International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.7, No.4 (2014), pp.13-22
http://dx.doi.org/10.14257/ijsip.2014.7.4.02

All Phase Biorthogonal Transform and Its Application in MPEG-4 Video Compression

Xiaoyan Wang[1], Baochen Jiang[1,2], Chengyou Wang[1,*], Zhiqiang Yang[1,3] and Chunxiao Zhang[1]

[1] School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264209, P. R. China
[2] School of Electronic Information Engineering, Tianjin University, Tianjin 300072, P. R. China
[3] Integrated Electronic Systems Lab Co. Ltd., Jinan 250100, P. R. China
swwxy00800313@163.com, jbc@sdu.edu.cn, wangchengyou@sdu.edu.cn, yangzhiqiang@ieslab.cn, sdwhzcx@126.com

Abstract

In this paper, we propose an efficient algorithm for encoding of MPEG-4 simple profile. In the proposed algorithm, I frame encoding adopts all phase biorthogonal transform (APBT); for P frames, each frame puts the reconstructed frame of previous coded one as the reference frame to perform motion estimation (ME) and motion compensation (MC), and then the encoding of residual frame uses discrete cosine transform (DCT). While the proposed algorithm applying on MPEG-4, the objective effect and visual quality of reconstructed video sequence are approximate to the one based on DCT at the same bit rates. But the advantage is that the APBT transform coefficients can be quantized uniformly. Therefore the computing time is shorted and hardware implementation becomes easier. The proposed algorithm uses low computational complexity to make it applicable to encoding of MPEG-4 simple profile.

Keywords: *Video Compression, MPEG-4, All Phase Biorthogonal Transform (APBT), Motion Estimation (ME), Motion Compensation (MC), Discrete Cosine Transform (DCT)*

1. Introduction

Discrete cosine transform (DCT) has been widely used in many image and video compression international standards, including JPEG [1], H.263 [2], and MPEG-4 [3-4]. DCT transform shows good characteristics in terms of energy compaction and coefficients decorrelation [5]. However, such techniques produce noticeable blocking artifacts along block boundaries and ringing artifacts near edges in the decompressed images especially at low bit rates. For video sequence, I frame encoding uses DCT in MPEG-4 simple profile. The reconstructed I frame has blocking artifacts and ringing artifacts along block boundaries at low bit rates. Each P frame sets the reconstructed frame of previous coded one as the reference frame to perform motion estimation (ME) and motion compensation (MC). Thereafter the residual frame proceeds to code. Therefore the blocking artifacts of previous frame are propagated to current frame and can exist in any position. Therefore, many kinds of post-filtering schemes for deblocking and deringing artifacts are proposed, which increases complexity of the process of decoding. For example, Kwon, *et al.,* proposed an efficient deblocking algorithm that used block boundary characteristics and adaptive filtering in spatial domain [6]. And in DCT-based image and video compression algorithm, the quantization Table is designed according to the visual characteristics of human eyes [7]. But adopting

ISSN: 2005-4254 IJSIP

different quantization steps for different DCT coefficients makes quantization Table complex. It will increase many multiplication operations when adjusting encoding bit rates.

All phase biorthogonal transform (APBT) which is based on Walsh-Hadamard transform (WHT), DCT and inverse discrete cosine transform (IDCT) proposed by Hou, *et al.*, is a new transform for image compression instead of DCT [7]. APBT coefficients can be quantized uniformly because of the high-frequency attenuation characteristics. Moreover, the image compression algorithm based on APBT achieves the same performance with that based on DCT especially at low bit rates both in objective and subjective effects. Based on the better performance of APBT in terms of image compression, this paper proposes an efficient algorithm for encoding of MPEG-4 simple profile. In the proposed algorithm, I frame encoding adopts all phase biorthogonal transform and uniform quantization operation; each P frame sets the reconstructed frame of previous coded one as the reference frame to perform motion estimation and motion compensation, and then the encoding of residual frame uses discrete cosine transform. Experimental results show that the objective effect and visual quality of reconstructed video sequence based on the proposed algorithm are approximate to that based on DCT at the same bit rates.

