ADAPTIVE OVERLAPPING APPROACH FOR DCT-BASED MOTION ESTIMATION*

Ut-Va Koc and *K. J. Ray Liu*

Electrical Engineering Department and Institute for Systems Research
University of Maryland at College Park
koc@eng.umd.edu and *kjrliu@eng.umd.edu*

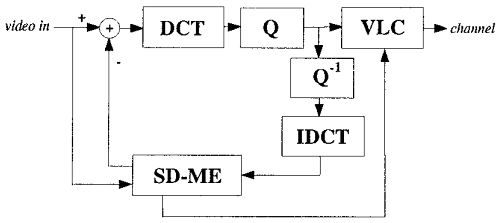
ABSTRACT

The feedback loop for temporal prediction of traditional implementations of an MPEG-compliant video coder requires conversion of images from the spatial domain to the transform domain and then back to the spatial domain due to the restriction of motion estimation schemes performed only in the spatial domain, severely limiting the throughput of the coder. However, availability of the low-complexity DCT-based motion estimation scheme (DXT-ME) that we proposed in the past can remove this bottleneck and concurrently reduce the complexity of the overall coder structure. In this paper, we adopt an adaptive overlapping approach to further improve this DCT-based motion estimation scheme, as demonstrated in the simulation results. Furthermore, the analysis of stability of DXT-ME will be presented and the case of a uniformly bright environment will also be discussed. Finally, the parallel architecture of DXT-ME will be outlined in this paper.

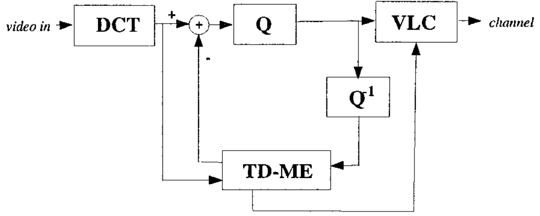
1. INTRODUCTION

In most international video coding standards such as CCITT H.261 [1] , MPEG [2] as well as the proposed HDTV standard, Discrete Cosine Transform (DCT) and block-based motion estimation are the essential elements to achieve spatial and temporal compression, respectively. Most implementations of a standard-compliant coder adopt the structure of Coder III (originally named in [3]) as shown in Fig. 1(a). The feedback loop for temporal prediction consists of a DCT, an Inverse DCT (IDCT) and a spatial-domain motion estimator (SD-ME) which is usually the Full Search Block Matching Approach (BMA-ME). This is undesirable. In addition to the additional complexity added to the overall architecture, this feedback loop limits the throughput of the coder and becomes the bottleneck of a real-time high-end video codec. A compromise is to remove the loop and perform open-loop motion estimation based upon original images instead of reconstructed images in sacrifice of the performance of the coder [4]. An alternative solution without degradation of the performance is to develop a motion estimation algorithm which can work in the DCT transform domain as remarked in [3]. In this way, the DCT and IDCT can be moved out of the loop as depicted in Fig. 1(b) to significantly increase the coder throughput. and at the same time reduces the complexity of the overall coder. Furthermore, different components can be jointly optimized if they operate in the same transform domain.

For the realization of the Coder II architecture to boost the system throughput and reduce the total number of components, a new DCT-based motion estimation algorithm, DXT-ME, was developed [5] to estimate motion in the DCT domain with very low computational complexity $O(N^2)$ as opposed to $O(N^4)$, the complexity of BMA-ME. In this paper, we present an adaptive overlapping approach that dynamically adjust the search area. The proposed scheme significantly improves the performance over the results in [5] and even almost as good as the results obtained by BMA-ME. We will also discuss the stability of the DXT-ME and the case of uniformly bright background. Finally, a parallel architecture that can efficiently implement the proposed scheme will be presented.



(a) Coder III



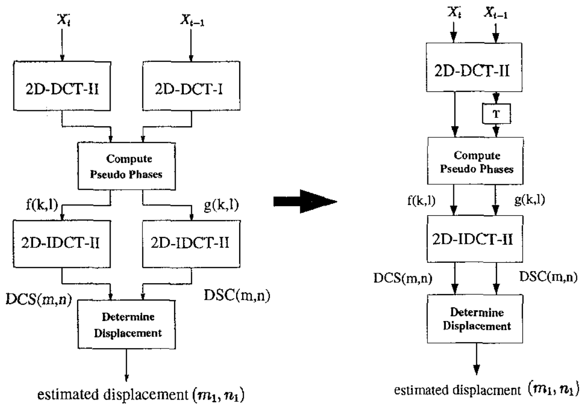
(b) Coder II

Figure 1: Coder structures: (a) Coder III is the motion-compensated DCT hybrid coder used in MPEG or H.261 standards with motion estimation done in the spatial domain. (b) Coder II is the motion-compensated DCT hybrid coder with motion estimation performed in the transform domain.

