

DCT-BASED SUBPIXEL MOTION ESTIMATION

Ut-Va Koc

K. J. Ray Liu

Electrical Engineering Department and Institute for Systems Research
University of Maryland, College Park, Maryland 20742
koc@eng.umd.edu and kjrlu@eng.umd.edu

ABSTRACT

Inter-pixel interpolation is required for most subpixel motion estimation schemes but it undesirably increases the overall complexity and data flow and deteriorates estimation accuracy. In this paper, we develop DCT-based techniques to estimate subpel motion at different desired levels of accuracy in DCT domain without interpolation by establishing subpel sinusoidal orthogonal principles and showing that subpixel motion information is preserved in the DCT of a shifted signal under some condition in the form of pseudo phases. Though applicable to other areas, the resulted algorithms from these techniques for video coding are flexible and scalable with very low complexity $O(N^2)$ compared to $O(N^4)$ for block matching approaches. Importantly, DCT-based approaches enable simplification of the heavily loaded feedback loop of conventional hybrid video coder design, resulting in a high-throughput, low-complexity fully DCT-based coder. Finally simulation results show comparable performance of the proposed algorithms with block matching approaches.

1. INTRODUCTION

The motion compensated DCT-based hybrid approach is the backbone of several international video coding standards such as CCITT H.261, MPEG1, MPEG2 and the emerging HDTV and H.263 standards. A standard-compliant coder is usually implemented in a hybrid way — spatial compression is achieved in the transform domain while motion compensation is performed in the spatial domain to remove temporal redundancy, as shown in Fig. 1(a). As a result, this hybrid structure needs three major components in the feedback loop: DCT (Discrete Cosine Transform), IDCT (Inverse DCT) and motion compensator. This heavily loaded feedback loop not only complicates hardware design with high computational requirement but also limits the throughput of the whole coder. However, if motion can be estimated and compensated in the transform domain, then DCT can be moved out of the loop and IDCT be eliminated [1] as in Fig. 1(b). Based on the notion of pseudo-phases carrying motion information and the sinusoidal orthogonal principles, a low-complexity ($O(N^2)$) DCT-based integer-pel motion estimation scheme (DXT-ME) was proposed in [2] so that a fully DCT-based video coder can be accomplished.

To further improve the coder gain, motion estimation with subpixel accuracy is essential because the frame-to-frame displacements of a video sequence are not necessarily multiples of the sampling grid distance. It is shown in [3] that prediction error variance is generally decreased by subpixel motion compensation but beyond a certain “critical accuracy” the possibility of further improving prediction by more accurate motion compensation is small. Most currently existing subpixel motion estimation schemes estimate displacements from interpolated images [4]. However,

interpolation not only increases the complexity but also the data flow of a coder. Above all, interpolated images are only the approximation to the actual images and thus the selection of an interpolating function affects the accuracy of motion estimates from interpolated images [4]. Moreover, a simple interpolation usually introduces spatial lowpass filtering adversely affecting the estimation accuracy. It is more desirable that subpixel accuracy of motion estimates can be obtained without interpolating the images. From subpixel motion information preserved in the pseudo phases and by means of the subpel sinusoidal orthogonal principles developed in this paper, we devise DCT-based subpixel motion estimation algorithms to estimate subpixel motion at different required accuracy without image interpolation at a low computational cost. In this paper we show in Section 2 that under the Nyquist condition, the pseudo-phases in the DCT coefficients of two consecutive frames contain subpixel motion information. In Section 3, we develop the subpel sinusoidal orthogonal principles to estimate this subpixel displacement from the pseudo phases. Applying these concepts and techniques to video coding, we then propose half-pel and quarter-pel DCT-based motion estimation algorithms in Section 4, even though these DCT-based techniques are equally applicable to other applications such as in time delay estimation, target tracking, non-contact measurement, remote sensing, image registration, and so on. The simulation results of these algorithms on two video sequences of different characteristics are given in Section 5. Finally we conclude this paper in Section 6.