The rest of this paper is organized as follows. Section 2 introduces DCT and APBT. In Section 3, the encoding scheme of I frame based on APBT is explained. And then Section 4 describes the encoding algorithm of P frame. The diamond search algorithm (DSA) included in the process of motion estimation is mainly introduced. Experimental results and comparisons between the proposed algorithm and the one based on DCT are given in Section 5. Conclusion of the paper and discussion for future work are given finally in Section 6.

2. DCT and APBT

2.1. Discrete Cosine Transform (DCT)

The conventional two-dimensional DCT transform is made usually by two one-dimensional DCT transform on row and column directions separately [1]. Let $\boldsymbol{X}$ and $\boldsymbol{C}$ represent an image block and DCT matrix with size of $N \times N$ respectively. After two-dimensional DCT transform, transform coefficients block $\boldsymbol{Y}$ can be denoted by

$$\boldsymbol{Y} = \boldsymbol{C}\boldsymbol{X}\boldsymbol{C}^{\mathrm{T}}, \tag{1}$$

$$\boldsymbol{C}(i, j) = \begin{cases} \sqrt{\dfrac{1}{N}}, & i = 0,\ j = 0,1,\cdots,N-1, \\ \sqrt{\dfrac{2}{N}}\cos\dfrac{i(2j+1)\pi}{2N}, & i = 1,2,\cdots,N-1,\ j = 0,1,\cdots,N-1. \end{cases} \tag{2}$$

where $\boldsymbol{C}^{\mathrm{T}}$ is the transpose matrix of $\boldsymbol{C}$.

Since DCT is an orthogonal transform, *i.e.*, $\boldsymbol{C}^{\mathrm{T}} = \boldsymbol{C}^{-1}$, we use

$$\boldsymbol{X} = \boldsymbol{C}^{-1}Y(\boldsymbol{C}^{\mathrm{T}})^{-1} = \boldsymbol{C}^{-1}Y(\boldsymbol{C}^{-1})^{\mathrm{T}} = \boldsymbol{C}^{-1}YC \tag{3}$$

to reconstruct the image, where $\boldsymbol{C}^{-1}$ is the inverse matrix of $\boldsymbol{C}$.

2.2. All Phase Biorthogonal Transform (APBT)

On the basis of all phase digital filtering [8], three kinds of all phase biorthogonal transforms based on the WHT, DCT and IDCT were proposed and the matrices of APBT were deduced in [7]. Similar to DCT matrix, it can be used in image compression transforming the image from spatial domain to frequency domain too.

Taking all phase inverse discrete cosine biorthogonal transform (APIDCBT) for example, the process of two-dimensional APBT is introduced as follows. Let $\boldsymbol{X}$ and $\boldsymbol{V}$ represent an image block and APIDCBT matrix with size of $N \times N$ respectively. After two-dimensional APIDCBT transform, transform coefficients block $\boldsymbol{Y}$ can be denoted by

$$\boldsymbol{Y} = \boldsymbol{V}\boldsymbol{X}\boldsymbol{V}^{\mathrm{T}}, \tag{4}$$

$$\boldsymbol{V}(m,n) = \begin{cases} \dfrac{1}{N}, & m = 0,\ n = 0,1,\cdots,N-1, \\ \dfrac{N-m+\sqrt{2}-1}{N^2}\cos\dfrac{m(2n+1)\pi}{2N}, & m = 1,2,\cdots,N-1,\ n = 0,1,\cdots,N-1. \end{cases} \tag{5}$$

where $\boldsymbol{V}^{\mathrm{T}}$ is the transpose matrix of $\boldsymbol{V}$. We use

$$\boldsymbol{X} = \boldsymbol{V}^{-1}\boldsymbol{Y}(\boldsymbol{V}^{-1})^{\mathrm{T}}, \tag{6}$$

to reconstruct the image, where $\boldsymbol{V}^{-1}$ is the inverse matrix of $\boldsymbol{V}$.