---

*This work is supported in part by the ONR grant N00014-93-1-0566, NSF grant MIP9309506, and MIPS/MicroStar.

0-8186-7310-9/95 $4.00 © 1995 IEEE



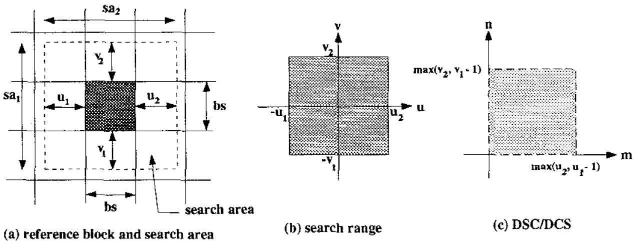
(a) flowchart (b) structure

Figure 2: Block diagram of DXT-ME

2. ADAPTIVE OVERLAPPING APPROACH FOR DXT-ME

As mentioned in [5], the procedure of the DXT-ME algorithm depicted in Fig. 2(a) can be summarized as follows: (1) Compute 2D DCT-II and DCT-I of the previous and current blocks respectively. (2) Compute the pseudo phases, $f(k,l)$ and $g(k,l)$ inherent in the DCT coefficients. (3) Perform an inverse 2D DCT-II on the pseudo phases to obtain $DCS(m,n)$ and $DSC(m,n)$. (4) Search for peak indices to determine estimated displacement vectors $(\hat{m}_1, \hat{n}_1)$. However, the computation of 2D DCT-I can be replaced by Block T, a coefficient transformation unit realizing the point-to-point relationship between 2D-DCT-I and 2D-DCT-II coefficients in the frequency domain, as shown in Fig. 2(b). To deal with complicated background, a preprocessing step is added to improve the performance and this extended algorithm is called Simplified Extended DXT-ME (SE-DXT-ME).

In the above procedure, the peaks of $DSC(m,n)$ and $DCS(m,n)$ are searched over a fixed index range of interest $\Phi = \{0, \dots, N/2\}^2$. However, if we follow the partitioning approach used in BMA-ME, then we may dynamically adjust Φ. First, partition the whole current frame into $bs \times bs$ non-overlapping reference blocks shown as the shaded area in Fig. 3(a). Each reference block is associated with a larger search area (of size sa) in the previous frame (the dotted region in the same figure) in the same way as for BMA-ME. From the position of a reference block and its associated search area, a search range $\mathcal{D} = \{(u,v) : -u_1 \le u \le u_2,\ -v_1 \le v \le v_2\}$ can then be determined as in Fig. 3(b). Differing from BMA-ME, DXT-ME requires that the reference block size and the search area size must be equal. Thus, instead of using the reference block, we use the block of the same size and position in the current frame as the search area of the previous frame. The peak values of DSC and DCS are then searched in a zigzag way over this index range $\Phi = \{0, \dots, \max(u_2, u_1 - 1)\} \times \{0, \dots, \max(v_2, v_1 - 1)\}$. In addition to the requirement that the new peak value must be larger than the current peak value by a preset threshold, it is necessary to examine if the motion estimate determined by the new peak index lies in the search region $\mathcal{D}$. Since search areas overlap on one another, the SE-DXT-ME architecture utilizing this approach is called Overlapping SE-DXT-ME (OSE-DXT-ME).



(a) reference block and search area (b) search range (c) DSC/DCS

Figure 3: Adaptive overlapping approach

3. SIMULATION RESULTS OF OVERLAPPING SE-DXT-ME

The simulation results on the small "Flower Garden" sequence are plotted in Fig. 5 for Overlapping SE-DXT-ME with either frame differentiation or edge extraction as the preprocessing function. As can be seen, the MSE values for Overlapping SE-DXT-ME of different block sizes are as small as the MSE values for BMA-ME. As shown in Fig. 6, similar results are obtained on the "Miss America" sequence which has the motion of mouth and eyes notably difficult to be detected. The performance of Overlapping SE-DXT-ME is comparable to that of BMA-ME while its computational complexity is much lower.

