Scalable Video Compression Framework With Adaptive Orientational Multiresolution Transform and Nonuniform Directional Filterbank Design

Hongkai Xiong, *Senior Member, IEEE,* Lingchen Zhu, Nannan Ma, and Yuan F. Zheng, *Fellow, IEEE*

***Abstract*—Although wavelet-based scalable video coding becomes the state-of-the-art video compression engine for its adaptability to heterogeneous networks and clients, a large number of attempts have been made to integrate local directionality onto discrete wavelet transform to explore the intrinsic geometrical structures. Taking into consideration that the contours and textures scattered in different scales change their directional resolutions as their curvatures change, we investigate adaptive directional resolutions along scales to achieve the dual (scale and orientation) multiresolution transform. This paper proposes nonuniform directional frequency decompositions for video representation and approximation, and exploits the nonuniformity of orientation multiresolution distribution and designs nonuniform directional filter banks to make the geometrical transform more sparse and efficient. The nonuniform directional frequency decomposition under arbitrary scales is fulfilled by a nonsymmetric binary tree (NSBT) topology structure with nonuniform directional filterbank design. In turn, the proposed scalable video coding framework, called DMSVC, is enriched with the dual multiresolution transform. Each temporal subband through motion compensated temporal filtering is further decomposed into multiscale subbands, and the highpass wavelet subspaces are divided into an arbitrary number of directional subspaces in alignment with the orientation distribution via phase congruency to establish NSBT. The paraunitary perfect reconstruction condition is provided through a polyphase identical form of filter bank. Comparing with the isolated wavelet basis, our transform provides a greater correlated set of localized and anisotropic basis functions. The spatio-temporal subband coefficients are coded by a 3-D ESCOT entropy coding algorithm which is adopted to match the structure of NSBT. Experimental results show that the reconstructed video frames DMSVC in the proposed DMSVC scheme have better visual quality than existing scalable video coding schemes. It could produce higher compression ratio on video sequences full of directional edges and textures.**

Manuscript received May 20, 2010; revised December 24, 2010; accepted February 1, 2011. Date of publication March 28, 2011; date of current version August 3, 2011. This work was supported in part by the National Natural Science Foundation of China, under Grants 60772099, 60928003, 60736043, and 60632040, and by the Program for New Century Excellent Talents in University, under Grant NCET-09-0554. This paper was recommended by Associate Editor D. S. Turaga.

H. Xiong and L. Zhu are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: xionghongkai@sjtu.edu.cn; extreme001@sjtu.edu.cn).

N. Ma is with Marvell Technology Group, Ltd., Shanghai 201203, China (e-mail: puppuppy@sjtu.edu.cn).

Y. F. Zheng is with the Department of Electrical and Computer Engineering, Ohio State University, Columbus, OH 43210 USA (e-mail: zheng@ece.osu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2011.2133310

***Index Terms*—Directional filter banks, motion compensated temporal filtering, multiscale geometric analysis, scalable video coding, sparse coding.**

I. Introduction

SCALABLE video coding (SVC) technique is now taking place of the traditional single operating point video coding schemes for its adaptability to the transmission on heterogeneous networks and clients. A typical SVC encoder provides a multi-dimension layer-dependent graph structure with continuous and discrete achievable rate regions. Each layer, in conjunction with all layers it depends on, forms one representation of the video signal at a certain spatio-temporal resolution and quality level. In the past years, several typical schemes have been proposed to MPEG for standardization, especially the H.264 scalable extension [1] and Barbell-lifting wavelet-based SVC [2]. In particular, wavelet-based approaches can produce high coding efficiency because of the inherent characteristics of multiresolution spatio-temporal representation and efficient approximation of 1-D piecewise smooth signals. Traditionally, a prevailing 2-D discrete wavelet transform (DWT) is implemented by the tensor product of separable 1-D filters in the vertical direction and horizontal direction so that the basis of wavelet spaces only provides limited directions such as LH, HL, and HH, which can only capture the scan-lines or the 1-D discontinuity on edge points, but cannot see the smoothness along the curves such as contours and textures. The nonlinear approximation (NLA) error decay of the best $\mathcal{M}$ wavelet coefficients for images containing 2-D discontinuities is $\mathcal{O}(\mathcal{M}^{-1})$ [3], which is due to the fact that the 2-D discontinuities result in many large coefficients in high frequency subbands. It turns out that DWT is not the most optimal solution for video coding and compression and there is quite plenty of room for improvement. Moreover, since the directional frequency distribution in natural 2-D signals is nonuniform, we need to find an optimal partition scheme which can catch such a nonuniform distribution dynamically in an adaptive manner. In order to represent the even amount of information with the least bits, a more efficient spatial decomposition should be investigated to represent the video signal, while preserving the characteristic of multiresolution so as to be compatible with the existing SVC framework.

