# A UNIFORM TRANSFORM DOMAIN VIDEO CODEC BASED ON DUAL TREE COMPLEX WAVELET TRANSFORM

*Kamakshi Sivaramakrishnan* †*and Truong Nguyen* ‡

Boston University
Electrical and Computer Engineering Department
Boston MA 02215
†kamakshi@bu.edu, ‡nguyent@engc.bu.edu

## ABSTRACT

This paper describes a uniform transform domain Video Codec where the motion estimation/compensation (ME) is performed in the transform domain. The estimation technique discussed here is a subpixel transform domain ME based on the Dual Tree Complex Wavelet Transform (DT CWT) and a maximum phase correlation technique. The DT CWT is a multiresolution fine-to-coarse bandpass filtered decomposition of each still frame and has desirable properties of ***shiftablility***, directional selectivity and perfect reconstruction (PR). Estimation is first performed at the finest resolution and successively proceeds to the coarse resolutions using a fine-to-coarse strategy. This gives multiresolution motion estimates enabling estimation of current frame transform coefficients from the corresponding ones in previous frame. The key difference of this approach is the transform domain error frames - a uniform transform domain Video Codec. This further simplifies the encoder and decoder resulting in computational savings with comparable performance to the standards.

## 1. INTRODUCTION

The intensity-based block matching (BKM) ME techniques are a good approximation of the underlying 3-D motion. However the ambiguity of the very measure of motion (pattern of intensity changes) may result in an erroneous representation of the 2-D projected motion. The conventional hybrid video coding techniques like the MPEG and H.263 comprise of a spatial domain ME (SD-ME) block which instates a heavily loaded feedback loop in the encoder shown in Fig.1(a). The Inverse transform ($\mathrm{T}^{-1}$), is solely for the purpose of spatial domain ME and has been long recognized as a major bottleneck of the video coding system for high speed real-time video networks. A possible solution to this problem is to perform the motion estimation in the transform domain (TD-ME), thus moving the transform (T) block out of the loop and removing the ($\mathrm{T}^{-1}$) block. This results in a uniform transform-domain Video codec circumventing the shortcomings in the conventional codec. In this paper, we propose a modified transform domain Video codec as shown in the Fig.1 (b).

An approach to the problem of transform domain ME is the phase-matching technique, which is feasible if local displacements in the spatial domain can be approximated by the linear phase-shifts in the transform domain. Adelson and Bergen et. al [1] designed a bank of spatio-temporal filters that decomposed the signal into channels that are tuned to different speeds and orientations and hence representative of the underlying motion components in the signal. This gives rise to the phase-based optical flow constraint

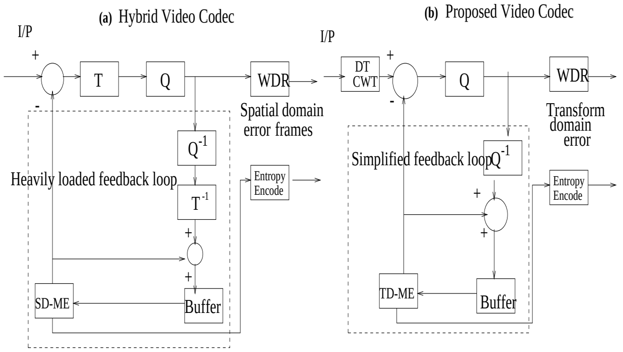


**Fig. 1**. Block diagram of (a) conventional and (b) Proposed Video encoder

$$(\nabla\phi_i)^T\mathbf{v} + \frac{\partial\phi_i}{\partial t} = 0 \tag{1}$$

where $\phi_i$ is the output (transform coefficients) of the channel $i$ and $\mathbf{v}$ is the corresponding component of the motion (displacement) vector. This idea was exploited by Magaurey et.al. [2] in a multiresolution coarse-to-fine strategy to increase the range of estimation. Their technique however suffers from the deficiencies of a coarse-to-fine ME strategy, viz:-

- Inaccurate ME at the coarsest resolution, due to the lack of detail and aliasing effects.
- ME performed using coded reference frames involves quantization noise which affects coarse resolutions the most.
- Finite extent subband filters introduce aliasing components at coarser resolutions preventing resolution scaling of translational motion.
- Not applicable for a complete uniform Video Codec as the error frame of the estimation is still in the

spatial domain.

