변환 영역에서 개선된 DCT를 기반으로 한 움직임 예측 및 보상

Motion Estimation and Compensation based on Advanced DCT

장 영, 조 효문, 조 상복
Young Jang, Hyo-Moon Cho, Sang-Bock Cho

Abstract - In this paper, we propose a novel architecture, which is based on DCT (Discrete Cosine Transform), for ME (Motion Estimation) and MC (Motion Compensation). The traditional algorithms of ME and MC based on DCT did not suffer the advantage of the coarseness of the 2-dimensional DCT (2-D DCT) coefficients to reduce the operational time. Therefore, we derive a recursion equation for transform-domain ME and MC and design the structure by using highly regular, parallel, and pipeline processing elements. The main difference with others is removing the IDCT block by using to transform domain. Therefore, the performance of our algorithm is more efficient in practical image processing such as DVR (Digital Video Recorder) system. We present the simulation result which is compare with the spatial domain methods, it shows reducing the calculation cost, compression ratio, and peak signal to noise ratio (PSNR).

Key Words : Motion estimation, Motion compensation, DCT, Recursion equation

1. INTRODUCTION

Recently, the video coding technologies are used in many fields such as HDTV, video streaming on the Inter-Net and mobile phone. The most video compression technologies are based on the DCT(Discrete Cosine Transform). The DCT is widely used to reduce the computational complex, because the motion estimation and compensation with mode decision took the much portion of the total computation in video compression such as H.264.

The motion estimation algorithms may be achieved by spatial domain approaches and/or transform based approaches. In spatial domain approaches of motion estimation, block matching methods and gradient based methods are use. Otherwise in the transform based approaches, phase correlation algorithms, wavelet-transform based algorithms, and DCT based algorithm are included.

This DCT based algorithms have some advantages than others in spatial domain approaches or transform based approaches. Because, it can reduce the computational complexity and be easy implemented by hardware or software.

However, the traditional DCT based algorithm have some problems such as design complexity, it means power consumption is so high and the computational complexity of itself is relatively high. Therefore, we introduce the advanced DCT based algorithm which solve these problems.

The main concept of our proposal is eliminating of IDCT(Inverse DCT) block for practical images, which have much motion informations. Thus, our algorithm is more efficient for dynamic moving pictures, but it is inefficient for relatively static images. Futhermore, we will add switching block, which switches between dynamic motion processor and static motion, to this proposed algorithm, and then we can obtain the powerful DCT based algorithm.

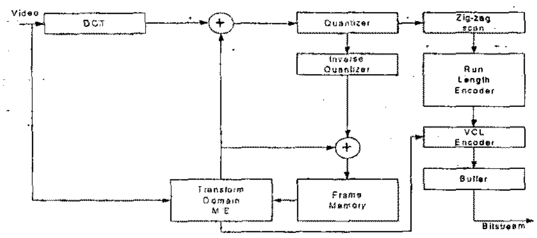
This paper is organized as follows. In section 2, the DCT based motion estimation and compensation algorithm is explained. In the section 3, a new advanced DCT algorithm is introduced. In the section 4, we describe the comparing result of the performance, such as computational complexity and PSNR(Peak Signal to Noise ratio), between the proposed algorithm and the conventional. And finally in the section 5, we briefly summarize our work and conclude this paper.

저자 소개
* 장 영 : 울산대학 전자공학과 석사과정
** 조효문: 울산대학 전자공학과 박사과정
*** 조상복: 울산대학 전자공학과 교수

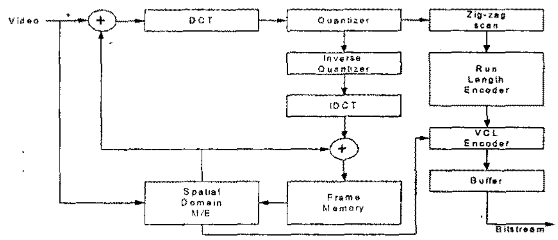
2. Conventional DCT-based motion estimation

The N x N DCT matrix $T=\{t(k, n)\}$, where $t(k, n)$ represents the matrix entry in the k-th row and the n-th column, is give as

$$t(k,n)=\begin{cases}\frac{1}{\sqrt{N}}, & k=0,\ 0\le n\le N-1\\ \sqrt{\frac{2}{N}}\cos\frac{\Pi(2n+1)k}{2N} & 1\le k\le N-1,\ 0\le n\le N-1\end{cases} \tag{1}$$



(a)



(b)

Fig.1 video encoder architecture (a) Conventional video encoder with motion estimation and compensation in the spatial domain (b) Simplified video encoder with motion estimation and compensation in the transform domain.

