defined CAE and VLC technique in the modern video codec. This also allows an easy way to improve an existing video codec by embedding the proposed MSDCS into it.

6. ACKNOWLEDGMENTS

The work described in this paper (or the equipment / facility) was supported by the Centre for Multimedia Signal Processing, Hong Kong Polytechnic University. K.C. Hui acknowledges the research grant G-V867 provided by the University and the Research Grants Council of the Hong Kong Special Administrative Region, China.

7. REFERENCES

- [1] Zixiang Xiong, Ramchandran K., Orchard M.T., Ya-Qin Zhang, "A comparative study of DCT and wavelet based coding", Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '98), Vol. 4, pp. 273 -276, May 1998.
- [2] de Faria S.M.M., Ghanbari M., "Low bit-rate video coding with spatio-temporal geometric transforms ", IEE Proceedings-Vision, Image and Signal Processing, Vol. 143, No. 3, pp. 164-170, June 1996.
- [3] Seung Chul Yoon, Ratakonda K., Ahuja N., "Low bit-rate video coding with implicit multiscale segmentation", IEEE

Transactions on Circuits and Systems for Video Technology, Vol. 9, No. 7, pp. 1115-1129, Oct. 1999.

[4] S. Eckart and C. Fogg, ISO/IEC MPEG-2 software video codec, Proc. SPIE 2419, 1995, pp. 100-118.

[5] Y. L. Chan and W. C. Siu, "New Adaptive Pixel Decimation for Block Motion Vector Estimation", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 6, No. 1, pp. 113-118, February 1996.

[6] Prabhudev Irappa Hosur and Kai-Kuang Ma, "Motion Vector Field Adaptive Fast Motion Estimation", Second International Conference on Information, Communications and Signal Processing (ICICS '99), Singapore, pp. 7-10, Dec. 1999.

[7] Wei-Ge Chen, Chuang Gu, Ming-Chieh Lee, "Repetitive and morphological padding for object-based video coding", Proceedings of International Conference on Image Processing (ICIP'97), Vol. 1, pp. 373-376, Oct. 1997.

[8] Ostermann J., "Efficient encoding of binary shapes using MPEG-4", Proceedings of International Conference on Image Processing (ICIP '98), pp. 295-298, Vol.1, 1998.

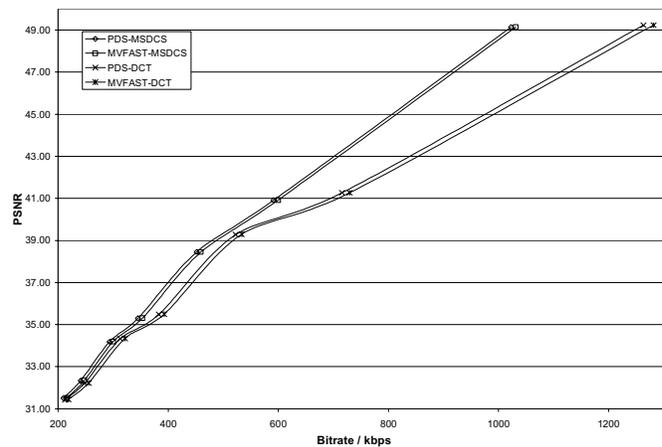


Figure 1. Coding performance for the video object "Segmented Stefan" with different quantization parameter, Qp ranged from 1 to 7.

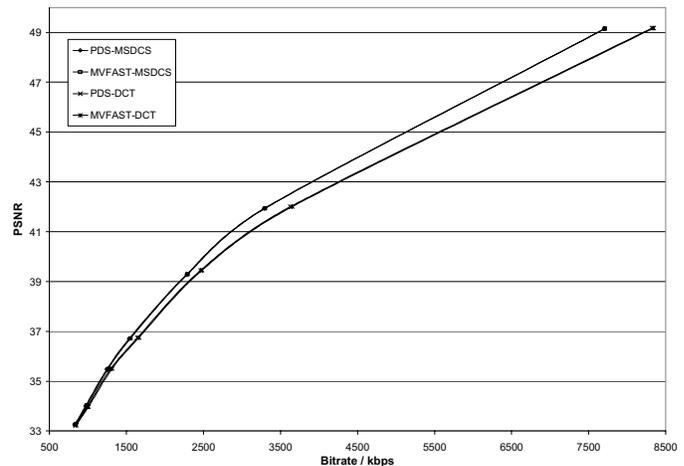


Figure 2. Coding performance for the sequence "Table Tennis" with different quantization parameter, Qp ranged from 1 to 7.