Also, the transform used in [2], is a complex-valued transform and hence not compatible with standards. These problems have been addressed in the fine-to-coarse DT CWT-based ME technique that we describe in this paper.

## 2. DUAL TREE COMPLEX WAVELET TRANSFORM (DT CWT)

In order to perform estimation in the transform domain, we need a "shift-invariant" transform. This eliminates the real Discrete Wavelet Transform (DWT) which does not exhibit shift-invariance as it violates the Nyquist sampling rate [3] and information moves from one subband to another under translation. It has been shown by Simoncelli et.al [3] that there must be a relaxation of the critical sampling condition to achieve approximate shift-invariance resulting in abandoning orthogonality. Kingsbury et.al in [4], has designed an approximately shift-invariant implementation of the wavelet transform, viz:- Dual Tree Complex Wavelet Transform. It has the added advantage of perfect reconstruction (PR) and directional selectivity. The PR property is very important for the ME technique suggested in this paper. The transform and its shiftability is briefly discussed in this section, the details of which can be found in [4], [3].

Approximate shift invariance is possible with a real DWT by "doubling" the sampling rate at each level of the tree. "Doubling" can be achieved by eliminating the downsampling operator after the first level of decomposition provided the samples are uniformly spaced. This is equivalent to two fully-decimated trees (two real DWTs) provided the filters in the two trees meet the delay criterion to maintain uniform intervals between samples. The dual tree of real filters is shown in the Fig. 2.

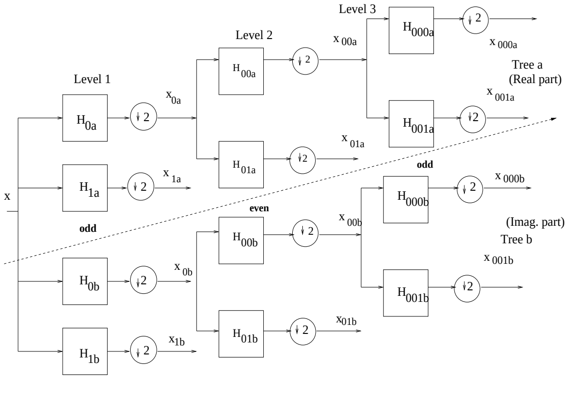


**Fig. 2**. Dual tree of real filters for the DT CWT, filters of even and odd length alternate at successive levels

The uniform sampling condition alongwith linear phase requirements impose conditions [4] on the length of the filters in the two trees. Greater symmetry considerations result in the alternating even and odd length filters in DT CWT. The filter responses are Gausssian-shaped resembling the real and imaginary parts of the complex filters in [2] and are shown in Fig. 3

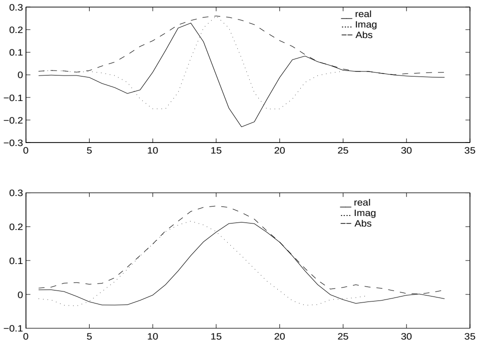


**Fig. 3**. Impulse responses at level 4 of the DT CWT scaling and wavelet function.

The transform is complex if the output from tree a is considered to be the real part and that from tree b, the imaginary part of the transform coefficient. Alternatively it could be considered to be a limited redundancy oversampled real transform.

### 2.1. Shiftability of CWT

Simoncelli et.al in [3] have shown that a transform is shiftable if and only if there exists a set of interpolation functions that *interpolate* the "non-translated" transform coefficients to give the translated coefficients for any arbitrary translation $x_0$ of the spatial domain input signal. This condition is summarized here for a periodic input signal, $f(x)$. The signal $f(x)$ when transformed using the projection functions corresponding to shifted copies (basis functions) of the kernel, $h(x)$ :

$$\{h(n - \Delta_x - x) | n = 0, 1, \ldots, N-1\}$$

over a period $[0, 2\pi]$ and sampling interval, $\Delta_x = \frac{2\pi}{N}$ gives transform coefficients, $C(n)$. Thus,

$$C(n) = \int_0^{2\pi} dx h(n\Delta_x - x) f(x), \; n \in 0, 1, \ldots, N-1 \qquad (2)$$

Simoncelli et.al in [3], have shown that the above transformation is shiftable if

$$h(x_0 - x) = \Sigma_{n=0}^{N-1} b_n(x_0) h(n\Delta_x - x) \qquad (3)$$

where $x_0$ is the arbitrary shift, $b_n(x_0)$ is the interpolant.