3. I Frame Encoding Algorithm

Video coding methods exploit both temporal and spatial redundancy to achieve compression. In the temporal domain, there is usually a high correlation between temporally adjacent frames, especially if the frame rate is high. In the spatial domain, there is a high correlation between pixels that are close to each other. A video encoding unit includes an I frame and many P frames for MPEG-4 simple profile. The I frame is a frame of video encoded independently, only reducing spatial redundancy. To reduce the temporal and spatial redundancy simultaneously, P frame puts the reconstructed frame of previous coded one as the reference frame to perform ME and MC, and then the residual frame proceeds with intra-frame encoding [9]. So it is obvious that the encoding of I frame affects not only the reconstructed effect of itself, but also that of the subsequent P frames. Based on the noticeable block artifacts of reconstructed video sequence using DCT-based MPEG-4 compression algorithm and the better performance of APBT at low bit rates, we decide to adopt APBT in I frame encoding of proposed algorithm.

Figure 1 shows the main procedures of encoding and decoding processes of I frame of video compression algorithm based on APBT for MPEG-4 simple profile. It illustrates the special case of a single-component video object plane (VOP). Since the input to MPEG-4 encoder is a YUV video sequence in 4:2:0 progressive formats, all processes operate on each VOP component (Y, U, V) independently. Substantially the same as basic steps of the baseline MPEG-4 I frame compression algorithm, there are only differences in the transform step and quantization step.

In the encoding process, the input I frame data are grouped into 8×8 blocks, and each block is transformed by APBT into 64 APBT coefficients. The coefficient in upper and left is direct current (DC) coefficient and the other 63 coefficients are alternating current (AC) coefficients. The APBT coefficients have high-frequency attenuation characteristics, that is to say, the transform coefficients of APBT have different frequency weighted in the transform process. When the APBT coefficients are quantized by uniform quantization step, it is equivalent to the low-frequency coefficients are quantized by small step and high-frequency coefficients are quantized by big step. Quantization results are similar to DCT which adopts complex quantization table. Therefore the uniform quantization step can be applied to encoding and decoding system based on APBT. After quantization, the APBT coefficients continue to go through the process of DC & AC prediction. The DC coefficient of current block is predicted from the DC coefficient of upper or left previously-coded 8×8 block. The direction of the smallest DC gradient is chosen as the prediction direction of the DC coefficient of current block. The prediction of AC coefficients is carried out in a similar way, with the first row or column of AC coefficients predicted in the direction determined for the DC coefficient. For example, if the prediction direction is from left block, the first column of AC coefficients in current block is predicted from the first column of left block. If the prediction direction is from upper block, the first row of AC coefficients in current block is predicted by the first row of upper block. We can determine whether the AC coefficients are predicted by the difference between the sum of the absolute value of AC coefficients and the sum of the absolute value which is got by the difference between AC coefficients and its prediction values. The AC coefficients will be predicted if the difference is larger than zero. After DC & AC prediction, the APBT coefficients are prepared by scan for variable length encoding. In DCT-based JPEG (DCT-JPEG) image compression algorithm, zigzag scan is used. Different from DCT-JPEG image compression algorithm, three kinds of scan patterns are used for I frame encoding method in MPEG-4 (as shown in Figure 2). If AC coefficients are not predicted, zigzag scan is selected. Otherwise, DC coefficient prediction direction is used to select a scan on block basis. For instance, if the DC prediction refers to the horizontally adjacent block, alternate-vertical scan is adopted for the current block. Otherwise, alternate-horizontal scan is used in the current block. Thereafter, the quantized coefficients are passed to the procedure of variable length encoding which compresses data further.

In contrast to the encoder, each step of decoding processes performs essentially the inverse of its corresponding main procedures within the encoder. The compressed I frame data go through the variable length decoding to get the scan sequence of quantized APBT coefficients. Then the one-dimensional scan sequence of quantized APBT coefficients becomes two-dimensional block by inverse scan. Thereafter the obtained block performs the step of inverse DC & AC prediction. After dequantization, APBT coefficients are transformed to 8×8 block of samples by inverse APBT. Therefore, the reconstructed I frame is obtained.