Attempts have been made to integrate local directionality into lifting-based DWT. For instance, adaptive directional

lifting [4] and direction-adaptive DWT [5] achieve directional lifting with adaptation to image direction features in local windows, and they outperform the traditional DWT in both subjective and objective quality for image coding. However, their division tree occupies a considerable portion in the bitstream so that they reveal an inferior performance and lose the scalability property on very low bit rates. Multiscale geometric analysis provides another category of image decompositions which focuses on the directional information on frequency domain, such as curvelet [6] and contourlet [7]. Curvelet is implemented on the continuous space and polar coordinates, such that it is really a challenge to convert it into the discrete world. Contourlet utilizes the local dependency of wavelet coefficients across scales and directions so that it employs a scale multiresolution Laplacian transform to split $L^2(\mathbb{R}^2)$ into self-nested complete subspaces and then employs a directional filter bank (DFB) on each subspace to combine all point discontinuities on the same direction into one coefficient. It also inherits the anisotropism of curvelets with NLA error decay rate of $\mathcal{O}(\mathcal{M}^{-2})$, but its main disadvantage lies on the 4/3 redundancy introduced by Laplacian pyramid, which renders it inapplicable to video coding directly. Later, a group of nonredundant multiscale geometric decompositions including CRISP-contourlets [8] and wavelet-based contourlets (WBCT) [9] were proposed to eliminate the transform redundancy by employing nonredundant multiresolution filter banks, but the approaches ignore the nonuniformity of orientation distribution of image spectrum and simply divide it into uniform directional subbands. They decompose an image by uniform directional filter banks (UDFB), which are not able to achieve more sparse representation adaptively and optimally. By only applying DFBs onto high-frequency regions of the wavelet subbands, a new family of transforms using hybrid wavelets and directional filter banks (HWD) [10] were developed to reduce the ringing artifacts which is introduced by applying DFBs to the low-frequency smooth regions of images.

From source coding perspective, it is known that sparse representation is sought to maximally capture interested features of a signal with maximally decimated coefficients. This representation is performed by a signal projection onto another space expanded by a complete orthogonal basis, whose efficiency is embodied by energy convergence to a few of large coefficients which are inner products of signal and basis. These nonredundant geometrical transforms provide a more reasonable choice than wavelet in scalable video coding and compression. Taking into account that the contours and textures scattered in different scales usually change their directional resolutions as their curvatures change, we investigate adaptive directional resolutions along scales to achieve the dual (scale and orientation) multiresolution transform. This paper proposes nonuniform directional frequency decompositions for image representation and approximation, and exploits the nonuniformity of orientation multiresolution distribution in scalable video coding and designs nonuniform directional filter banks (NUDFB) to make the geometrical transform more sparse and efficient. It is worth mentioning that NUDFB is imperative to achieve an orientation multiresolution under a certain scale. Despite extensive methods to design 1-D nonuniform filter banks [11], the advance on 2-D nonuniform filter banks has been hindered by complicated issues in design process. For example, universal anti-aliasing filter bank for arbitrary downsampling matrices, some of which may be irrational, may not be accessible. A non-symmetric binary tree (NSBT) structured filter bank is proposed to fulfill NUDFB, because of the following advantages: 1) minimum branches or channels at each node to reduce the design complexity, especially for 2-D nonseparable filter banks; 2) more flexible to choose an appropriate frequency division; and 3) convenient to elaborate the binary tree structure, which is important to acquire the decomposition structure if a nonuniform decomposition is used. Although biorthogonal filter bank is less constrained in perfect reconstruction, orthogonal filter bank is chosen in this context owing to its attractive properties in subband coding applications [12]. For 2-D filters design in a filter bank, there are mainly two methods: to design a 2-D filter directly, and to get the target 2-D filter from a 1-D prototype filter [13]. The latter is employed to simplify both the design procedure and the implementation process to reduce the implementation complexity to $\mathcal{O}(\mathcal{N})$ other than $\mathcal{O}(\mathcal{N}^2)$. The 2-D nonuniform filter bank with NUDFB structure is of maximal decimation and paraunitary perfect reconstruction.