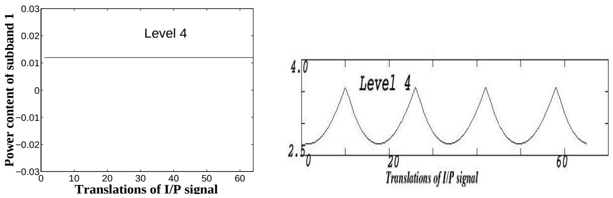
That is, the arbitrary shifted kernel $h(x_0 - x)$ can be expressed as a linear combination of the basis functions $h(x - n\Delta_x)$. This implies that the sampled basis set spans the entire subspace of all translations of the kernel. Equivalently, the linear subspace spanned by the sample basis is invariant to translation.

Using equation (2) in equation (3), we get

$$C(n_0 - n) = \Sigma_{k=0}^{N-1} b_k(n_0 * \frac{2\pi}{N}) C(k - n) \qquad (4)$$

where $n_0 = x_0 * \frac{N}{2\pi}$ is the corresponding shift in the transform domain.

If a solution to the above equation exists, then the transform is shiftable. In an aliased transform (real DWT), the response power depends on the signal position: translation of the input signal generally results in a re-distribution of the power content amongst the various frequency subbands. The shiftability constraint is equal to the fact that power of the transform coefficients in the subband is preserved when the input signal is shifted in position [3]. This is demonstrated by the DT CWT while the real DWT exhibits a periodic variation in the power content of the transform coefficients with translations of the input signal and is shown in the Fig. 4(a) and (b). This constraint is used in the approximation of the spatial motion to a linear phase change, giving a closed form expression for the motion vector estimate.



(a) (b)

**Fig. 4**. (a) Constant power of DT CWT coefficients (b) Periodic power variation of Real DWT coefficients v/s translation of the spatial domain signal

*Thus, the DT CWT gives us a "real-valued" transform that is approximately shift-invariant !*

## 3. FINE-TO-COARSE MOTION ESTIMATION

The solution to equation (4) is the key to the problem of ME in transform domain and more importantly, it is necessary that we look for a computationally inexpensive interpolant, $b_n(x_0)$. The zero-order interpolation technique discussed by Magaurey et.al [2] describes a simple single subpel interpolation technique which is good only for small range motion vectors, $\mathbf{n_0}$. Here, we describe a higher order interpolation technique that is more robust and it averages out the noise effects in the transform coefficients. The error measure is the subband squared difference ($SD$), similar to that in [2]

$$SD^{(n,m)}(\mathbf{n},\mathbf{n_0}) = |C_1^{(n,m)}(\mathbf{n}+\mathbf{n_0}) - C_2^{(n,m)}(\mathbf{n_0})|^2 \qquad (5)$$

where $C_1^{(n,m)}$ is the subpel of the reference frame (unshifted) at a level of decomposition $m$ and subband $n$ centered around the position $\mathbf{n} = (n_1, n_2)$ and $C_2^{(n,m)}$ is the subpel of the current frame (shifted) with similar attributes. Here, $\mathbf{n_0}$ is the translation in the transform domain corresponding to a motion in the spatial domain.

From equations (3) and (2)

$$C^{(n,m)}(\mathbf{n}+\mathbf{n_0}) = \Sigma_k B_{\mathbf{n_0}}^{(n,m)}(\mathbf{k}) C^{(n,m)}(\mathbf{n}+\mathbf{k}) \qquad (6)$$

where $B_{\mathbf{f}}^{(n,m)}$ is the interpolant at subband $n$ and resolution $m$ Because of the Gaussian-shaped impulse response of the CWT filters, Magaurey et.al showed that the interpolant can be expressed as

$$B_{\mathbf{n_0}}^{(n,m)}(\mathbf{k}) = K_{\mathbf{n_0}}(\mathbf{k}) e^{j2^m[\boldsymbol{\Omega}^{(n,m)}]^T(\mathbf{n_0}-\mathbf{k})} \qquad (7)$$

