Iterative Thresholding-based Sparse Directional Representation for Efficient Low Bit-Rate Embedded Video Coding

Lingchen Zhu† Hongkai Xiong†
†Department of Electronic Engineering
Shanghai Jiao Tong Univ., Shanghai, 200240, China,
{extreme001, xionghongkai}@sjtu.edu.cn

Abstract

The contourlet transform provides a flexible directional image decomposition by employing Laplacian pyramid and uniform directional filter banks. Although it is able to efficiently capture the 2-D piecewise smooth functions with line or curve discontinuities, its major drawbacks of 4/3 redundancy and non-ideal filter banks in Laplacian pyramid are the main obstacles for high performance image/video coding. In this paper, we propose a video coding scheme based on the sharp frequency localized contourlet transform (contourlet/SFL) under the sparse representation framework, in which the iterative thresholding algorithm is applied to get an l_1 norm sparser version of the transform coefficients. In the meantime, considering the fact that strong inter-band and inter-scale dependencies exist in the contourlet/SFL coefficients, a directional embedded image coding system is proposed to propagate the significance status by using the neighbor, cousin and parent significant coefficients as seeds. Experimental results show that the coding performance and visual quality of the sparse contourlet/SFL based video compression scheme are superior to the wavelet-based one, especially on those sequences full of directional structures at low bit rates.

1 Introduction

The traditional video compression frameworks such as H.264/MPEG-4 AVC [1] are based on a hybrid temporal-spatial representation of video sequences. A typical video compression framework consists of several stages sequentially, e.g., after one block of frames are obtained, a motion estimation/compensation module is applied to decouple them into temporal subbands, then a spatial transformation is carried on each temporal subband to decompose the pixels into coefficients. The sparseness of coefficients in the transform domain determines its effectiveness in the following entropy coding stage. It is therefore crucial in the low bit-rate video coding since much of the acquired information would be discarded after encoding.

The discrete wavelet transform (DWT) has been successfully employed as a core technique in some video coding frameworks. 3-D wavelet-based scalable video codec

The work has been partially supported by the NSFC grants (No. 60632040, No. 60772099, No. 60802019, No. 60928003, No. 60736043), and the Program for New Century Excellent Talents in University (No. NCET-09-0554).

1068-0314/11 $26.00 © 2011 IEEE
DOI 10.1109/DCC.2011.16

introduced by Microsoft [2] serves as a good example because of the inherent characteristics of multiscale representation in the 2-D spatial wavelet decomposition along with the temporal scalability supported by 1-D temporal wavelet transform. Basically, a separable 2-D wavelet filter commonly used in image compression applications is extended from two 1-D filters in the vertical and horizontal direction via tensor product. Therefore 2-D wavelet basis can only capture the scan-lines or the 1-D discontinuity on edge points, but cannot efficiently model the smoothness along the curves such as contours and directional textures which are abundant in natural images.

To solve the problem of lacking directionality of wavelet basis, a category of transforms which employ 2-D nonseparable directional filter banks are proposed. Contourlet [3] is one of the most typical example among these transforms for it can provide multiscale and directional image representations by employing the Laplacian pyramid and uniform nonseparable directional filter banks (UDFB). The best M-term approximation error decay rate for contourlets is $O((\log M)^3 M^{-2})$, which is much better than the rate of $O(M^{-1})$ for wavelets. Conceptually, the Laplacian pyramid iteratively split $L^2(\mathbb{R}^2)$ into self-nested and complete subspaces while the UDFBs are applied to each highpass subspace to combine all point discontinuities on the same direction into one coefficient. Thus, any modification to the choice of pyramid or directional filter banks will derive a unique transform with its own characteristic. For example, by substituting critical subsampling Mallat pyramid for the Laplacian pyramid which has a redundancy factor of 4/3, a nonredundant version of contourlet transform called wavelet-based contourlet (WBCT) is introduced [5, 6]. However, due to the frequency scrambling existing in all the levels of highpass wavelet subbands, WBCT suffers from severe pseudo-Gibbs phenomena artifacts when some coefficients become zero during NLA and coding. Furthermore, taking account of the fact that the stopband frequency of the Laplacian pyramid filter is over $[-\pi/2, \pi/2]^2$ which also results in fuzzy artifacts during image coding in the original contourlet transform [7], sharp frequency localized contourlet (contourlet/SFL) [8] is another example which significantly reduces the frequency scrambling by employing a new multiscale decomposition which can flexibly configure the width of transition band of the lowpass filters. Contourlet/SFL provides a considerable potential to further improve the coding performance on the trade-off between transform redundancy and frequency scrambling.