Two main contributions of this paper are the proposal of nonuniform directional frequency decompositions under arbitrary scales which are fulfilled by a NSBT topology structure with nonuniform directional filterbank design, and the development of a novel generic scalable video coding framework with the dual (nonredundant scale and orientational) multiresolution transform, called DMSVC. The proposed NUDFB provides a multiresolution on directions as well as wavelet filters provide a multiresolution on scales, and it is more flexible to statistically utilize the directional information of contours and textures in video frames to achieve a more efficient filter bank partition scheme. In the underlying dual multiresolution transform context, orientation resolution is regarded as an isolated variable from scale resolution. The wavelet basis function in each scale is converted to an adaptive set of nonuniform directional basis. Through the nonuniform frequency division, we can get arbitrary orientation resolution l at a direction of $c2^{-l}$ under a target scale. NUDFB is fulfilled by arraying the topology structure of a NSBT, as a symmetric extension from a two channel filter bank. The paraunitary perfect reconstruction condition is provided through a polyphase identical form of filter bank, in terms of 2-D nonseparable filters from a 1-D prototype. Comparing with the isolated wavelet basis, our transform provides a greater correlated set of localized and anisotropic basis functions with video frames, which can capture contours and textures with sparse coefficients.

As a prospective application, the proposed DMSVC consists of three main stages: the temporal dependencies of source frames are eliminated along the motion trajectories by lifting-based motion compensated temporal filtering (MCTF); in the spatial stage, each temporal subband is further decomposed into multiscale subbands, and the orientation distribution is estimated via phase congruency in the overcomplete wavelet domain to establish a NSBT for each scale, which is used

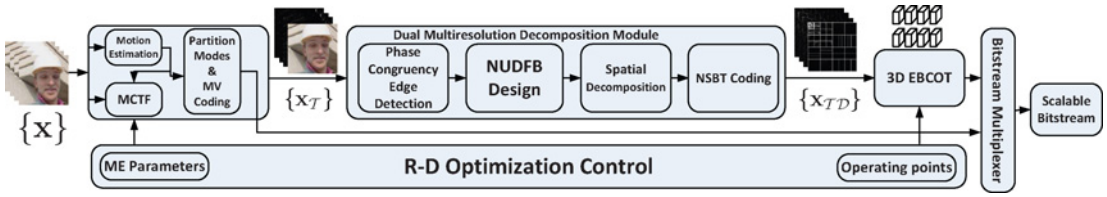

Fig. 1. DMSVC general framework.

as skeleton structure of NUDFB; through the orientation multiresolution decomposition via NUDFB, the highpass wavelet subspaces are divided into an arbitrary number of directional subspaces in alignment with the orientation distribution. Finally, the spatio-temporal subband coefficients are coded by a 3-D ESCOT entropy coding algorithm [14] which is adopted to match the structure of NSBT. It is conceptually similar to EBCOT [15] but employs a new 3-D context table that is more suitable for the SVC trace. Experimental results show that the reconstructed video frames DMSVC in the proposed DMSVC scheme have better visual quality than other SVC schemes. It could also produce higher compression ratio especially on those sequences full of directional edges and textures, and reveal a better performance on smooth curve representation and energy compaction.