$K_{\mathbf{n_0}}(\mathbf{k})$ is the interpolating kernel and $\boldsymbol{\Omega}^{(n,m)}$ is the center frequency [2], along which the phase is a constant and equal to the orientation of the subband. If we weight the surrounding subpels by their phase contribution (measured by the inner product of the center frequency and position in the subband), we get a simple higher order and robust interpolation technique, i.e.,

$$K_{\mathbf{n_0}}^{(n,m)}(\mathbf{k}) \quad = \quad e^{j2^m[\Omega^{(n,m)}]^T(\mathbf{k})}$$

Now,

$$B_{\mathbf{n_0}}^{(n,m)}(\mathbf{k}) = e^{j2^m(\Omega^{(n,m)})^T\mathbf{k}}(\mathbf{k}) e^{j2^m[\Omega^{(n,m)}]^T(\mathbf{n_0}-\mathbf{k})} \qquad (8)$$

Equation (6) now becomes

$$C^{(n,m)}(\mathbf{n}+\mathbf{n_0}) \approx e^{j2^m(\Omega^{(n,m)})^T(\mathbf{n_0})}\Sigma_k C^{(n,m)}(\mathbf{n}+\mathbf{k}) \qquad (9)$$

The above interpolation formula and the constant subband power (section 2.1) is used in locating the minimum of the SSD surface (equation 5), $SD^{(m)}(\mathbf{n},\mathbf{n_0})$. Expanding equation(5) it can be shown that minimizing the SSD is equivalent to maximizing the phase correlation of the complex coefficients. Using the model for the interpolated phase (equation(9)), the motion estimate, $\mathbf{n_0}$ is expressed as a linear relation to the interpolated phase, $\theta^{(n,m)}(\mathbf{n})$

$$2^m(\Omega^{(n,m)})^T\mathbf{n_0} \quad = \quad \theta^{(n,m)}(\mathbf{n}) \qquad (10)$$

$$\text{where, } \theta^{(n,m)}(\mathbf{n}) \quad = \angle \frac{C_2^{(n,m)}(\mathbf{n})}{\Sigma_k C_1^{(n,m)}(\mathbf{n}+\mathbf{k})} \qquad (11)$$

In 2-D, since we have six subbands from the DT CWT decomposition, the motion estimate $\mathbf{n_0}$ corresponding to the subpel $\mathbf{n}$ at any resolution is obtained by averaging over all the six subbands and is given by

$$\mathbf{v} = \arg\min_{\mathbf{n}} \Sigma_{n=1}^{6} SD^{(n,m)}(\mathbf{n},\mathbf{n_0}) \qquad (12)$$

Using the linear phase model, a closed-form expression for the minimizer of $SD^{(n,m)}$ is obtained [5], [2]. The advantage of the phase model in equation (10) is that is not limited by the range of the motion vector $\mathbf{n_0}$. Hence we need not perform any further refinement to obtain a true minimum.

### 3.1. Algorithm Structure

The block diagram of the transform domain ME is shown in Fig. 5 which is self-explanatory in itself. The motion compensation is done using a wavelet based lowpass/bandpass interpolation [6] technique. As shown in Fig. 5, the motion compensation is performed using the motion vectors and the reference frame transform coefficients to generate the predicted frame transform coefficients. The transform

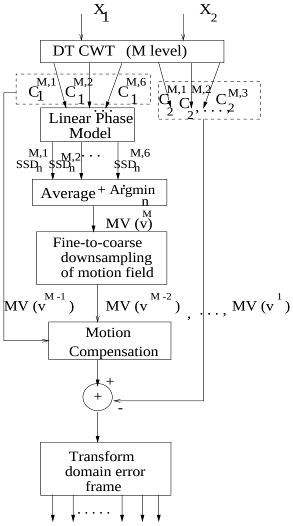


**Fig. 5**. Block Diagram of the proposed ME technique

domain error frames thus generated are coded using the Wavelet Difference Reduction [7] algorithm. *Thus, we have transform domain error frames and motion vectors (MV) of transform coefficients!*