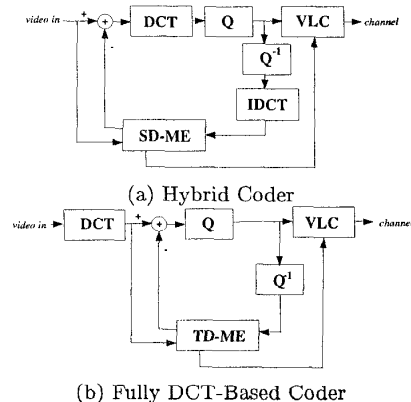


Figure 1. Coder structures: (a) Commonly used motion-compensated DCT hybrid coder performs motion estimation in the spatial domain. (b) Fully DCT-based coder estimates motion in the transform domain.

2. SUBPIXEL MOTION IN PSEUDO-PHASES

2.1. One Dimensional Signal Model

Without loss of generality, let us consider the one dimensional model in which a continuous signal $x_c(t)$ and its

This work is supported in part by the ONR grant N00014-93-1-0566, NSF grant MIP9457397, and MIPS/MicroStar.

shifted version $x_c(t-d)$ are sampled at a sampling frequency $1/T$ to generate two sample sequences $\{x_1(n) = x_c(nT)\}$ and $\{x_2(n) = x_c(nT-d)\}$, respectively. Define the DCT and DST coefficients of the second kind [5] as

$$X_i^C(k) \triangleq \text{DCT}\{x_i\} = \frac{2C(k)}{N} \sum_{n=0}^{N-1} x_i(n) \cos \frac{k\pi}{N} \left(n + \frac{1}{2}\right), \quad (1)$$

$$X_i^S(k) \triangleq \text{DST}\{x_i\} = \frac{2C(k)}{N} \sum_{n=0}^{N-1} x_i(n) \sin \frac{k\pi}{N} \left(n + \frac{1}{2}\right), \quad (2)$$

for $i = 1, 2$ where $C(k) = \frac{1}{\sqrt{2}}$ for $k = 0$ or N but $C(k) = 1$ otherwise. We can show that the DCT/DST and DFT coefficients are related as follows:

$$X_i^C(k) = \frac{C(k)}{N} [\tilde{X}_i^Z(-k)e^{j\frac{k\pi}{2N}} + \tilde{X}_i^Z(k)e^{-j\frac{k\pi}{2N}}], \quad (3)$$

for $k = 0, \dots, N-1$;

$$X_i^S(k) = \frac{C(k)}{jN} [\tilde{X}_i^Z(-k)e^{j\frac{k\pi}{2N}} - \tilde{X}_i^Z(k)e^{-j\frac{k\pi}{2N}}], \quad (4)$$

for $k = 1, \dots, N$,

where $\{\tilde{X}_i^Z(k)\}$ is the DFT of the zero-padded sequence $\{x_i^Z(n) : x_i^Z(n) = x_i(n)$ for $n = 0, \dots, N-1$; $x_i^Z(n) = 0$ for $n = N, \dots, 2N-1\}$ so that $\tilde{X}_i^Z(k) \triangleq \text{DFT}\{x_i^Z\} = \sum_{n=0}^{N-1} x_i(n)e^{-j\frac{2\pi kn}{2N}}$, for $k = 0, \dots, 2N-1$.

From the sampling theorem, the Discrete Time Fourier Transform (DTFT) of sequences $x_1(n)$ and $x_2(n)$ are related to the Fourier Transform (FT) of $x_c(t)$, $X_c(\Omega)$, in the following way:

$$X_1(\omega) \triangleq \text{DTFT}\{x_1\} = \frac{1}{T} \sum_l X_c\left(\frac{\omega - 2\pi l}{T}\right), \quad (5)$$

$$X_2(\omega) \triangleq \text{DTFT}\{x_2\} = \frac{1}{T} \sum_l e^{-j(\frac{\omega - 2\pi l}{T})d} X_c\left(\frac{\omega - 2\pi l}{T}\right). \quad (6)$$