The contourlet/SFL basis form an overcomplete dictionary which seems contradictory to the goal of image/video compression, thus it is a crucial task to seek a good approximation which uses the minimal number of significant coefficients before compression. For a given signal, we cannot have a well-posed representation based upon overcomplete dictionaries but nevertheless such a redundancy provides freedom to find the sparsified solution which is friendly to data compression. In recent years, a series of iterative, greedy search algorithms such as matching pursuit (MP) [9], orthogonal MP [10] and iterative thresholding (IT) [11, 12] were proposed to seek sparse solutions to underdetermined linear systems. Since the contourlet transform is linear, and therefore the coefficients can be further sparsified to get a better coding performance.

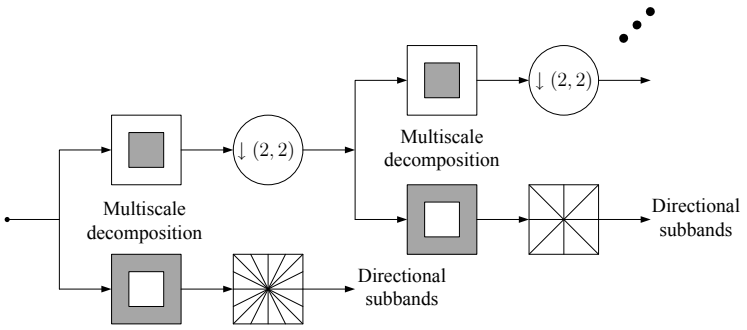
In this paper, we propose a low bit rate video compression framework which is based on the sparsified contourlet/SFL transform. It consists of four main stages: in the temporal decomposition stage, the temporal dependencies of source frames have been eliminated along the motion trajectories by lifting-based motion compensated temporal filtering (MCTF); in the spatial decomposition stage, each temporal subband is further decomposed into multiscale subbands via a wavelet-contourlet/SFL hybrid transform. The transform coefficients are then sparsified by IT in accordance with the tolerant error bound. Finally, the sparsified spatio-temporal subband co-

efficients are coded by a directional embedded image coding system. Experimental results of the sparse representation based video codec present a higher performance than the wavelet-based scalable video codec [2] with a comparable computational complexity.

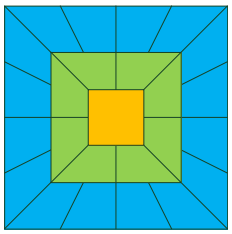
The remainder of this paper is organized as follows. Section 2 briefly introduces the contourlet/SFL transform. How to sparsify the redundant directional transform via IT is presented in Section 3. An embedded coding method which can exploit the inter-band and inter-scale dependency of the directional subbands is introduced in Section 4. The application as a scalable video codec as well as the experimental results are shown in Section 5, followed by the conclusion in Section 6.

2 Contourlet/SFL Transform

A generic operation flow within contourlet transform family can be illustrated by Fig. 1(a). A critically sampling Laplacian pyramid iteratively decomposes 2-D frequency spectrum into multiscale bandpass and lowpass subbands, followed by UDFBs applied on each bandpass subband. Finally the image will be decomposed into wedge-shaped divisions as shown in 1(b).



(a) Contourlet Transform Operation Flow



(b) Frequency Division

Figure 1: The contourlet operation flow and one possible resulting frequency division

In practice, the potential performance of the contourlet transform may be affected by the non-ideal filters in the multiscale decomposition stage. The prevalent 2-D lowpass filters used in the Laplacian pyramid are separable and 9/7 biorthogonal filters with their passband and transition band not restricted within $[-\pi/2, \pi/2]^2$, which fail to satisfy the Nyquist criteria with respect to critical sampling. Hence frequency scrambling appears after each bandpass is filtered by UDFBs and it will degrade the coding performance.