4. STABILITY OF DXT-ME AND NON-DARKEN BACKGROUND

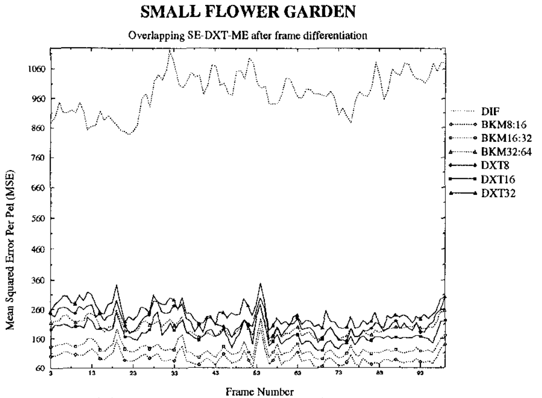
The stability of this motion estimator depends upon the property of the determinant of the system matrix $|\mathbf{Z}_{t-1}(k,l)|$ used as the denominator in the computation of the pseudo phase functions $f(k,l)$ and $g(k,l)$ for $k,l = 1, \dots, N-1$:

$$|\mathbf{Z}_{t-1}(k,l)| = \begin{vmatrix} Z^{cc}_{t-1}(k,l) & -Z^{cs}_{t-1}(k,l) & -Z^{sc}_{t-1}(k,l) & Z^{ss}_{t-1}(k,l) \\ Z^{cs}_{t-1}(k,l) & Z^{cc}_{t-1}(k,l) & -Z^{ss}_{t-1}(k,l) & -Z^{sc}_{t-1}(k,l) \\ Z^{sc}_{t-1}(k,l) & -Z^{ss}_{t-1}(k,l) & Z^{cc}_{t-1}(k,l) & -Z^{cs}_{t-1}(k,l) \\ Z^{ss}_{t-1}(k,l) & Z^{sc}_{t-1}(k,l) & Z^{cs}_{t-1}(k,l) & Z^{cc}_{t-1}(k,l) \end{vmatrix}$$

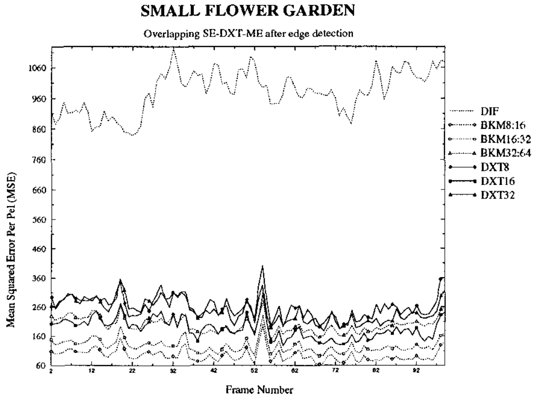
A zero or near-zero value of $|\mathbf{Z}_{t-1}(k,l)|$ may jeopardize the performance. However, it can be proved analytically that

Stage	Component	Computational Complexity
1	2D-DCT-II	$O_{dct} = O(N)$
	Coeff. Transformation Unit (T)	$O(N^2)$
2	Pseudo Phase Computation	$O(N^2)$
3	2D-IDCT-II	$O_{dct} = O(N)$
4	Peak Searching	$O(N^2)$
	Estimation	$O(1)$

Figure 4: Computational complexity of each stage in DXT-ME

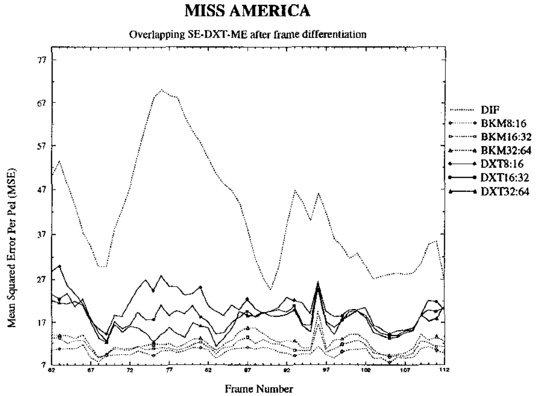


(a) Frame differentiation for preprocessing

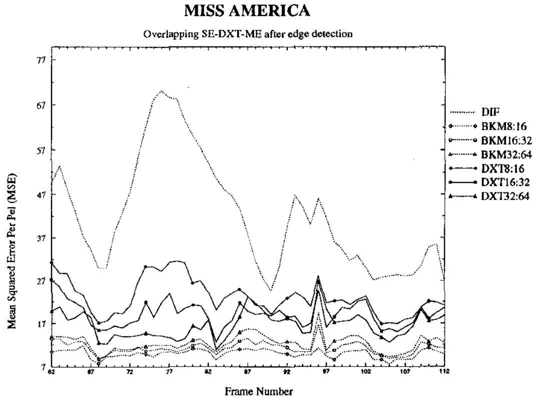


(b) Edge extraction for preprocessing

Figure 5: Comparison of Overlapping SE-DXT-ME with BMA-ME on small "Flower Garden"