Furthermore, if $X_c(\Omega)$ is bandlimited in the baseband $(-\frac{\pi}{T}, \frac{\pi}{T})$, then for $\Omega = \frac{\omega}{T} \in (-\frac{\pi}{T}, \frac{\pi}{T})$, then

$$X_1(\Omega T) = \frac{1}{T} X_c(\Omega), \quad X_2(\Omega T) = \frac{1}{T} X_c(\Omega) e^{-j\Omega d}.$$

Thus, the DFT of $x_1(n)$ and $x_2(n)$ are

$$\tilde{X}_1(k) \triangleq \text{DFT}\{x_1\} = \sum_{n=0}^{N-1} x_1(n) e^{-j\frac{2\pi kn}{N}} = \frac{1}{T} X_c\left(\frac{2\pi k}{NT}\right),$$

$$\tilde{X}_2(k) \triangleq \text{DFT}\{x_2\} = \sum_{n=0}^{N-1} x_2(n) e^{-j\frac{2\pi kn}{N}} = \frac{1}{T} X_c\left(\frac{2\pi k}{NT}\right) e^{-j\frac{2\pi kd}{N}}.$$

whereas the DFT of $x_1^Z(n)$ and $x_2^Z(n)$ become

$$\tilde{X}_1^Z(k) = X_1\left(\frac{\pi k}{N}\right) = \frac{1}{T} X_c\left(\frac{\pi k}{NT}\right), \quad (7)$$

$$\tilde{X}_2^Z(k) = X_2\left(\frac{\pi k}{N}\right) = \frac{1}{T} X_c\left(\frac{\pi k}{NT}\right) e^{-j\frac{\pi kd}{N}}. \quad (8)$$

Therefore,

$$X_2\left(\frac{\pi k}{N}\right) = X_1\left(\frac{\pi k}{N}\right) e^{-j\frac{\pi kd}{N}}. \quad (9)$$

Substituting (9) back into (3)-(4), we get

$$X_2^C(k) = \frac{C(k)}{N} [\tilde{X}_1^Z(-k)e^{j\frac{k\pi d}{N}} e^{j\frac{k\pi}{2N}} + \tilde{X}_1^Z(k)e^{-j\frac{k\pi d}{N}} e^{-j\frac{k\pi}{2N}}],$$

for $k = 0, \dots, N-1$,

$$X_2^S(k) = \frac{C(k)}{jN} [\tilde{X}_1^Z(-k)e^{j\frac{k\pi d}{N}} e^{j\frac{k\pi}{2N}} - \tilde{X}_1^Z(k)e^{-j\frac{k\pi d}{N}} e^{-j\frac{k\pi}{2N}}],$$

for $k = 1, \dots, N$.

Finally we obtain the relationship between $x_1(n)$ and $x_2(n)$ in DCT/DST domain:

$$X_2^C(k) = \frac{2C(k)}{N} \sum_{n=0}^{N-1} x_1(n) \cos \frac{k\pi}{N} \left(n + \frac{d}{T} + \frac{1}{2}\right), \quad (10)$$

for $k = 0, \dots, N-1$,

$$X_2^S(k) = \frac{2C(k)}{N} \sum_{n=0}^{N-1} x_1(n) \sin \frac{k\pi}{N} \left(n + \frac{d}{T} + \frac{1}{2}\right), \quad (11)$$

for $k = 1, \dots, N$.