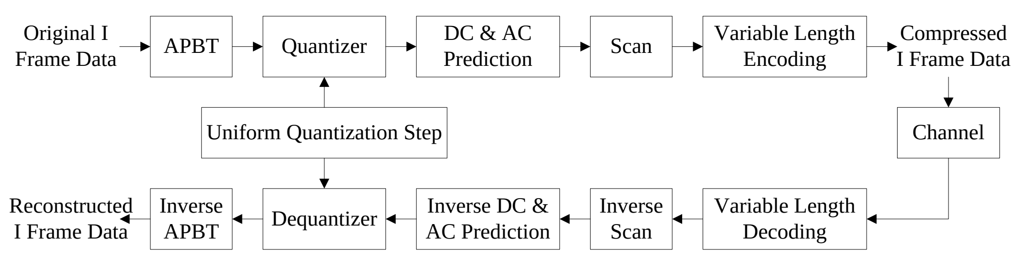

Figure 1. Simplified Diagram of I Frame of Video Encoder and Decoder based on APBT for MPEG-4 Simple Profile

0	1	2	3	10	11	12	13
4	5	8	9	17	16	15	14
6	7	19	18	26	27	28	29
20	21	24	25	30	31	32	33
22	23	34	35	42	43	44	45
36	37	40	41	46	47	48	49
38	39	50	51	56	57	58	59
52	53	54	55	60	61	62	63

(a)

0	4	6	20	22	36	38	52
1	5	7	21	23	37	39	53
2	8	19	24	34	40	50	54
3	9	18	25	35	41	51	55
10	17	26	30	42	46	56	60
11	16	27	31	43	47	57	61
12	15	28	32	44	48	58	62
13	14	29	33	45	49	59	63

(b)

0	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63

(c)

Figure 2. Three scan patterns: (a) Alternate-horizontal scan, (b) Alternate-vertical scan, (c) Zigzag scan

4. P Frame Encoding Algorithm

Figure 3 shows the proposed encoding framework based on APBT for MPEG-4 simple profile. The corresponding decoding framework is shown in Figure 4. Different from I frame encoding, P frame sets the reconstructed frame of previous coded one as the reference frame to perform ME and MC before intra-frame encoding [10]. By reducing temporal redundancy and spatial redundancy, the compression capability of P frame is much better than I frame.

The coding process of P frame is introduced in detail as follows. Firstly each P frame puts the reconstructed frame of previous coded one as the reference frame to perform ME and MC. Basic motion compensation scheme is block-based compensation of 16×16 pixel macroblocks. The DSA is adopted in motion estimation process. DSA is a search algorithm based on large diamond search and small diamond search (Figure 5). The large diamond search has nine search points, the step of which is two. The small diamond search has five search points, the step of which is one. While the DSA applying to search process, large diamond search pattern is firstly used. If the optimal point is not located in the center of large diamond, the large diamond are moved along the direction of the optimal point and then the same search process proceeds until the optimal point is located in the center of large diamond. Thereafter the search method is switched to small diamond search. The minimum error point of small diamond search is the best matching point. The offset between the current macroblock and the compensation region in the reference frame may have half-pixel resolution. The predicted samples at sub-pixel positions are calculated using bilinear interpolation between samples at integer-pixel positions. Therefore, the motion vector (MV) and predicted block of current block are obtained by motion estimation and motion compensation. Then the predicted frame is subtracted by current frame to get the residual frame. The DCT-based intra-frame coding is performed on residual frame. The residual frame is divided into 8×8 blocks, and each block is transformed by forward DCT (FDCT) into 64 DCT coefficients. The 64 DCT coefficients are quantized by quantization table in MPEG-4 standard. And the luminance quantization table and chrominance quantization table are different. The subsequent procedures of P frame encoding are basically the same with those of I frame encoding except the scan process. The 8×8 blocks of transform coefficients are scanned in zigzag scanning sequence for P frame encoding.

The decoding process of P frame is relatively simple, compared to the encoding process of P frame. Firstly the received code stream passes the demultiplexer to get MV and texture information. The MV is decoded for motion compensation process. The texture information is entropy decoded and then inverse zigzag scanned. Thereafter the resulting quantization coefficients block performs the step of inverse DC & AC prediction. After dequantization, the DCT coefficients block is transformed by inverse DCT to obtain the residual block. Then the reconstructed frame of the previous coded one together with MV is used for motion compensation. The resulting predicted frame is added to residual frame to get the reconstructed current frame.