As a solution to overcome the scrambling, Lu and Do introduce the contourlet/SFL transform [8] which designs a new multiscale pyramid with zero response out of $[-\pi/2, \pi/2]^2$. Raised-cosine filters are employed in the new multiscale pyramid for the first level and all other levels by providing extra free parameters which can flexibly configure the width of transition band of the low-pass filters to cancel out the scrambling components. Suppose the separable 2-D lowpass filters $L_i^{(2D)}(\omega)$ used in the level i can be written as $L_i^{(2D)}(\omega) = L_i^{(1D)}(\omega_1) L_i^{(1D)}(\omega_2)$, while each $L_i^{(1D)}(\omega)$ is a 1-D filter with passband frequency $\omega_{p,i}$ and stopband frequency $\omega_{s,i}$, defined as

$$L_i^{(1D)}(\omega) = \begin{cases} 1 & \text{for} \quad |\omega| \le \omega_{p,i} \\ \frac{1}{2} + \frac{1}{2}\cos\frac{(|\omega|-\omega_{p,i})\pi}{\omega_{s,i}-\omega_{p,i}} & \text{for} \quad \omega_{p,i} \le |\omega| \le \omega_{s,i} \\ 0 & \text{for} \quad \omega_{s,i} \le |\omega| \le \pi \end{cases} \tag{1}$$

where the passband frequency $\omega_{p,i}$ and the stopband frequency $\omega_{s,i}$ are set to $4\pi/21$ and $10\pi/21$, respectively [8]. Since the output signal of the lowpass filters are always downsampled by $(2,2)$, the resulting redundancy ratio of the entire transform keeps the same as $4/3$ as the original one.

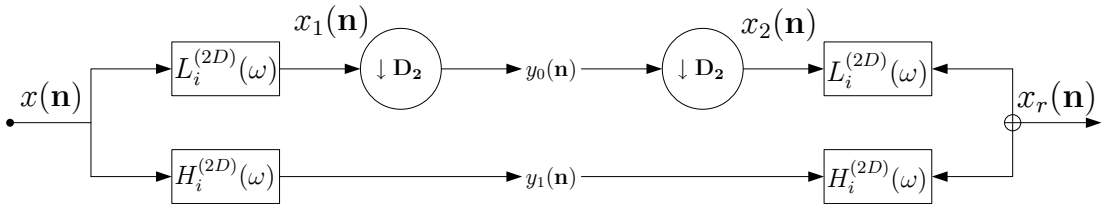


Figure 2: Type-II 2-channel filter banks

As shown in Fig. 2, each level of pyramid should fulfill the requirement of perfect reconstruction (PR) to allow the entire multiscale pyramid achieve PR property. According to the theory of general two-dimensional multirate filter banks, we have

$$X_2(\omega) = Y_1(\mathbf{D_2}^T\omega) = \frac{1}{|det(\mathbf{D_2})|}\sum_{\mathbf{k}\in\mathcal{N}(\mathbf{D_2}^T)} X_1(\omega - 2\pi\mathbf{D_2}^{-T}\mathbf{k}) \tag{2}$$

where $\mathbf{D_2} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ is the critical sampling matrix and $\mathbf{k} \in \mathcal{N}(\mathbf{D_2}^T)$ is the coset vectors as $k_0 = [0,0]^T$, $k_1 = [0,\pi]^T$, $k_2 = [\pi,0]^T$ and $k_3 = [\pi,\pi]^T$. Since $X_1(\omega) = X(\omega)L_i^{(2D)}(\omega)$, we have

$$X_2(\omega) = \frac{1}{4}\sum_{i=0}^{3} X(\omega - 2\pi\mathbf{D_2}^{-T}k_i)L_i^{(2D)}(\omega - 2\pi\mathbf{D_2}^{-T}k_i) \tag{3}$$

Obviously,

$$X_r(\omega) = X_2(\omega)L_i^{(2D)}(\omega) + Y_1(\omega)H_i^{(2D)}(\omega) \tag{4}$$

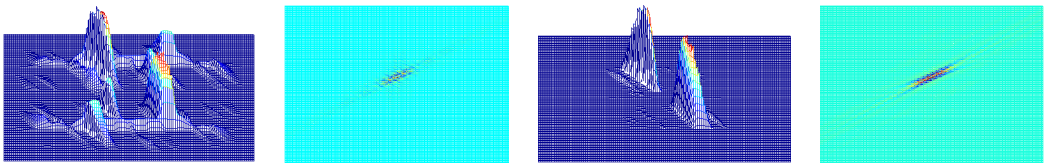
put Eq. 3 and $Y_1(\omega) = X(\omega)H_i^{(2D)}(\omega)$ into 4, we have

$$\begin{aligned} X_r(\omega) = X(\omega)&\left[\left|H_i^{(2D)}(\omega)\right|^2 + \frac{1}{4}\left|L_i^{(2D)}(\omega)\right|^2\right] \\ &+ \frac{1}{4}\sum_{i=1}^{3} X(\omega - 2\pi\mathbf{D_2}^{-T}k_i)L_i^{(2D)}(\omega - 2\pi\mathbf{D_2}^{-T}k_i)L_i^{(2D)}(\omega) \end{aligned} \tag{5}$$