We conclude the result in the following theorem:

THEOREM 1 If a continuous signal $x_c(t)$ is $\frac{\pi}{T}$ -bandlimited and the sampled sequences of $x_c(t)$ and $x_c(t-d)$ are $\{x_c(nT)\}$ and $\{x_c(nT-d)\}$, respectively, then their DCT and DST are related by

$$\text{DCT}\{x_c(nT-d)\} = \text{DCT}_{\frac{d}{T}}\{x_c(nT)\}, \quad (12)$$

$$\text{DST}\{x_c(nT-d)\} = \text{DST}_{\frac{d}{T}}\{x_c(nT)\}, \quad (13)$$

where

$$\text{DCT}_{\alpha}\{x\} \triangleq \frac{2C(k)}{N} \sum_{n=0}^{N-1} x(n) \cos \frac{k\pi}{N} \left(n + \alpha + \frac{1}{2}\right), \quad (14)$$

$$\text{DST}_{\beta}\{x\} \triangleq \frac{2C(k)}{N} \sum_{n=0}^{N-1} x(n) \sin \frac{k\pi}{N} \left(n + \beta + \frac{1}{2}\right), \quad (15)$$

are the DCT and DST with α and β shifts in their kernels, respectively. Here d is the shift amount and T is the sampling interval, but d/T is not necessarily an integer.

2.2. Two Dimensional Image Model

Consider a moving object casting a continuous intensity profile $I_t(u, v)$ on a camera plane of the continuous coordinate (u, v) where the subscript t denotes the frame number. This intensity profile is then digitized on the fixed sampling grid of the camera with a sampling distance d to generate the current frame of pixels $x_t(m, n)$ where m and n are integers. Further assume that the displacement of the object between the frames $t-1$ and t is (d_u, d_v) such that $I_t(u, v) = I_{t-1}(u-d_u, v-d_v)$ where $d_u = (m_u + \nu_u)d = \lambda_u d$ and $d_v = (m_v + \nu_v)d = \lambda_v d$. Here m_u and m_v are the integer components of the displacement, and ν_u and $\nu_v \in (-\frac{1}{2}, \frac{1}{2}]$. Therefore, $x_t(m, n) = I_t(md, nd) = I_{t-1}(md - d_u, nd - d_v)$, $x_{t-1}(m, n) = I_{t-1}(md, nd)$. Unlike the case of integer-pel movement, the displacement is not necessarily multiples of the sampling distance d . In other words, ν_u and ν_v do not necessarily equal zero.

For non-integer pel movement, if the Fourier transform of the continuous intensity profile $I_t(u, v)$ is $\frac{\pi}{d}$ -bandlimited and $I_t(u, v) = I_{t-1}(u-d_u, v-d_v)$, then according to Theorem 1, we can obtain the following 2-D relations:

$$X_t^{cc}(k, l) = \text{DCT}_{\lambda_u} \circ \text{DCT}_{\lambda_v} \{x_{t-1}(m, n)\}, \quad (16)$$

$k, l \in \{0, \dots, N-1\}$;

$$X_t^{cs}(k, l) = \text{DCT}_{\lambda_u} \circ \text{DST}_{\lambda_v} \{x_{t-1}(m, n)\}, \quad (17)$$

$k \in \{0, \dots, N-1\}, l \in \{1, \dots, N\}$;

$$X_t^{sc}(k, l) = \text{DST}_{\lambda_u} \circ \text{DCT}_{\lambda_v} \{x_{t-1}(m, n)\}, \quad (18)$$

$k \in \{1, \dots, N\}, l \in \{0, \dots, N-1\}$;

$$X_t^{ss}(k, l) = \text{DST}_{\lambda_u} \circ \text{DST}_{\lambda_v} \{x_{t-1}(m, n)\}, \quad (19)$$

$k, l \in \{1, \dots, N\}$,

where $\{X_t^{\sigma}(k, l)\}$; $\sigma = cc, cs, sc, ss$ are the DCT coefficients of the second kind of x_t [5] and $\text{DCT}_{\lambda_u} \circ \text{DST}_{\lambda_v}$ is the

composite function of DCT_{λ_u} and DST_{λ_u} defined in (14)-(15). Thus, rearranging the equations, we can obtain the *system equation* at the subpixel level [2]:

$$\mathbf{Z}_{t-1}(k, l) \cdot \vec{\theta}_{\lambda_u, \lambda_v}(k, l) = \vec{x}_t(k, l), \text{ for } k, l \in \mathcal{N}, \quad (20)$$