(a) Frame differentiation for preprocessing



(b) Edge extraction for preprocessing

Figure 6: Comparison of Overlapping SE-DXT-ME with BMA-ME on "Miss America"

this determinant is zero only when either

$$\sum_{m,n=0}^{N-1} x_{t-1}(m,n)\sin[\frac{\pi}{N}(km \mp ln)] = 0$$

or

$$\sum_{m,n=0}^{N-1} x_{t-1}(m,n)\cos[\frac{\pi}{N}(km \mp ln)] = 0.$$

This in turn requires $x_{t-1}(m,n) \equiv 0$. Therefore it is very unlikely that this determinant is zero and as such, the DXT-ME is a stable estimator. Even so, if $\mathbf{Z}_{t-1}(k,l) = 0$ really happens due to limitation of machine precision or $\mathbf{Z}_{t-1}(k,l)$ is less than a threshold, then we can let $f(k,l) = g(k,l) = 1$, which is equivalent to the situation when $x_{t-1}(m,n) \equiv 0$. In this way, the catastrophical effect of computational precision of a certain implementation on the stability of DXT-ME will be kept to minimum.

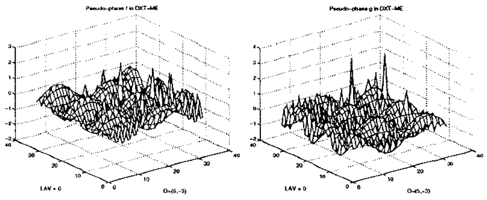
What if an object is moving in a uniformly bright background instead of a completely dark environment? It can be shown analytically and empirically that uniformly bright background introduces only very small spikes which does not affect the accuracy of the estimate. Concisely, in addition to a large peak $\vec{\delta}_{m_1n_1}(\cdot,\cdot)$ whose position is used to estimate the motion vector, several small spikes and a noise term $\vec{\mathbf{n}}(\cdot,\cdot)$ will appear in the 2D-IDCT-II domain of the pseudo phase function in a uniformly bright background with background pixel value c:

$$\begin{aligned}\vec{\mathbf{d}}(m,n) &= \vec{\delta}_{m_1n_1}(m,n) + \vec{\mathbf{n}}(m,n) \\ &+ \frac{c}{4(s+c)}[\vec{\mathbf{E}}_{00}(m,n) - \vec{\mathbf{E}}_{m_1n_1}(m,n)] \quad (1)\end{aligned}$$

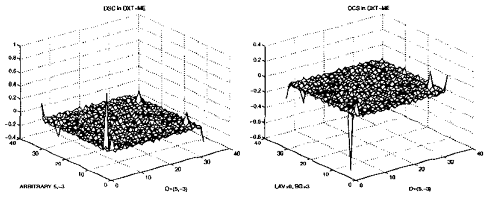
where $\vec{\mathbf{E}}_{00}(\cdot,\cdot)$ and $\vec{\mathbf{E}}_{m_1n_1}(\cdot,\cdot)$ contain four small spikes each, and s is a constant depending on the relative brightness of a moving object to the background. The noise term $\vec{\mathbf{n}}$ for the case of $c = 3$ in Fig. 7(b) is observable but very small and can be regarded as pure noise whereas $\vec{\mathbf{n}}$ is practically absent as in Fig. 7(c) when $c = 255$. However, eight very small spikes are present in both cases.

5. PARALLEL ARCHITECTURE OF DXT-ME

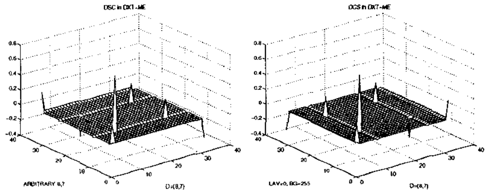
For a DCT unit of computational complexity O_{dct}, the overall complexity of DXT-ME is $O(N^2) + O_{dct}$ with the complexity of each component summarized in Table 4. The computational complexity of the pseudo phase computation component is only $O(N^2)$ for an $N \times N$ block and so is the unit to determine the displacement. For the computation of the pseudo phase functions $f(\cdot,\cdot)$ and $g(\cdot,\cdot)$ which can be found in [5, 6], DSCT, DCST and DSST coefficients (regarded as DST coefficients) must be calculated in addition to DCCT coefficients (i.e. the usual 2D DCT). However all