Assuming all the spectrum aliasing of the lowpass filters have been cancelled out, in order to keep the PR property, i.e., $X_r(\omega) = X(\omega)$, the relationship between $H_i^{(2D)}(\omega)$ and $L_i^{(2D)}(\omega)$ should be

$$\left|H_i^{(2D)}(\omega)\right|^2 + \frac{1}{4}\left|L_i^{(2D)}(\omega)\right|^2 \equiv 1 \tag{6}$$

Fig. 3 shows the difference of the frequency and spatial basis images between the original contourlet transform and the contourlet/SFL transform. For the contourlet basis images, Fig. 3(a) indicates an obvious impact of the scrambling on the frequency domain and therefore the basis does not have a good spatial regularity as shown in Fig. 3(b). For the contourlet/SFL basis images, the scrambling problem is greatly suppressed in Fig. 3(c) and the spatial regularity shown in Fig. 3(d) is much better.



(a) (b) (c) (d)

Figure 3: Comparison of basis images. From left to right: Contourlet frequency domain basis, Contourlet spatial domain basis, Contourlet/SFL frequency domain basis, Contourlet/SFL spatial domain basis

3 An Iterative Thresholding Approach

Redundancy of the contourlet/SFL transform would be a major impediment to the coding efficiency, so that it is crucial to sparsify the transform coefficients as much as possible. The contourlet/SFL transform provides a multiresolution directional tight frame with frame bounds equal to 1 if both the multiscale pyramid and UDFBs use orthogonal filters [3, 4]. In this case, the inverse transform would be the dual frame operator to achieve the linear reconstruction. Define the input signal by $\mathbf{y} \in \mathbb{R}^n$, the contourlet/SFL coefficients by $\mathbf{x} \in \mathbb{R}^m$, and the redundant system can be represented as $\mathbf{\Psi x} = \mathbf{y}$ and $\mathbf{\Psi}^\dagger \mathbf{y} = \mathbf{x}$, where $\mathbf{\Psi} \in \mathbb{R}^{n \times m} (n < m)$ is the synthesis operator and $\mathbf{\Psi}^\dagger = \mathbf{\Psi}^T (\mathbf{\Psi \Psi}^T)^{-1}$ is the analysis operator which is the pseudo-inverse of the $\mathbf{\Psi}$. For the inverse transform $\mathbf{\Psi x} = \mathbf{y}$, infinite number of solutions $\mathbf{x}$ exist to this underdetermined system.

A straightforward model to find the sparsified $\mathbf{x}$ is to minimize the number of its significant coefficients, i.e., $\min \|\mathbf{x}\|_0$ s.t. $\mathbf{\Psi x} = \mathbf{y}$. However, [13] proves such a combinational search problem is NP-hard, and reduces it into a convex optimization problem by relaxing the l_0 norm into l_1 norm as follows:

$$\min \|\mathbf{x}\|_1 \text{ s.t. } \|\mathbf{\Psi x} - \mathbf{y}\|_2 \leq e \tag{7}$$

The iterative thresholding (IT) algorithm, is firstly developed for signal denoising [14] by shrinking the small coefficients without distorting the signals of interest too much. IT attracts people for its computational-friendly to the large-scale problems such as image/video processing. A generic IT algorithm can be described as follows [13]:

In the algorithm, $\mathbf{x}^k = \text{Shrink}(\mathbf{x}; \lambda)$ can be determined by either hard thresholding (set all but the largest λ elements of $|\mathbf{x}|$ to 0) or soft thresholding ($\max(|\mathbf{x}| - \lambda, 0) \cdot \frac{\mathbf{x}}{|\mathbf{x}|}$). The shrinkage step diminishes those small coefficients into zero, and makes the coefficients more sparse, also introduces some distortion. Define space $\mathbb{S} = \{\mathbf{x} = \mathbf{\Psi}^\dagger \mathbf{y}\}$, $\mathbb{S}^\perp$ is the orthogonal complementary space of $\mathbb{S}$, and $\mathbf{P} = \mathbf{\Psi}^\dagger \mathbf{\Psi}$ is the projector onto $\mathbb{S}$ while $\mathbf{P}^\perp = \mathbf{I} - \mathbf{\Psi}^\dagger \mathbf{\Psi}$ is the projector onto $\mathbb{S}^\perp$ since $\mathbf{P}^2 = \mathbf{P}$ and $\mathbf{P}^{\perp 2} = \mathbf{P}^\perp$. In each iteration k, the intermediate result $\mathbf{x}^k = \widehat{\mathbf{x}}^k + \epsilon^k$, where $\widehat{\mathbf{x}}^k$ achieves perfect