The rest of this paper is organized as follows. Initially, we summarize all the math symbols throughout this paper in Table I. The proposed DMSVC framework is presented in Section II. Section III formulates the design process of the orthonormal basis of NUDFB used in the dual multiresolution transform and proves that the whole filter bank achieves perfect reconstruction. Extensive experiments are validated for nonlinear approximation and scalable video coding in Section IV. Finally, we conclude this paper in Section V.

$$\mathbf{x}_{\mathcal{T}}^{H_i} = \mathbf{x}_{2i+1} - \frac{1}{2}[\mathbf{x}_{2i}(\cdot + \mathbf{mv}_{2i+1,2i}) + \mathbf{x}_{2i+2}(\cdot + \mathbf{mv}_{2i+1,2i+2})] \quad (1)$$

$$\mathbf{x}_{\mathcal{T}}^{L_i} = \mathbf{x}_{2i} + \frac{1}{4}[\mathbf{x}_{\mathcal{T}}^{H_{i-1}}(\cdot + \mathbf{mv}_{2i,2i-1}) + \mathbf{x}_{\mathcal{T}}^{H_i}(\cdot + \mathbf{mv}_{2i,2i+1})]. \quad (2)$$

II. DMSVC Framework

The DMSVC framework is derived from the current wavelet-based scalable video coding (WSVC) video coding schemes, which can be categorized into two traces: to perform the MCTF on the full resolution video frame before the spatial decomposition, which is also called "T+2-D;" to perform the spatial decomposition on the full resolution video frame and then execute MCTF on each subband, which is often referred as "2-D+T."

Fig. 1 gives the entire process of DMSVC encoder framework which might apply to both schemes. If the pre-2-D-decomposition part is null, motion estimation and temporal filtering are applied to the full resolution frames to separate them into temporal lowpass subbands $\mathbf{x}_{\mathcal{T}}^{L}$ and highpass subbands $\mathbf{x}_{\mathcal{T}}^{H}$; otherwise, MCTF is performed on the subbands of transformed frames. In turn, a post 2-D dual multiresolution transform is applied to each temporal subband to decompose it into nonuniform directional subspaces. Suppose the input signal $\mathbf{x}$ goes through an N-level dual multiresolution transform and is indicated as $\mathbf{x}_{\mathcal{D}}$, which consists of a set of scale multiresolution subbands $\mathbf{x}_{\mathcal{D}} = \{\mathbf{x}_{\mathcal{D}}^{c}, \mathbf{x}_{\mathcal{D}}^{d_s(N)}, \cdots, \mathbf{x}_{\mathcal{D}}^{d_s(1)} | s = 1, 2, 3$ for LH, HL, HH subbands, respectively$\}$. Each scale subband $\mathbf{x}_{\mathcal{D}}^{d_s(k)}$ is composed of a set of nonuniform directional subbands $\mathbf{x}_{\mathcal{D}}^{d_s(k)} = \{\mathbf{x}_{\mathcal{D}}^{d_s^0(k)}, \mathbf{x}_{\mathcal{D}}^{d_s^1(k)}, \cdots, \mathbf{x}_{\mathcal{D}}^{d_s^{L-1}(k)}\}$. The overall spatial and temporal decomposition structure can be seen in Fig. 2, where MCTF is realized by dyadic DWT transform. MCTF can be implemented by a lifting structure involving the *predict* and *update* steps, and it also enables perfect reconstruction with sub-pixel motion alignment [16]. With the lifting structure, any traditional motion model that establishes a pixel-mapping relationship between two adjacent frames can be easily adopted by the motion aligned temporal filtering. Moreover, the lifting structure ensures perfect reconstruction under the condition of complex motion fields and fractional pixel motion vectors. We preserve the MCTF operation in the DMSVC framework to make the energy concentrated on the temporal lowpass bands and it will make the spatial transform more effective.