where $\mathcal{N} = \{1, \dots, N-1\}$, $\{Z_{t-1}^\sigma(k, l); \sigma = cc, cs, sc, ss\}$ are the DCT coefficients of the first kind of x_{t-1} [5], and $\vec{\theta}_{\lambda_u, \lambda_v} = [g_{\lambda_u, \lambda_v}^{CC}, g_{\lambda_u, \lambda_v}^{CS}, g_{\lambda_u, \lambda_v}^{SC}, g_{\lambda_u, \lambda_v}^{SS}]^T$. A similar relationship between the DCT coefficients of $x_t(m, n)$ and $x_{t-1}(m, n)$ at the block boundary can be obtained in the same way.

In (20), the pseudo phase vector $\vec{\theta}_{\lambda_u, \lambda_v}(k, l)$ contains the information of the subpixel movement (λ_u, λ_v) . In an ideal situation where one rigid object is moving translationally within the block boundary without observable background and noise, we can find $\vec{\theta}_{\lambda_u, \lambda_v}(k, l)$ explicitly in terms of λ_u and λ_v as such:

$$\vec{\theta}_{\lambda_u, \lambda_v}(k, l) = \begin{bmatrix} \cos \frac{k\pi}{N} (\lambda_u + \frac{1}{2}) \cos \frac{l\pi}{N} (\lambda_v + \frac{1}{2}) \\ \cos \frac{k\pi}{N} (\lambda_u + \frac{1}{2}) \sin \frac{l\pi}{N} (\lambda_v + \frac{1}{2}) \\ \sin \frac{k\pi}{N} (\lambda_u + \frac{1}{2}) \cos \frac{l\pi}{N} (\lambda_v + \frac{1}{2}) \\ \sin \frac{k\pi}{N} (\lambda_u + \frac{1}{2}) \sin \frac{l\pi}{N} (\lambda_v + \frac{1}{2}) \end{bmatrix}. \quad (21)$$

3. SUBPEL SINUSOIDAL ORTHOGONALITY PRINCIPLES

Estimation of integer-pel displacements in DCT domain is based on the sinusoidal orthogonal principles [2]:

$$\begin{aligned} & \frac{2}{N} \sum_{k=0}^{N-1} C^2(k) \cos\left[\frac{k\pi}{N} \left(m + \frac{1}{2}\right)\right] \cos\left[\frac{k\pi}{N} \left(n + \frac{1}{2}\right)\right] \\ &= \delta(m-n) + \delta(m+n+1), \end{aligned} \quad (22)$$

$$\begin{aligned} & \frac{2}{N} \sum_{k=1}^N C^2(k) \sin\left[\frac{k\pi}{N} \left(m + \frac{1}{2}\right)\right] \sin\left[\frac{k\pi}{N} \left(n + \frac{1}{2}\right)\right] \\ &= \delta(m-n) - \delta(m+n+1), \end{aligned} \quad (23)$$

where $\delta(n)$ is the discrete impulse function, and m, n are integers. This is no longer valid in the subpixel level. However, we define

$$L_c(u, v) \triangleq \sum_{k=0}^{N-1} \cos \frac{k\pi}{N} \left(u + \frac{1}{2}\right) \cos \frac{k\pi}{N} \left(v + \frac{1}{2}\right), \quad (24)$$

$$L_s(u, v) \triangleq \sum_{k=0}^{N-1} \sin \frac{k\pi}{N} \left(u + \frac{1}{2}\right) \sin \frac{k\pi}{N} \left(v + \frac{1}{2}\right), \quad (25)$$

then we can show that

$$L_c(u, v) = \frac{1}{2} [\xi(u-v) + \xi(u+v+1)], \quad (26)$$

$$L_s(u, v) = \frac{1}{2} [\xi(u-v) - \xi(u+v+1)], \quad (27)$$

where

$$\xi(x) \triangleq \sum_{k=0}^{N-1} \cos\left(\frac{k\pi}{N}x\right) = \frac{1}{2} [1 - \cos \pi x + \sin \pi x \cdot \frac{\cos \frac{\pi x}{2N}}{\sin \frac{\pi x}{2N}}]. \quad (28)$$