(a) f and g



(b) DSC and DCS



(c) another DSC/DCS

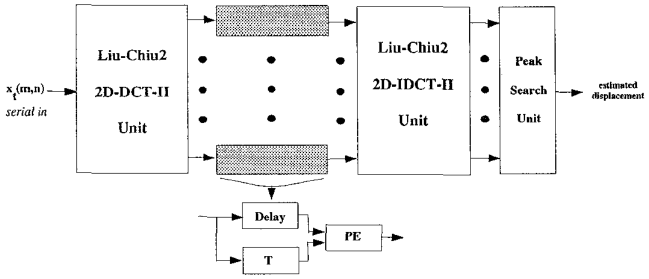
Figure 7: (a)(b) An object is moving in the direction (5, -3) in a uniformly bright background ($c = 3$). (c) Another object is moving northeast (8,7) for background pixel values $= c = 255$.

these coefficients can be generated with little overhead in the course of computing 2D DCT coefficients. As a matter of fact, a parallel and fully-pipelined 2D DCT (Liu-Chiu2) lattice structure has been developed [7, 8, 9] to generate 2D DCT coefficients at a cost of $O(N)$ operations. This DCT structure computes DCT and DST coefficients dually due to its internal lattice architecture. These internally generated DST coefficients can be output to the module for pseudo phase computation. This same lattice structure can also be modified as a 2D IDCT which also has $O(N)$ complexity. In conclusion, the overall complexity of this DXT-ME is only $O(N^2)$, much lower than the $O(N^4)$ complexity of BMA-ME.

Due to its intrinsic highly parallel nature, its structure can be mapped to the parallel architecture as shown in Fig. 8 where PE is a processing element to compute the pseudo-phases $f(k,l)$ and $g(k,l)$ at (k,l) and the DCT and IDCT units adopt the Liu-Chiu2 lattice structure. In this way, a modular and parallel architecture can be implemented to achieve high throughput.

6. CONCLUSION

Our simulation results have shown that the adaptive DCT-based motion estimation algorithm, called *Overlapping SE-DXT-ME,* can achieve low MSE values comparable to the BMA-ME approach which is optimized in terms of MSE values. On the other hand, this DCT-based motion estimation scheme has very low complexity $O(N^2)$ in contrast



PE computes pseudo-phases at (k, l)

Figure 8: Parallel architecture of DXT-ME

to the notoriously high complexity $O(N^4)$ of BMA-ME. In addition, motion estimation in the DCT domain simplifies the feedback loop and thus significantly improves the coder throughput and reduces the overall complexity as well. Further analysis of this algorithm ensures its stability and robustness to motion in uniformly bright background. Finally the computation of pseudo phases element by element suggests the feasibility of a highly parallel implementation of this scheme. Future research may focus on hardware implementation and integration with other processing units to form a video coder of high throughput and low complexity for HDTV and multimedia applications.

REFERENCES

[1] CCITT Recommendation H.261, *Video Codec for Audiovisual Services at p × 64 kbit/s*, CCITT, August 1990.

[2] CCITT Recommendation MPEG-1, *Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s*, ISO/IEC 11172, Geneve Switzerland, 1993.

[3] 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.

[4] J. S. McVeigh and S.-W. Wu, "Comparative study of partial closed-loop versus open-loop motion estimation for coding of HDTV", in *Proc. IEEE Workshop on Visual Signal Processing and Communications*, New Brunswick, September 1994, pp. 63–68.

[5] U. V. Koc and K. J. R. Liu, "Discrete-Cosine/Sine-Transform based motion estimation", in *Proceedings of IEEE International Conference on Image Processing (ICIP-94)*, Austin, Texas, November 1994, vol. 3, pp. 771–775.

[6] 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.

[7] C. T. Chiu and K. J. R. Liu, "Real-time parallel and fully pipelined two-dimensional DCT lattice structures with applications to HDTV systems", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 2, no. 1, pp. 25–37, March 1992.

[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.

[9] K. J. R. Liu, C. T. Chiu, R. K. Kologotla, and J. F. JaJa, "Optimal unified architectures for the real-time computation of time-recursive Discrete Sinusoidal Transforms", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 4, no. 2, pp. 168–180, April 1994.