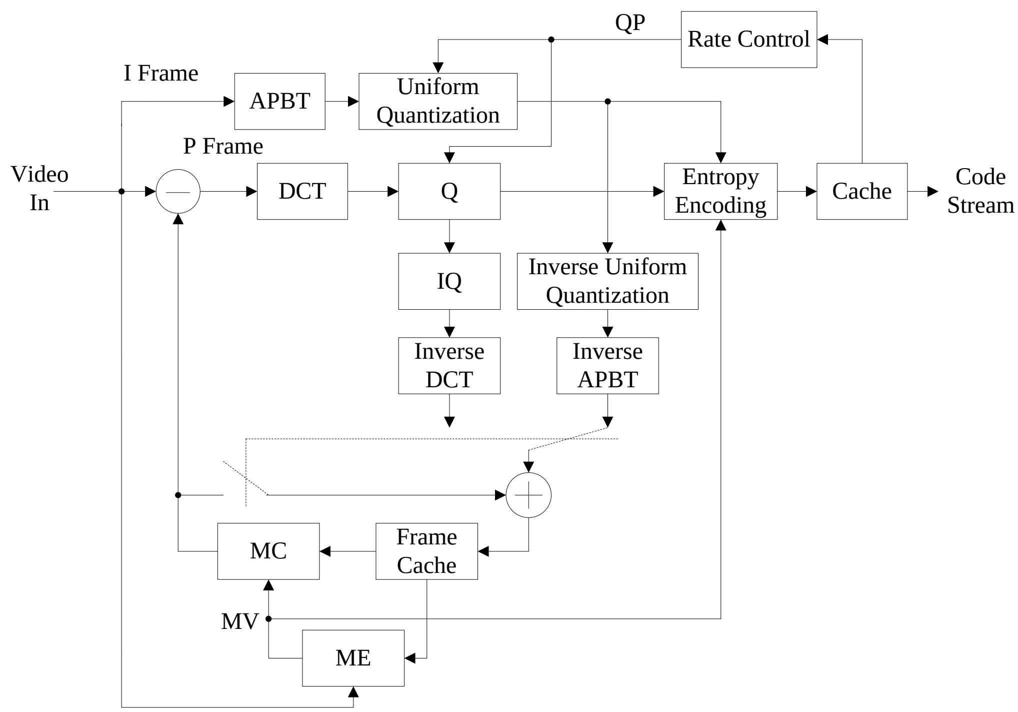

Figure 3. Proposed Encoding Framework based on APBT for MPEG-4 Simple Profile

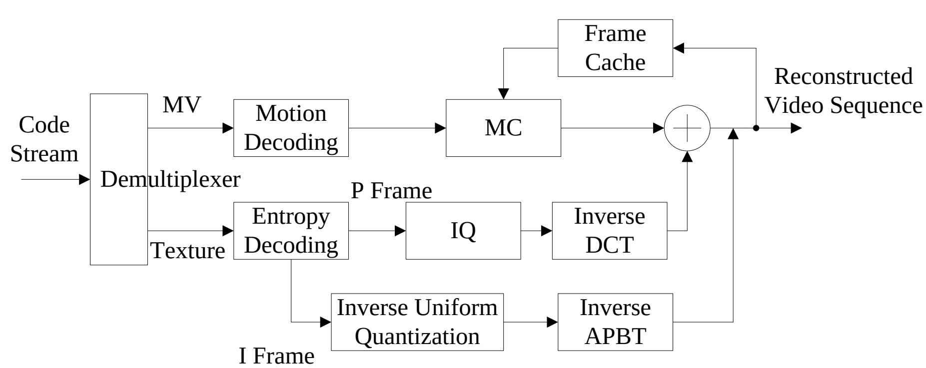

Figure 4. Proposed Decoding Framework based on APBT for MPEG-4 Simple Profile

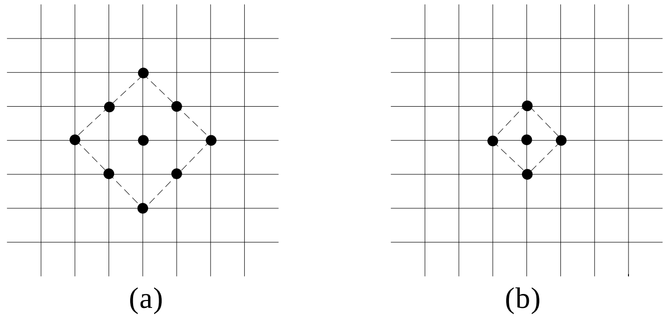

Figure 5. Diamond Search Algorithm: (a) Large Diamond Search, (b) Small Diamond Search

5. Experimental Result

All experimental results in this paper are achieved by C language in the environment of VC++ 6.0. The test video sequences are in YUV format, with the size of 352×288. The frame rate is 25 frames/s and the bit rate is 300kbps. The interval between two I frames is 15 frames.

To test the proposed algorithm and compare with MPEG-4 based on DCT, the video sequence football is applied in simulation experiments. Figure 6 shows that the experimental