If $\frac{(\pi x)}{(2N)}$ is so small that the second and higher order terms of $\frac{(\pi x)}{(2N)}$ can be ignored, then $\cos \frac{\pi x}{2N} \approx 1$, $\sin \frac{\pi x}{2N} \approx \frac{\pi x}{2N}$. Thus,

$$\xi(x) \approx \frac{1}{2} [1 - \cos \pi x] + N \text{sinc}(x), \quad (29)$$

where $\text{sinc}(x) \triangleq \sin(\pi x)/(\pi x)$. For large N , $\xi(x)$ is approximately a sinc function whose largest peak can be identified

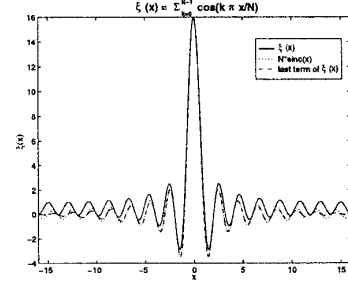


Figure 2. Plot of $\xi(x) = \sum_{k=0}^{N-1} \cos\left(\frac{k\pi}{N}x\right)$ for $N = 16$.

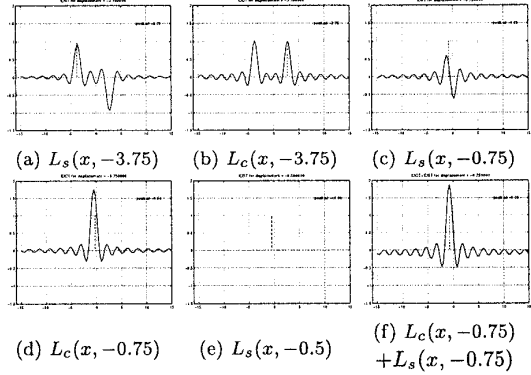


Figure 3. Illustration of sinusoidal orthogonal principles at the subpixel level for different displacements.

easily at $x = 0$ as depicted in Fig. 2, where $\xi(x)$ closely resembles $N \cdot \text{sinc}(x)$, especially when x is small.

Equations (26)-(27) are the equivalent form of the sinusoidal orthogonal principles (22)-(23) at the subpixel level. The sinc functions at the right hand side of the equations are the direct result of the rectangular window inherent in the DCT transform [6]. Fig. 3 (a) and (b) illustrate $L_s(x, -3.75)$ and $L_c(x, -3.75)$ respectively where two ξ functions are interacting with each other but their peak positions clearly indicate the displacement. However, when the displacement v is small (in the neighborhood of -0.5), $\xi(u-v)$ and $\xi(u+v+1)$ move close together and addition/subtraction of $\xi(u-v)$ and $\xi(u+v+1)$ changes the shape of L_s and L_c . As a result, neither L_s nor L_c looks like two ξ functions and the peak positions of L_s and L_c are different from those of $\xi(u-v)$ and $\xi(u+v+1)$, as demonstrated in Fig. 3 (c) and (d) respectively where the peak positions of $L_s(x, -0.75)$ and $L_c(x, -0.75)$ are -1.25 and -0.5 , differing from the true displacement -0.75 . In the extreme case, $\xi(u-v)$ and $\xi(u+v+1)$ cancel out each other when the displacement is -0.5 such that $L_s(x, -0.5) \equiv 0$ as shown in Fig. 3(e).