## 4. RESULTS

The result of the proposed ME algorithm applied to the test sequence of 'Miss America' is shown in Fig. 6 as a motion field representation. The comparison of the proposed ME technique with standard BKM techniques and coarse-to-fine complex multiresolution ME (CDWT2) in [2] is shown in Fig. 7(a) and Fig. 7(b) show the PSNR of the reconstructed frames for various bit allocations of coding the error frame. The significant improvement in the SNRs of the reconstructed frames is attributed to the improved phase model (averaged contribution of the spatial neighbouring coefficients), the fine-to-coarse ME strategy and the higher order wavelet-based interpolation technique. Also, the fine-to-coarse ME strategy engenders a better estimation accuracy. The grouping of coefficients in the estimation of the motion vectors into blocks gives higher computational speed in addition to averaging out the noise effects in estimation. The wavelet-based bandpass interpolation technique gives a high subpixel accuracy in the order of dyadic fractions which is very useful in estimating fractional displacements of motion. Additionally, since TD-ME results in a true representation of the underlying motion which is independent of the intensity, light effects and the acquisition methods of capture of the video sequence. Thus, we conclude that we have an efficient uniform transform domain Video Codec with higher accuracy in reduction of temporal redundancy.

## 5. REFERENCES

[1] E.H. Adelson and J. R. Bergen, "Spatiotemporal energy models for the perception of motion," *J. Opt. Soc. Amer. A.*, vol. 2, pp. 284–299, 1985.

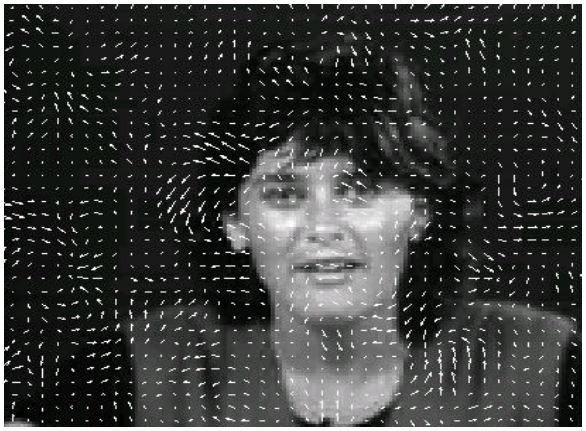


**Fig. 6**. Motion field of the "Miss America" sequence

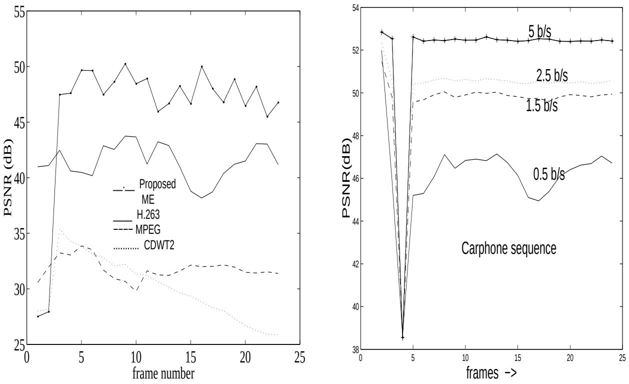


**Fig. 7**. (a) Comparison of the PSNR of the standards with proposed technique (b)PSNR of the reconstructed sequence at various bpp of coding the error frames

[2] Julian Magaurey and Nick Kingsbury, "Motion estimation using a complex-valued wavelet transform," *IEEE Trans on Signal Processing*, vol. 46, no. 4, pp. 1069–84, April 1998.

[3] E.P. Simoncelli, W.T. Freeman, E.H. Adelson, and David J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. on Information Theory*, vol. 38, no. 2, pp. 587–607, March 1992.

[4] Nick Kingsbury, "Image processing using complex wavelets," *Phil. Trans. Royal Society London A*, September 1999.

[5] Kamakshi Sivaramakrishnan, "A wavelet-based transform domain motion estimation technique for video compression," *Masters Thesis, Dept. of Electrical and Computer Engg., Boston University*, August 2000.

[6] R. A. Gopinath and C.S. Burrus, "Wavelet-based lowpass/bandpass interpolation," *Proc. of ICASSP '92*, vol. 4, pp. 384–388, March 1992.

[7] James S. Walker and Truong Nguyen, ," *Wavelet-Based Image Compression, Chapter of CRC Press book: CRC Handbook of Transforms and Data Compression (to be published).*