Problem: Find $\mathbf{x}$ that solves: $\min_{\mathbf{x}} \lambda\|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{y} - \mathbf{\Psi x}\|_2^2$
begin: Initialize $k = 0$, solution $\mathbf{x}^k = \mathbf{0}$
repeat
  Back Projection: Compute $\mathbf{e}^k = \mathbf{\Psi}^\dagger(\mathbf{y} - \mathbf{\Psi x}^k)$
  Line Search: Find μ_k which minimizes the function $\|\mathbf{x}^k + \mu_k \mathbf{e}^k\|_1$ by gradient descent algorithm
  Shrinkage: $\mathbf{x}^{k+1} = \text{Shrink}(\mathbf{x}^k + \mu_k \mathbf{e}^k; \lambda)$ with threshold λ
  Increment k **by 1**
until $\|\mathbf{x}^k - \mathbf{x}^{k-1}\|_2^2 \leq \epsilon$;

Algorithm 1: Iterative Thresholding Algorithm

reconstruction and ϵ^k is the noise caused by shrinkage. According to the Algorithm 1, we have

$$\begin{aligned}
\widehat{\mathbf{x}}^{k+1} &= \mathbf{x}^k + \mu\mathbf{\Psi}^\dagger(\mathbf{y} - \mathbf{\Psi x}^k) \\
&= \widehat{\mathbf{x}}^k + \epsilon^k + \mu\mathbf{\Psi}^\dagger(\mathbf{\Psi}\widehat{\mathbf{x}}^k - \mathbf{\Psi x}^k) \\
&= \widehat{\mathbf{x}}^k + (\mathbf{I} - \mathbf{\Psi}^\dagger\mathbf{\Psi})\epsilon^k + (1-\mu)\mathbf{\Psi}^\dagger\mathbf{\Psi}\epsilon^k \\
&= \widehat{\mathbf{x}}^k + \mathbf{P}^\perp\epsilon^k + (1-\mu)\mathbf{P}\epsilon^k \\
&= \widehat{\mathbf{x}}^0 + \sum_{i=0}^{k} \mathbf{P}^\perp\epsilon^i + (1-\mu)\mathbf{P}\epsilon^i
\end{aligned} \tag{8}$$

Through the back projection process, the initial result of $\widehat{\mathbf{x}}^0$ is sparsified by introducing orthogonal components to $\mathbb{S}$ with each iteration. The normalized scalar μ is used to control the convergence speed. With $\mu = 1$, none of the non-zero coefficients in $\mathbb{S}$ will go back to the reconstructed coefficients in the next iteration, therefore the shrinkage process reduces the number of non-zero coefficients. When $\mu > 1$, the convergence speed of reconstructed image energy is much faster but at the price of introducing more components in $\mathbb{S}$. The transform coefficients achieve sparseness until $\mathbf{x}^k \simeq \mathbf{x}^{k-1}$.

4 Embedded Directional Subband Coding

The image is decomposed into two sorts of subbands, the low frequency subbands and the high frequency subbands after the contourlet/SFL transform. The former ones carry the approximations of one image which contain most of its energy, while the latter ones carry the details of one image which play a vital role in reconstruction quality. Unlike the wavelet transform which has weak correlation among the cousin and parent subbands, subbands of the contourlet/SFL transform exhibit much higher inter-band and inter-scale dependencies [15]. Based on the measured statistics research, contourlet coefficients tend to be clustered in spatial and frequency domains, so it would be more probable to find a significant coefficient where it has significant neighbors.

Traditional wavelet-based embedded entropy coding algorithms such as EBCOT [16] used in the JPEG2000 and its 3-D counterpart 3D-ESCOT [17] used in video coding only take intra-band dependencies into account. They mainly consist of three coding passes: significant propagation (SP) pass, magnitude refinement (MR) pass and cleanup (normalization) pass. To exploit the intra-band dependency in the wavelet

subbands, the SP pass is established based on the observation that the significant coefficients always tend to be clustered in the high frequency subbands, and the previously detected significant coefficients can be regarded as seeds to search new significant ones.