Fortunately we can eliminate the adverse interaction of the two ξ functions by simply adding L_c and L_s together since $L_c(x, v) + L_s(x, v) = \xi(x-v)$ as depicted in Fig. 3(f) where the sum $L_c(x, -0.75) + L_s(x, -0.75)$ behaves like a sinc function and its peak position coincides with the displacement. Furthermore, due to the sharpness of this ξ function, we can accurately pinpoint the peak position under a noisy situation and in turn determine the motion estimate. This property enables us to devise flexible and scalable subpixel motion estimation algorithms in the sub-sections.

4. SUBPIXEL MOTION ESTIMATION

The DCT-based half-pel motion estimation algorithm (HDXT-ME) requires two steps [7]: (1) The integer component of the displacement is estimated as (\hat{m}_u, \hat{m}_v) by the DXT-ME algorithm. (2) The pseudo phases $g_{\lambda_u, \lambda_v}^{CS}$, $g_{\lambda_u, \lambda_v}^{SC}$ from DXT-ME are used to produce $\overline{DCS}(u, v)$ and $\overline{DSC}(u, v)$ for $u \in \{\hat{m}_u - 0.5, \hat{m}_u, \hat{m}_u + 0.5\}$ and $v \in \{\hat{m}_v - 0.5, \hat{m}_v, \hat{m}_v + 0.5\}$ where

$$\overline{DCS}(u, v) = \sum_{k=0}^{N-1} \sum_{l=1}^{N-1} g_{\lambda_u, \lambda_v}^{CS}(k, l) \cos \frac{k\pi}{N} \left(u + \frac{1}{2}\right) \sin \frac{l\pi}{N} \left(v + \frac{1}{2}\right),$$

$$\overline{DSC}(u, v) = \sum_{k=1}^{N-1} \sum_{l=0}^{N-1} g_{\lambda_u, \lambda_v}^{SC}(k, l) \sin \frac{k\pi}{N} \left(u + \frac{1}{2}\right) \cos \frac{l\pi}{N} \left(v + \frac{1}{2}\right).$$

The peak position of either \overline{DCS} or \overline{DSC} determines the motion estimate $(\hat{\lambda}_u, \hat{\lambda}_v)$. The computational complexity of this algorithm is $O(N^2)$.

Similarly, the quarter-pel motion estimation algorithm (QDXT-ME) follows the same procedure except that $\overline{DCS}(u, v)$ and $\overline{DSC}(u, v)$ are computed at $u = \hat{m}_1 + 0.25i$ and $v = \hat{m}_1 + 0.25i$ for $i = -3, \dots, 3$ and the peak of $\overline{DCS} + \overline{DSC}$ is searched instead since this sum annihilates $\xi(m - \lambda_u)\xi(n + \lambda_v + 1)$ and $\xi(m + \lambda_u + 1)\xi(n - \lambda_v)$ to reduce the interaction effect of ξ functions on estimation precision [7]. However, if \overline{DCC} and \overline{DSS} are also available, then $D_4(u, v) \triangleq \overline{DCC}(u, v) + \overline{DCS}(u, v) + \overline{DSC}(u, v) + \overline{DSS}(u, v) = \xi(u - m_1 - \nu_u)\xi(v - n_1 - \nu_v)$. This leads to another quarter-pel algorithm (Q4DXT-ME) which searches the peak of D_4 to estimate motion [7]. As a matter of fact, the four functions in computing $D_4(u, v)$ can be generated naturally at the same time using the computing algorithms and architectures in [8].

5. SIMULATION

Simulation is run on two video sequences of different characteristics: Small Flower Garden (HFG) containing fast moving scene and Miss America (MS) with slow head and shoulder movement accompanying with occasional eye and mouth opening. The performance is compared with Full Search Block Matching Algorithm (BKM-ME) and its subpixel versions, Full Search Half-Pel Block Matching Algorithm (FHBKM-ME) and Full Search Quarter-Pel Block Matching Algorithm (FQBKM-ME), in terms of mean square error per pixel (MSE) defined as $MSE = \frac{1}{N^2} \sum_{m,n} [\hat{x}(m, n) - x(m, n)]^2$ where $\hat{x}(m, n)$ is the reconstructed image predicted from the original image $x(m, n)$ based upon the estimated displacement vector $\hat{\lambda} = (\hat{\lambda}_u, \hat{\lambda}_v)$. For all the MSE values computed in the experiment, bilinear interpolation is used in reconstructing displaced images as specified in MPEG for MPEG-compliant decoders. The simulation results in Fig. 4 and 5 show similar performance of the proposed algorithms to that of block matching approaches.