A typical biorthogonal 5/3 wavelet lifting structure adopted in MCTF firstly splits the input frames into even components $\mathbf{x_{2i}}$ and odd components $\mathbf{x_{2i+1}}$, and then two immediate neighboring frames are needed to establish a bi-directional *predict* or *update* signal. The motion vectors $\mathbf{mv}$ are obtained by block-based bidirectional motion estimation at each $\mathbf{x_{2i}}$ frame using two neighboring $\mathbf{x_{2i+1}}$ frames as references. We can obtain the high-pass temporal subbands $\mathbf{x}_{\mathcal{T}}^{H_i}$ in the *predict* step by (1) and the low-pass temporal subbands $\mathbf{x}_{\mathcal{T}}^{L_i}$ in the *update* step by (2). It can be seen that no matter what the distribution of the motion field is, MCTF based on the lifting structure can ensure the condition of perfect reconstruction.

After all the frames are decomposed into high-pass and low-pass temporal subbands, they are pushed into the dual multiresolution decomposition module. The first multiresolution, scale multiresolution, is achieved by wavelet decomposition using a simple syntax in the configuration file. Based on the scale multiresolution, orientation multiresolution is carried out adaptively according to the estimation result of the orientation distribution in the overcomplete wavelet space, and then NUDFB with the NSBT structure is formed to decompose and reposition the frames into spatial subbands. To determine the decomposition structure in an adaptive topology, we estimate the orientation distribution by using phase congruency metric within the overcomplete wavelet subspace. Initially, we build a full binary tree structure with deterministic depth where each leaf represents the distribution density of the directions in a uniform interval. Through the tree-pruning, it comes to be an NSBT where each leaf is not in the same depth anymore and

TABLE I
NOMENCLATURE TABLE

	Multi-Dimensional Symbols
$\mathcal{V}_j^2$	Approximation space with scale j
$\mathcal{W}_j^2$	Wavelet detail space with scale j
$\mathcal{W}_j^{2,k}$	Wavelet detail subspace with scale j and index $k = 1, 2, 3$
$\varphi_{j,\mathbf{k}}^{(2)}(\mathbf{t})$	2-D basis of $\mathcal{V}_j^2$
$\psi_{j,\mathbf{k}}^{(2),i}(\mathbf{t})$	2-D basis of $\mathcal{W}_j^2$
$\mathcal{D}_{j,p}^{(l)}$	Directional subspace with scale j and orientation resolution 2^{-l} and index $0 \le p < 2^l$
$\lambda_{j,\mathbf{k},p}^{(l)}$	2-D basis of $\mathcal{D}_{j,p}^{(l)}$
$h[\mathbf{n}]$	Filter's impulse response in discrete-time domain
$H[\omega]$	Filter's impulse response in frequency domain
Matrix M	Sampling matrix: a $d \times d$ nonsingular matrix of integers
$\mathcal{N}(M)$	Set of all integers $M\mathbf{x}$ with $\mathbf{x} \in [0, 1)^d$

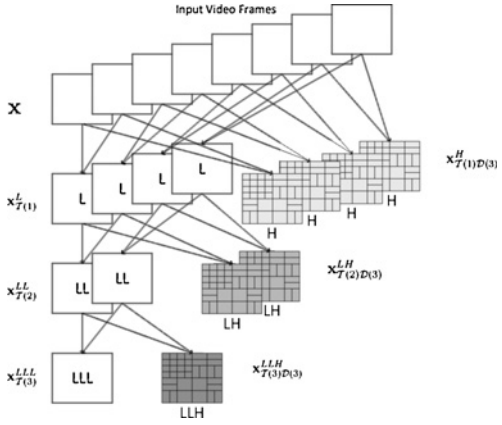

Fig. 2. Spatial and temporal subbands.