results with the DCT, APIDCBT, APDCBT and APWBT in terms of Y component's peak signal to noise ratio (PSNR) for different frames.

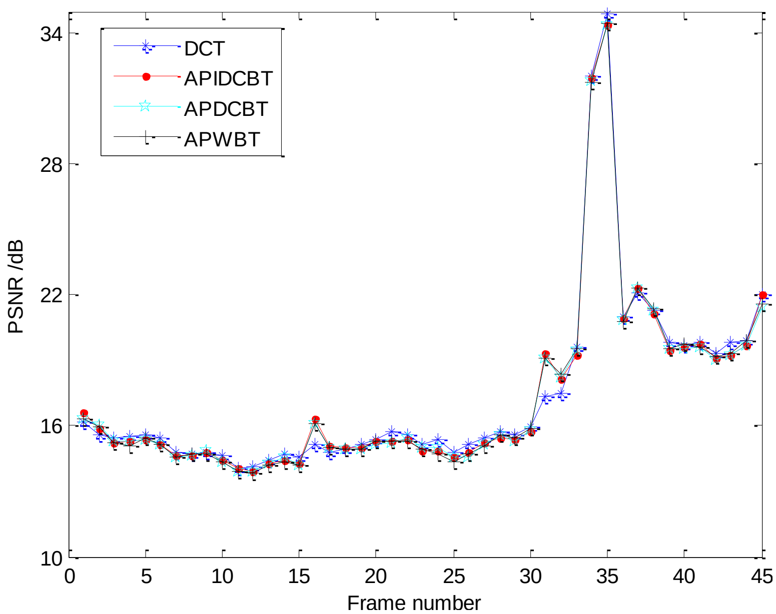

Figure 6. Results of Different Algorithms Applied to Video Sequence Football

From the experimental results in Figure 6, we conclude that the performance of proposed algorithm based on APBT is close to the MPEG-4 based on DCT in video compression. In order to compare compression performance subjectively, Figure 7 shows the 87th frame of reconstructed football video sequence obtained by using DCT, APIDCBT, APDCBT and APWBT. It can be seen that the block artifacts of the 87th frame of reconstructed football video sequence based on DCT are obvious, while the block artifacts of the 87th frame of reconstructed football video sequence based on APBT are reduced greatly.