And the 1-Dimensional DCT of a sequence x(n), 0≤n≤N-1 is given as

$$y(k) = \sum_{n=0}^{N-1} t(k,n)x(n), \quad 0 \le k \le N-1 \tag{2}$$

In case of N=8, the 8 × 8 block DCT is used according to video compression standards. The shift matrix-based technique, which is proposed by Plompen et al., improves the image quality by using the integer-pixel accurate motion estimation and compensation. That is, more accurate predictions increase the compression ratio and improve PSNR. And the improving utilizing the prediction accuracy is based on sub-pixel. Generally, the pixel values on the half-pixel locations, which is the nearest pixel values of it, are obtained by bilinear interpolation.

A predicted block f_{pred} can be represented as a summation of horizontal and vertical shifted version of the four surrounding blocks f_0, f_1, f_2, and f_3. The eq. (3) can be calculated by applying vertical shifting matrix V_i and horizontal shifting matrix H_i. The currently block f_{curr} is predicted from the best estimated f_{curr} and the corresponding motion vector as indicated in Fig. 2.

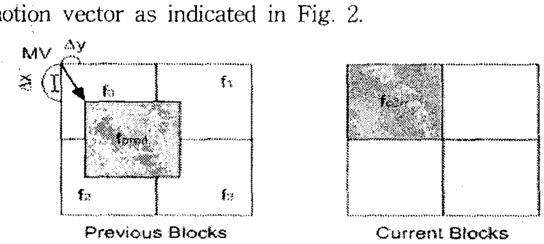


Fig. 2. Motion estimation

The predicted block f_{pred} can be obtain by

$$f_{pred} = \sum_{i=0}^{3} V_i f_i H_i \tag{3}$$

The vertical shifting matrix V_i and the horizontal shifting matrix H_i are given as

$$V_0 = D_{8-\Delta x},\ V_1 = D_{8-\Delta x},\ V_2 = D^t_{\Delta x},\ V_3 = D^t_{\Delta x} \tag{4}$$

$$H_0 = D^t_{8-\Delta y},\ H_1 = D_{\Delta y},\ H_2 = D^t_{8-\Delta y},\ H_3 = D_{\Delta y} \tag{5}$$

In (4) and (5), displacement matrix Dn is defined as

$$\left[\begin{array}{c|ccc} & & I_n & \\ - & - & - & - \end{array}\right] \tag{6}$$

where I_n is n×n identity matrix.

The motion estimation predicted error can be reduced by using the sub-pixel. To reduce this error, the motion vectors can be specified to half-pixel accuracy. However, to do this increases the computational complex and storage requirements.

Therefore, as shown in (7) - (10), we can derive the generalized vertical shifting matrix V_i and horizontal shifting matrix H_i to support half-pixel accurate motion estimation method in the transform domain without additional expense. The calculation quantity diminishes.

$$V_0 = V_1 = \frac{1}{2}(D_{8-\lfloor\Delta x\rfloor} + D_{8-\lceil\Delta x\rceil}) \tag{7}$$

$$V_2 = V_3 = \frac{1}{2}(D^t_{\lfloor\Delta x\rfloor} + D^t_{\lceil\Delta x\rceil}) \tag{8}$$

$$H_0 = H_2 = \frac{1}{2}(D^t_{8-\lfloor\Delta y\rfloor} + D^t_{8-\lceil\Delta y\rceil}) \tag{9}$$

$$H_1 = H_3 = \frac{1}{2}(D_{\lfloor\Delta y\rfloor} + D_{\lceil\Delta y\rceil}) \tag{10}$$