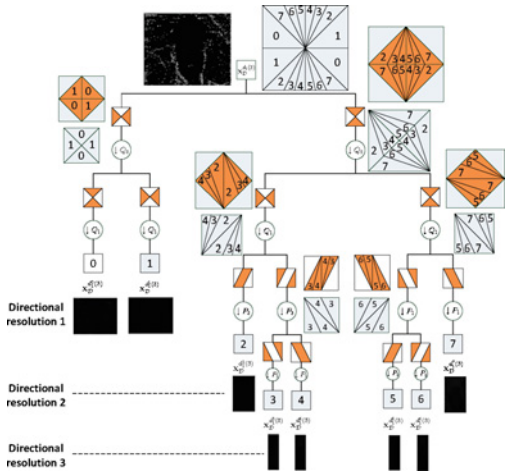

Fig. 3. Analysis part of NUDFB matching with the finest LH subband in one frame of *Foreman*.

represents basically equivalent directional distribution density. NUDFB, as the core component of the dual multiresolution transform, is fulfilled by applying 2-D nonseparable quadrature mirror filter banks (QMFB) on every two leaves extended from one common parent in NSBT. The two-channel QMFB in the 2-D case can be implemented by quincunx and parallelogram filters and allow for the perfect reconstruction. To show how NUDFB decomposes the wavelet subband into nonuniform directional subbands with different directional resolutions, Fig. 3 illustrates an example of the analysis part of NUDFB which fits for the finest LH subband in a frame of the *Foreman* sequence. Fig. 9 shows the impulse response of the NUDFB in both frequency and time domains. It can be seen that the basis functions in dual multiresolution transform are adaptive to the source and obey the anisotropy scaling law, so that the magnitude of coefficients is significantly reduced in these subbands. After the spatio-temporal modules, the coefficients are organized into 3-D blocks and coded with 3-D ESCOT. All the details will be discussed in the following sections.

III. DUAL MULTIRESOLUTION TRANSFORM WITH NUDFB

Strictly speaking, all the geometric transforms achieve the multiresolution on scale but ignore the nonuniformity of orientation distribution of curve smoothness such as contours and textures, so that they only divide the highpass subspaces into uniform directional subspaces. For example, the contourlet transform and WBCT, only decompose the image into 2^l directional subbands at each scale with fixed directional resolutions l via UDFB; therefore, they cannot get an adaptive and optimal representation of the source image. The directionality of the spectrum are obvious in the high frequency and the low frequency portions would also leak into several adjacent directional subbands. Furthermore, the contours and textures scattered in different scales usually change their directional resolutions as their curvatures change. Thus, adaptive directional resolutions are required for different scales. To take the advantage of the nonuniformity in directions, we introduce another multiresolution approach called orientation multiresolution to achieve the dual multiresolution transform. After we estimate the orientation distribution in the overcomplete wavelet subspaces, the orientation multiresolution is achieved by implementing NUDFB, which is represented by a NSBT structure. We prove that in this transform, any orientation resolution under a given scale can be achieved, and we can select proper analysis and synthesis filters in every pair of siblings of NSBT to fulfill the requirement of perfect reconstruction.

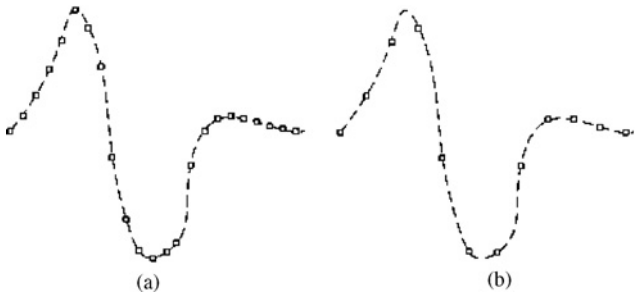

Fig. 4. Illustration of edge profile blurring by decimation. (a) Sampling of the edge profile. (b) Blurring by decimation.