Here we propose a three-dimensional embedded directional subband coding with optimized truncation (3D-EDSCOT) which can further exploit the inter-band and inter-scale dependencies. Unlike the regular 3D-ESCOT which only focuses on the context modeling of the significant neighborhood, 3D-EDSCOT utilizes the conditional probability relationship of one coefficient with its cousin and parent by introducing two more coding passes: cross-scale prediction (CSP) pass and cross-neighbor prediction (CNP) pass. The new coding method retains random access property. We summarize the coding procedure in Algorithm 2 which consists of five consecutive passes in each bit-plane.

• **Initialization:** Divide the contourlet subbands into independent $N \times N$ blocks, find the coefficient with the highest non-zero bitplane M and set current processing bitplane $n \leftarrow M$.
• **Significance Propagation (SP) Pass:** If $n == M$, go to **MR Pass**.
For those coefficients that are neither significant nor scanned, check its 3-by-3 horizontal, vertical and temporal neighbors. If at least one significant neighbor exists, we use ZC primitive to code the bits, if any of them become significant, the SC primitive to encode their signs.
• **Cross-scale Prediction (CSP) Pass:** If current subband is a wavelet subband, go to **MR Pass**.
For those coefficients that are neither significant nor scanned, check its parent coefficient within a 5-by-5 cross shaped horizontal and vertical neighbors. Similar to SP pass, if at least one of them has already been significant, encode the newly scanned bits by ZC primitive, if any of them become significant, encode the signs by SC primitive.
• **Cross-neighbor Prediction (CNP) Pass:** If current subband is a wavelet subband, go to **MR Pass**.
For those coefficients that are neither significant nor scanned, check its left and right cousin coefficient within 3-by-3 horizontal, vertical neighbors. Encode the candidate bits with preferred cousins and if the newly scanned coefficients become significant, encode their signs.
• **Magnitude Refinement (MR) Pass:** Use MR primitive to encode the nth bit of the coefficients that has been detected as significant in previous bit-planes.
• **Normalization Pass:** Encode all the remaining coefficients which are neither significant nor scanned by ZC or SC primitives.
• **Process a New Bit-plane:** Reset scanning record, set $n \leftarrow n - 1$ if $n \geq 0$, and go to *SP Pass*.

Algorithm 2: Three-dimensional Embedded Directional Subband Coding with Optimized Truncation (3D-EDSCOT)

5 Experimental Results

5.1 Experiment Setup

The wavelet-based scalable video coding (WSVC) framework *VidWav* [2] is used as the reference software. It consists of three main process stages: the temporal dependencies of source frames have been eliminated along the motion trajectories by

lifting-based motion compensated temporal filtering (MCTF), and then each temporal subband frame is further decomposed into wavelet subbands. Consequently the coefficients in the spatio-temporal subbands are coded by 3D-ESCOT. In our experiment, the latter two are replaced by the sparse contourlet/SFL transform and 3D-EDSCOT, as shown in Fig. 4. More exactly, the video frames in one GOP are temporally decomposed into four temporal subbands, and each temporal subband is further spatially decomposed by the sparse hybrid wavelet and contourlet/SFL transform into four scales for CIF test sequence (*Walk* and *Foreman*). The aliasing-free multiscale pyramid decomposes the image into three scales, and then UDFBs with ladder structures [18] are used at the two highest scales. For the two lowest scales, 9/7 DWT is employed because it is reported that UDFBs do not perform as well as DWT in the low frequency subbands [7]. We use the iterative shrinkage algorithm based on hard thresholding, the sparseness parameter λ is optimized in terms of video frames and the coding bit-rate operating points by experiments. The predetermined threshold in the stopping criterion is $\epsilon = 0.01$.