The main characteristics of the DCT are orthogonality and separability. Thus we can obtain as belows

$$DCT(f) \cong \hat{f} = TfT^t \tag{11}$$

$$T \cdot T^t = I \tag{12}$$

$$DCT(ab) = DCT(a)DCT(b) = \hat{a}\hat{b}. \tag{13}$$

Therefore, the DCT coefficients of the predicted block f_{pred} can be obtained by using the DCT coefficients of the previous surrounding blocks f_0, f_1, f_2, and f_3, the pre-calculated DCT coefficients of the horizontal shifting matrix H_i and the vertical shifting matrix V_i can be obtained as shown in (14)

$$\widehat{f_{pred}} = DCT(f_{pred}) = DCT(\sum_{i=0}^{3} V_i f_i H_i) = \sum_{i=0}^{3} DCT(V_i)DCT(f_i)DCT(H_i) \tag{14}$$

$$\hat{V}_i = TV_iT^t,\ \hat{f}_i = Tf_iT^t,\ \hat{H}_i = TH_iT^t \tag{15}$$

And also, the motion estimation and compensation are derived by eq. (15).

3. Proposed DCT-based algorithm

The main defect of the conventional DCT-based motion estimation is a high computational complexity, since it is constructed by eight matrix multiplications and three matrix addition to obtain the DCT coefficients of the predicted block f_{pred}.

To reduce this computational complexity, we propose a new approach with the energy concentration property in the frequency domain.

Table I. Average Number of Non-zero DCT coefficients per Block in the P frame

QP		3	6	9	12	15	18	21	24	27	30
Foreman	QCIF	18.3	8.5	5.7	3.6	3.0	2.1	1.8	1.3	1.3	1.0
News	QCIF	11.6	6.3	5.7	4.0	4.0	3.1	2.9	2.3	2.3	1.8
Paris	CIF	15.3	8.5	5.7	5.6	5.1	4.1	3.8	3.1	2.9	2.4

Table II. Simulation Results of Spatial Domain Motion Estimation and Transform Domain Motion Estimation

QP			3	6	9	12	15	18	21	24	27	30
Spatial Domain	FormanQ CIF	PSNR	40.3520	36.3084	33.3973	31.9589	30.4664	29.7497	28.7782	28.2382	27.5203	27.1183
		Total Bits /P frame	39194	19997	15395	11583	10301	9337	8809	8460	8274	8037
	News QCIF	PSNR	41.5328	37.1467	33.3314	32.3314	30.6977	29.8065	28.7960	28.1416	27.3975	26.8663
		Total Bits /P frame	27525	17943	16606	14121	13506	12421	11840	11063	10631	10440
	Paris CIF	PSNR	40.8560	36.3569	33.0068	31.3658	29.6853	28.7348	27.7015	27.0420	26.3094	25.8387
		Total Bits /P frame	140876	89170	80744	66997	62784	56985	55004	50415	49132	46406
Transform Domain	Forman QCIF	PSNR	40.6514	36.5406	33.6242	32.2298	30.7994	30.0861	29.1231	28.5348	27.9319	27.6576
		Total Bits /P frame	38774	19512	14847	11347	10127	9181	8583	8189	7968	7603
	News QCIF	PSNR	41.6089	37.2106	32.4588	34.4588	30.8692	29.9322	28.9918	28.3727	27.6876	27.1765
		Total Bits /P frame	27395	18027	16472	13931	13597	12434	11927	11102	10673	10481
	Paris CIF	PSNR	40.9251	36.4184	33.1009	31.4607	29.8147	28.8653	27.8550	27.2132	26.4814	26.0336
		Total Bits /P frame	138825	88142	79663	66134	62057	56309	54264	50179	48960	46229