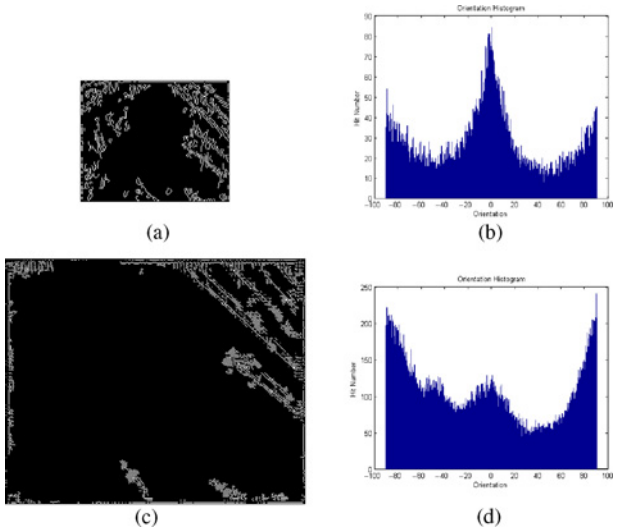

Fig. 5. Edge detection on wavelet and overcomplete wavelet HL subspaces. (a) Canny detection of the overcomplete wavelet HL subspace. (b) Orientation distribution estimated in the overcomplete wavelet HL subspace. (c) Canny detection of the wavelet HL subspace. (d) Orientation distribution estimated in the wavelet HL subspace.

A. Orientation Distribution Estimation in the Overcomplete Wavelet Domain

In order to achieve a more sparse representation, NUDFB is used to combine significant wavelet coefficients around curve discontinuities. The design of NUDFB obeys the principle that all direction pixels (dirpixels) contained in the source frame are scattered uniformly among the nonuniform filter banks, and the NSBT structure of NUDFB is obtained from the estimation result of orientation distribution at multiscale wavelet subspaces via common feature extraction methods such as edge detection. Hereafter, the extracted edge feature regarding orientation are called direction pixel (dirpixel).

In natural images and video frames, intensity within a local neighborhood of an edge tends to change slowly along the edge direction, and rapidly but smoothly along the direction vertical to the edge. The regular 2-D separable wavelet decomposition system consists of pre-filtering and decimation operations. Although the smoothness and phase continuity of edges along different directions can be retained after the linear pre-filtering system, they are damaged after the nonlinear decimation system. Furthermore, since the pre-filtering system is not ideal, aliasing always happens at the frequencies higher than one half of the Nyquist frequency. Since most high frequency components in 2-D signals are composed of edges and contours, they are more sensitive to frequency aliasing such that the estimation may not be accurate.

Generally speaking, the edge detection methods can be categorized into two classes: gradient-based methods which calculate gradient of every pixel within a small region [18] and phase-based methods that estimate the edges throughphase congruency [19]. Accuracy of typical gradient-based edge detection methods, such as Canny operator, relies on the adjacent pixels along the gradient directions. Fig. 4(a) shows a typical example of edge profile with rapid changing on intensity, but after decimation we can see from Fig. 4(b) that most pixels reflecting intensive changing along the edges are lost, such that the gradient based on the sampled version cannot show the real information of this edge. Since decimation also causes frequency aliasing, phase-based edge detection methods are not effective. Inspired by the motion compensation technique performed in the overcomplete wavelet space [20], orientation distribution is estimated in the overcomplete wavelet space to avoid problems brought by decimation.

To measure the accuracy of edge detection between the wavelet and the overcomplete wavelet subbands by the gradient-based method, Fig. 5(a) and (b) shows the Canny edge detection result on the wavelet and the overcomplete HH subband of one frame in *Foreman*, respectively, and Fig. 5(b) and (d) shows the histograms of orientation distribution in the wavelet and the overcomplete HH subband, respectively. From these figures, we can see that the decimation system in the regular wavelet decompositions blurs all the continuities along the edges and thus the histogram of distribution is hardly accurate whereas all the details and continuities are completely retained with few blurs in the overcomplete wavelet domain.