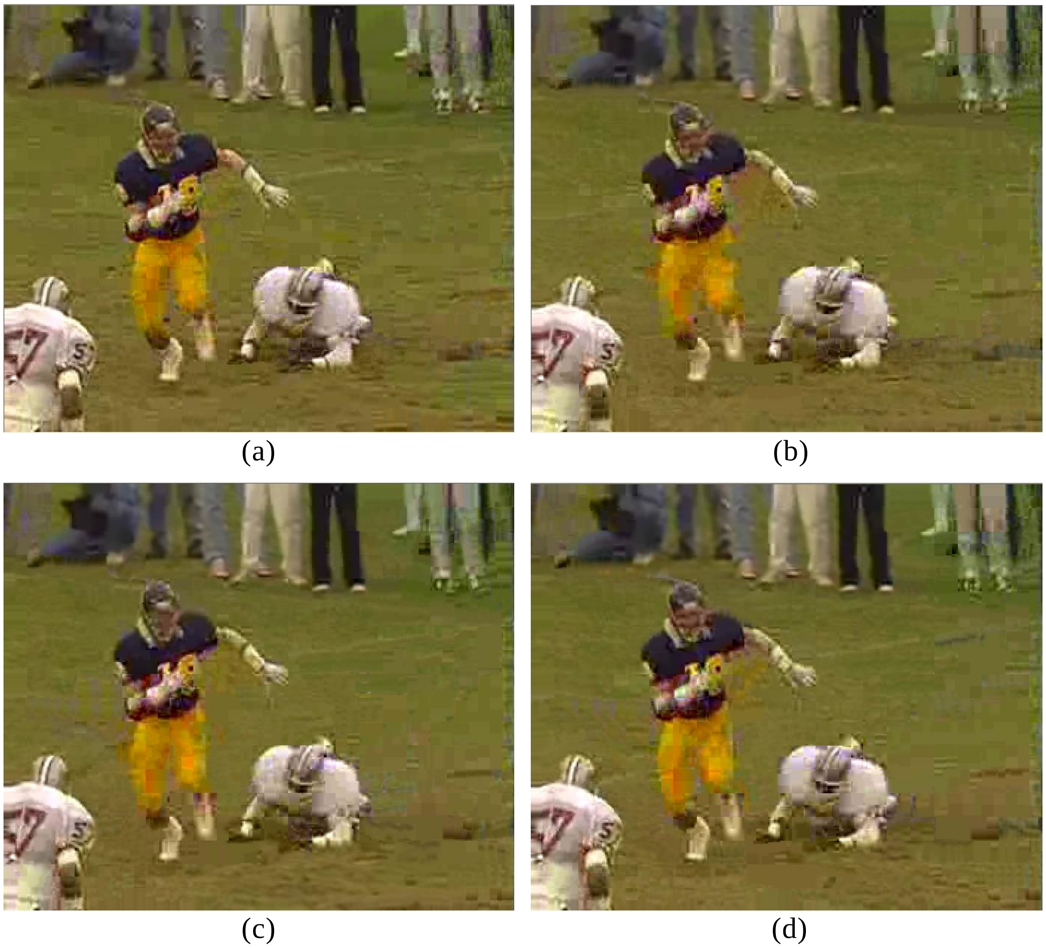

Figure 7. The 87th Frame of Reconstructed Football Video Sequence based on Different Algorithms: (a) DCT, (b) APIDCBT, (c) APDCBT, (d) APWBT

In conclusion, the proposed algorithm based on APBT performs close to MPEG-4 based on DCT in terms of objective and subjective effects. Complex quantization table is adopted in MPEG-4 based on DCT, which increases the memory space of quantization table and 63 multiplication operations between quantization factor and quantization Table for each image block. With the use of uniform quantization step, the proposed algorithm based on APBT can simplify the quantization process and save the computing time.

6. Conclusion

In this paper, an efficient algorithm based on APBT is proposed for encoding of MPEG-4 simple profile. In the proposed algorithm, I frame encoding uses APBT and uniform quantization operation; for P frames, each frame sets the reconstructed frame of previous coded one as the reference frame for ME and MC, and then the encoding of residual frame adopts DCT. The experiments to typical test video sequence are done in VC++6.0. Compared with MPEG-4 based on DCT, similar performance is obtained. Because of the use of uniform quantization step, the proposed algorithm saves computation time and is easier for hardware implementation. In the future, we will apply shape adaptive all phase biorthogonal transform (SA-APBT) [5] to the encoding of MPEG-4 advanced coding efficiency profile.

Acknowledgements

This work was supported by the Graduate Innovation Foundation of Shandong University at WeiHai (GIFSDUWH) (Grant No. yjs12028), the National Natural Science Foundation of China (Grant No. 61201371) and the promotive research fund for excellent young and middle-aged scientists of Shandong Province, China (Grant No. BS2013DX022). The authors would like to thank Songzhao Xie for his help and valuable suggestions. The authors also thank the anonymous reviewers and the editor for their valuable comments to improve the presentation of the paper.