Quantized block $\hat{f}(m, n)$ is calculated using $V_i(m)$, which is the m-th column of matrix V_i and the row vector $H_i(n)$, which is the nth row of matrix H_i. Therefore, (14) can be modified to the DCT coefficient block f_{pred}, as shown in (16) :

$$\widehat{f_{pred}} = \sum_{i=0}^{3} \hat{V}_i \hat{f}_i \hat{H}_i = \sum_{i=0}^{3} \sum_{m=0}^{7} \sum_{n=0}^{7} \hat{f}_i(m,n)\, \hat{V}_i(m,n) \hat{H}_i(m,n)$$
$$= \sum_{m=0}^{7} \sum_{n=0}^{7} \sum_{i=0}^{3} \hat{f}_i(m,n)\, \hat{V}_i(m,n) \hat{H}_i(m,n) \qquad (16)$$

to this equation, the outer product is used because the outer products can be omitted by zero values of $\hat{f}_i(m, n)$. According to the DCT theory, the low frequency components are non zero, but high frequency components are almost likely zero. Therefore, the computational complexity can be reduced. Eq. (16) can be re-organized in zigzag scan order as shown in (17)

$$\widehat{f_{pred}} = \sum_{m=0}^{7} \sum_{n=0}^{7} \sum_{i=0}^{3} \hat{f}_i(m,n)\, \hat{V}_i(m,n) \hat{H}_i(m,n)$$
$$= \hat{f}_0(0,0)\, \hat{V}_0(0) \hat{H}_0(0) + \hat{f}_1(0,0)\, \hat{V}_1(0) \hat{H}_1(0)$$
$$+ \hat{f}_2(0,0)\, \hat{V}_2(0) \hat{H}_2(0) + \hat{f}_3(0,0)\, \hat{V}_3(0) \hat{H}_3(0)$$
$$+ \cdots + \hat{f}_0(7,7)\, \hat{V}_0(7) \hat{H}_0(7) + \hat{f}_1(7,7)\, \hat{V}_1(7) \hat{H}_1(7)$$
$$+ \hat{f}_2(7,7)\, \hat{V}_2(7) \hat{H}_2(7) + \hat{f}_3(7,7)\, \hat{V}_3(7) \hat{H}_3(7) \qquad (17)$$

4. Experimental results

This paper has described advanced DCT-based motion estimation and compensation. As the quantization parameter increases, the number of non-zero DCT coefficients decrease significantly. But as many non-zero DCT coefficients increases, the image quality is decreased, and quantization cost is increased as shown in table I. Table II shows the results of spatial domain motion estimation and transform domain motion estimation using the full search method. In this simulation, all the 8 x 8 DCT coefficients of the block are used to perform DCT-based motion estimation. Using transfer domain methods. we can obtain the high quality. The results of experimental, better visual quality in the aspects of PSNR and compression ratio are obtained.

5. Conclusion

A new algorithm, which uses the transform domain DCT-based motion estimation, is proposed. As the quantization parameter increases, the number of non-zero DCT coefficients decrease. The calculation cost and the number of non-zero DCT coefficients have tradeoff relationship. The proposed approach provides many useful features for power-aware video encoding despite the intensive computation, when compared to the spatial domain approaches. Simulation on video sequences of different characteristics shows the compression efficiency of the transfer domain approach with respect to the PSNR and the compression ratio are higher than that of the spatial domain approach.

Acknowledgement

This work was partly sponsored by ETRI SoC Industry Promotion Center, Human Resource Development Project for IT-SoC Architect and NARC (Network-based Automation Research Center) and 2 level BK21 (EVERDEC(e-Vehicle Research & human Resource Development Center)) in MOE (Ministry of Education & Human Resources Development).

참 고 문 헌

[1] P. Kuhn, Algorithms, Complexity Analysis and VLSI Architectures for MPEG-4 Motion Estimation. Boston, MA: Kluwer Academic, 1999.

[2] S. Cheng and H. Hang, "A comparison of block-matching algorithms mapped to systolic-array implementation," IEEE Trans. Circuits Syst. Video Technol., vol. 7, no. 5, pp. 741 - 757, Oct. 1997.

[3] A. Chimienti, C. Ferraris, and D. Pau, "A complexity-bounded motion estimation algorithm," IEEE Trans. Image Process., vol. 11, no. 4, pp. 387 - 392, Apr. 2002.

[4] N. Merhav and V. Bhaskaran, "A fast algorithm for DCT-domain inverse motion compensation," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., vol. 4, May 1996, pp. 2307 - 2310.