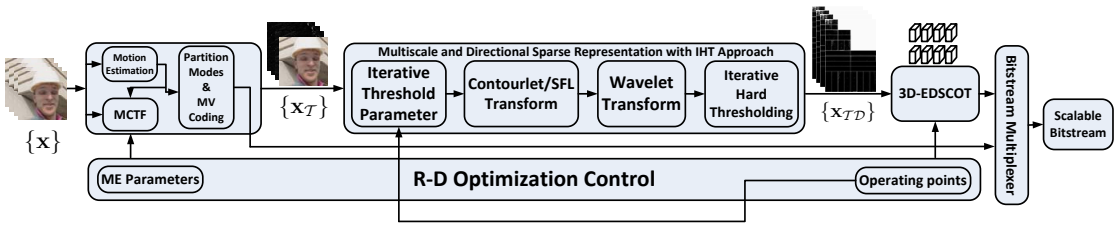
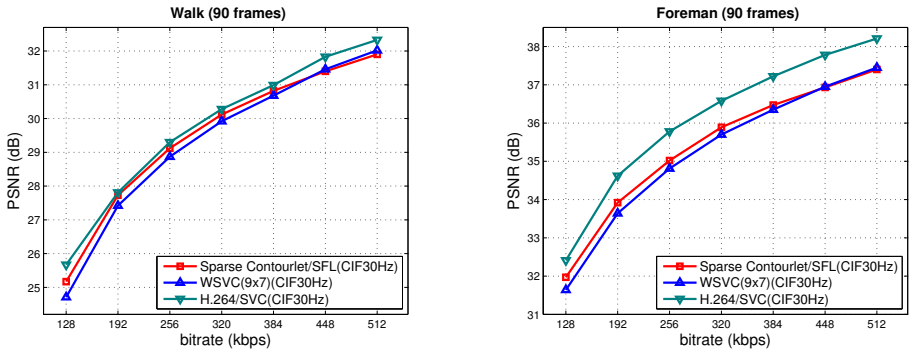


Figure 4: Video compression scheme in the experiment



(a) *Walk* (b) *Foreman*

Figure 5: Coding performance comparison by PSNR

5.2 Numerical Results

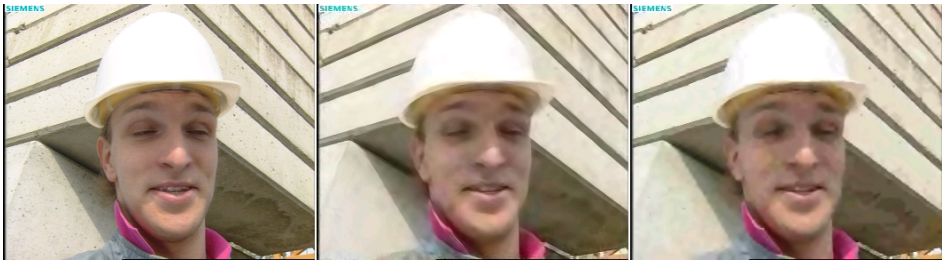
Fig. 5 and Table 1 give the rate-distortion performance comparison in terms of PSNR and SSIM [19] between the wavelet-based *VidWav*, the H.264-based *JSVM* and the sparse representation based video coding scheme. SSIM can be served as an effective metric when qualifying video frames decoded at low bit rates, which assesses image quality by measuring the structural similarity. We can see that the proposed framework outweighs *VidWav* consistently in the range of low bit rates and proves the advantage of applying sparse representation on redundant transforms. Particularly, in the case of *Walk* sequence, our scheme achieves up to 0.5dB PSNR improvement over

VidWav. Fig. 6 shows the original CIF frame in *Walk* and *Foreman* and the visual effects of reconstructed frame coded by our scheme and *VidWav* at 128kbps, from left to right respectively. Since the sparse contourlet/SFL representation in our video coding scheme captures directional curves more efficiently, more details of directional information have been well preserved after reconstruction. The open-loop control of MCTF and the missing of intra-prediction reduce the coding efficiency gains when comparing to *JSVM*. In fact, it is not relevant to the attention of this paper with a sparsified directional decomposition and representation.

Table 1: Coding performance comparison by SSIM

	SSIM of *Walk* CIF @ 30Hz						
bit rate(kbps)	128	192	256	320	384	448	512
DMSVC	0.765	0.845	0.879	0.902	0.912	0.923	0.927
WSVC(9×7)	0.741	0.833	0.873	0.899	0.910	0.926	0.931
H.264/SVC	0.776	0.843	0.878	0.902	0.915	0.927	0.935
	SSIM of *Foreman* CIF @ 30Hz						
bit rate(kbps)	128	192	256	320	384	448	512
DMSVC	0.891	0.916	0.929	0.937	0.943	0.950	0.952
WSVC(9×7)	0.878	0.906	0.921	0.932	0.938	0.947	0.951
H.264/SVC	0.887	0.923	0.938	0.947	0.945	0.956	0.958

(a) The original and reconstructed *Walk* frame



(b) The original and reconstructed *Foreman* frame

Figure 6: Reconstructed visual quality. Each group from left to right: the original frame, the reconstructed frame by sparse representation based scheme and the reconstructed frame by *Vidwav*.