References

[1] G. K. Wallace, "The JPEG still picture compression standard", Communications of the ACM, vol. 34, no. 4, (**1991**), pp. 30-44.
[2] ITU-T Recommendation, "H.263 Version 2: Video Coding for Low Bit Rate Communication", Draft I, (**1998**).
[3] ISO/IEC JTC1/SC29/WG11, N2502a, "Generic Coding of Audio-Visual Objects Visual 14496-2", Final Draft IS, (**1998**).
[4] K. Jinzenji, H. Watanabe, S. Okada and N. Kobayashi, "Very low bit rate video compression using MPEG-4 sprite coding", Electronics and Communications in Japan, Part II: Electronics, vol. 86, no. 6, (**2003**), pp. 83-93.
[5] B. C. Jiang, A. P. Yang, C. Y. Wang and Z. X. Hou, "Shape adaptive all phase biorthogonal transform and its application in image coding", Journal of Communications, vol. 8, no. 5, (**2013**), pp. 330-336.
[6] K. K. Kwon, S. H. Im and D. S. Lim, "Deblocking algorithm in MPEG-4 video coding using block boundary characteristics and adaptive filtering", Proceedings of the IEEE International Conference on Image Processing, Genova, Italy, vol. 3, (**2005**) September 11-14, pp. 541-544.
[7] Z. X. Hou, C. Y. Wang and A. P. Yang, "All phase biorthogonal transform and its application in JPEG-like image compression", Signal Processing : Image Communication, vol. 24, no.10, (**2009**), pp. 791-802.
[8] Z. X. Hou and X. Yang, "The all phase DFT filter", Proceedings of the 10th IEEE Digital Signal Processing Workshop and the 2nd IEEE Signal Processing Education Workshop, Pine Mountain, Georgia, USA, (**2002**) October 13-16, pp. 221-226.
[9] H. C. Chang, L. G. Chen, M. Y. Hsu and Y. C. Chang, "Performance analysis and architecture evaluation of MPEG-4 video codec system", Proceedings of the IEEE International Symposium on Circuits and Systems, Geneva, Switzerland, vol. 2, (**2000**) May 28-31, pp. 449-452.
[10] Z. X. Hou, B. C. Jiang, Y. P. Cao, A. P. Yang and C. Y. Wang, "Video compression based on wavelet transform and DBMA with motion compensation", Journal of Communications, vol. 8, no. 7, (**2013**), pp. 440-448.

Authors

Xiaoyan Wang, she was born in Shandong province, China in 1990. She received her B.S. degree in electronic information science and technology from Shandong University, Weihai, China, in 2012. Now she is pursuing her M.E. degree in circuits and systems in Shandong University, Weihai, China. Her research interests concentrate on image and video processing.

Baochen Jiang, he was born in Shandong province, China in 1962. He received his B.S. degree in radio electronics from Shandong University, China, in 1983 and his M.E. degree in communication and electronic systems from Tsinghua University, China, in 1990. Now he is a professor in the School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai, China. His current research interests include signal and information processing, image and video processing, and smart grid technology.

Chengyou Wang, he was born in Shandong province, China in 1979. He received his B.E. degree in electronic information science and technology from Yantai University, China, in 2004 and his M.E. and Ph.D. degree in signal and information processing from Tianjin University, China, in 2007 and 2010 respectively. Now he is an associate professor in the School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai, China. His current research interests include digital image and video processing, multidimensional signal and information processing, and smart grid technology.

Zhiqiang Yang, he was born in Shandong province, China in 1954. He received his B.S. degree and M.E. degree in electronics from Shandong University, China. Now he is an associate professor in the School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai, China and he is also with the Integrated Electronic Systems Lab Co. Ltd., Jinan, China. His current research interests include electronic information system, software and power system automation.

Chunxiao Zhang, he was born in Shandong province, China in 1990. He received his B.S. degree in electronic information science and technology from Shandong University, Weihai, China, in 2012. Now he is pursuing his M.E. degree in circuits and systems in Shandong University, Weihai, China. His research interests concentrate on image and video processing.