6. CONCLUSION

We develop the DCT-based subpixel motion estimation techniques in this paper and devise the HDXT-ME, QDXT-ME, and Q4DXT-ME algorithms to estimate motion with half-pel and quarter-pel accuracy without interpolation for video coding. This results in significant savings in hardware / complexity for interpolation and far less data flow compared to the conventional block matching methods on

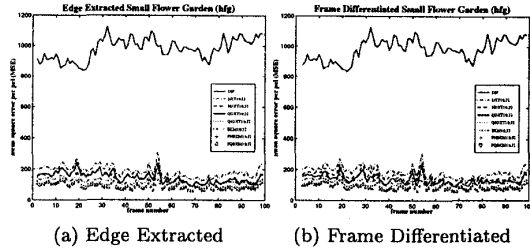


Figure 4. Simulation Results on HFG sequence

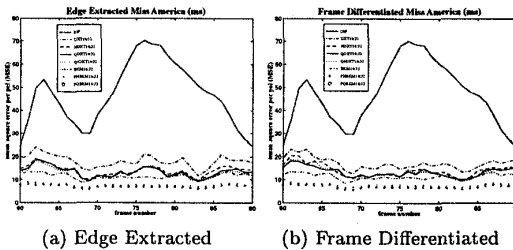


Figure 5. Simulation Results on MS sequence

interpolated images. Therefore, the proposed algorithms are more suitable for VLSI implementation. Furthermore, it avoids the deterioration of the estimate accuracy due to interpolation required in most current subpixel motion estimation schemes. In addition, the proposed approaches are scalable in the sense that higher estimation accuracy can be provided easily by applying the same subpel sinusoidal orthogonal principles without re-computation of pseudo phases. Therefore, flexible fully DCT-based codec design is possible because the same hardware can support different levels of required accuracy. Meanwhile, their computational complexity is only $O(N^2)$ compared to $O(N^4)$ for block matching approaches. Furthermore, computation of pseudo-phases is local and thus a highly parallel architecture design is feasible. Finally, the algorithms are DCT-based, enabling us to build a low-complexity, high-throughput fully DCT-based video coder.

REFERENCES

- [1] H. Li, A. Lundmark, and R. Forchheimer, "Image sequence coding at very low bitrates: A review", *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 589-608, September 1994.
- [2] U.-V. Koc and K. J. R. Liu, "DCT-based motion estimation", Tech. Rep. T.R.95-1, Institute for Systems Research, University of Maryland at College Park, 1995.
- [3] B. Girod, "Motion compensation: Visual aspects, accuracy, and fundamental limits", in *Motion Analysis and Image Sequence Processing*, M. I. Sezan and R. L. Lagendijk, Eds., chapter 5. Kluwer Academic Publishers, 1993.
- [4] B. Girod, "Motion-compensating prediction with fractional-pel accuracy", *IEEE Trans. Communications*, vol. 41, no. 4, pp. 604, April 1993.
- [5] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, and Applications*, Academic Press, Inc., 1990.
- [6] A. Papoulis, *Signal Analysis*, McGraw-Hill, Inc., 1977.
- [7] U. V. Koc and K. J. R. Liu, "Exact subpixel motion estimation in DCT domain", Tech. Rep., Institute for Systems Research, University of Maryland at College Park, 1995.
- [8] K. J. R. Liu and C. T. Chiu, "Unified parallel lattice structures for time-recursive Discrete Cosine/Sine/Hartley transforms", *IEEE Trans. Signal Processing*, vol. 41, no. 3, pp. 1357-1377, March 1993.