6 Conclusions

In this paper, we exploit video coding based on the sharp frequency localized contourlet (contourlet/SFL) transform with sparse representation. By using the con-

tourlet/SFL transform with non-aliasing multiscale pyramid, the video reconstruction performance can be further improved. Although the redundancy of the contourlet/SFL transform seems to be the main obstacle for video coding, it provides extra freedom for signal sparse approximation by iterative thresholding algorithm. Since the strong inter-band and inter-scale dependencies among the directional subbands can be utilized to pursue better performance, a three-dimensional embedded directional subband coding method is used for constructing final bitstream. Experimental results validate a superior coding performance and visual quality over the wavelet-based video coding framework at low bit rates.

References

[1] T. Wiegand, G. J. Sullivan, and A. Luthra, "Draft ITU-T Rec. H.264/ISO/IEC 14496-10 AVC," presented at the JVT ISO/IEC MPEG, ITU-T VCEG, Doc. JVT-G050r1, 2003.

[2] R. Xiong, X. Ji, D. Zhang, *et al.*, "Vidwav Wavelet Video Coding Specifications," *ISO/IEC JTC1/SC29/WG11/M12339*, Poznan, Jul. 2005.

[3] M. N. Do and M. Vetterli, "Contourlets," in *Beyond Wavelets*, G. V. Welland, Ed. Amsterdam, The Netherlands: Academic, Ch.4, pp. 83-105, 2003.

[4] Do, M.N.; Vetterli, M.; , "Framing pyramids," *Signal Processing, IEEE Transactions on*, vol.51, no.9, pp. 2329-2342, Sept. 2003.

[5] R. Eslami and H. Radha, "Wavelet-based contourlet transform and its application to image coding," *Image Processing, 2004. ICIP '04. 2004 International Conference on*, vol.5, pp. 3189-3192, Oct. 2004.

[6] R. Eslami and H. Radha, "A New Family of Nonredundant Transforms Using Hybrid Wavelets and Directional Filter Banks," *Image Processing, IEEE Transactions on*, vol.16, no.4, pp.1152-1167, Apr. 2007.

[7] T. T. Nguyen and S. Oraintara, "On the aliasing effect of the contourlet filter banks," in *Proceedings of the 14th European Signal Processing Conference (EUSIPCO '06)*, Florence, Italy, Sept. 2006.

[8] Y. Lu; M. N. Do; , "A New Contourlet Transform with Sharp Frequency Localization," *Image Processing, 2006 IEEE International Conference on*, pp.1629-1632, 8-11 Oct. 2006.

[9] Mallat, S.G.; Zhifeng Zhang; , "Matching pursuits with time-frequency dictionaries," *Signal Processing, IEEE Transactions on*, vol.41, no.12, pp.3397-3415, Dec. 1993.

[10] J. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *Information Theory, IEEE Transaction on*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[11] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal. Applicat.*, vol. 14, no. 5, pp. 629-654, Dec. 2008.

[12] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, pp. 1413-1457, 2004.

[13] A.M. Bruckstein, D.L. Donoho, and M. Elad, "From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images", *SIAM Review*, Vol. 51, No. 1, Pages 34-81, Feb. 2009.

[14] Donoho, D.L.; , "De-noising by soft-thresholding," *Information Theory, IEEE Transactions on*, vol.41, no.3, pp.613-627, May 1995.

[15] Po, D.D.-Y.; Do, M.N.; , "Directional multiscale modeling of images using the contourlet transform," *Image Processing, IEEE Transactions on*, vol.15, no.6, pp.1610-1620, June 2006.

[16] D. Taubman, "High performance scalable image compression with EBCOT," *Image Processing, IEEE Transactions on*, vol.9, no.7, pp.1158-1170, Jul. 2000.

[17] J. Xu, Z. Xiong, S. Li, *et al.*, "Three-Dimensional Embedded Subband Coding with Optimized Truncation (3D ESCOT)," *Applied and Computational Harmonic Analysis: Special Issue on Wavelet Applications in Engineering*, vol. 10, pp. 290-315, May 2001.

[18] S. M. Phoong, C. W. Kim, P. P. Vaidyanathan, *et al.*, "A new class of two-channel biorthogonal filter banks and wavelet bases," *Signal Processing, IEEE Transactions on*, vol. 43, no. 3, pp. 649-665, Mar. 1995.

[19] Z. Wang, A. C. Bovik, H. R. Sheikh *et al.*, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol.13, no.4, pp.600-612, Apr. 2004.