Phase-based edge detection methods such as phase congruency metric determines dirpixels through the points where the Fourier components are highly consistent in phase, and its 2-D directional version is shown in (3) [19]. In our scenario, the local overcomplete Gabor wavelet component at location $\mathbf{x}$ and direction d can be described by complex vectors which add head to tail with amplitudes $A_{dn}(\mathbf{x})$, phase angles $\phi_{dn}(\mathbf{x})$ and weighted mean phase angle $\bar{\phi}_d(\mathbf{x})$. The term $W_d(\mathbf{x})$ is a weighted factor of frequency spread, and ϵ is a small constant incorporated to avoid division by zero. T_{dn} is used as a threshold to cancel the noise influence since the operator $\lfloor\cdot\rfloor$ only preserves the positive operand, and otherwise returns zero

$$PC(\mathbf{x}) = \frac{\sum_d \sum_n W_d(\mathbf{x}) \lfloor A_{dn}(\mathbf{x}) \Delta\Phi_{dn}(\mathbf{x}) - T_{dn} \rfloor}{\sum_d \sum_n A_{dn}(\mathbf{x}) + \epsilon} \tag{3}$$

where the sensitive phase deviation function $\Delta\Phi_{dn}$ is defined as

$$\Delta\Phi_{dn}(\mathbf{x}) = \cos\left(\phi_{dn}(\mathbf{x}) - \bar{\phi}_d(\mathbf{x})\right) - \left|\sin\left(\phi_{dn}(\mathbf{x}) - \bar{\phi}_d(\mathbf{x})\right)\right|. \tag{4}$$

Once obtaining the dirpixels, we set up a full binary tree with 2^p leaves to represent the cumulative histogram of dirpixels in each uniform directional interval from $(-\frac{\pi}{2} +$

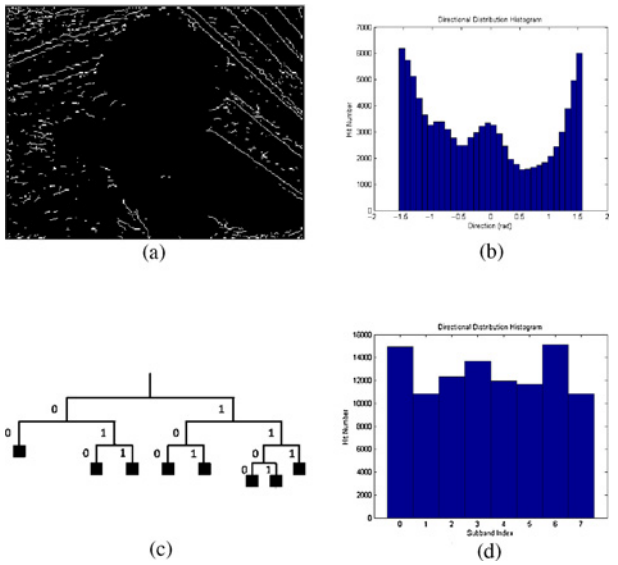

Fig. 6. Directional distribution and NSBT shaping. (a) Phase congruency measurement in the overcomplete HH subband. (b) Directional distribution estimated at 32 uniform directional intervals in the overcomplete HH subband. (c) NSBT derived from directional distribution estimation. (d) Directional distribution on NSBT.

$i\frac{\pi}{2^p}, -\frac{\pi}{2}+(i+1)\frac{\pi}{2^p}](i=0,1,\cdots,2^p-1)$ and then do tree-pruning to equalize the dirpixels on each leaf. In our tree-pruning process, two adjacent leaves with the least dirpixels which are extended from a common parent node are always selected to merge till the number of leaves reaches the target value.

Fig. 6(b) shows the directional distribution of one overcomplete HH subband [see Fig. 6(a)] in the *Foreman* sequence. The full binary tree is initially set to $2^5 = 32$ leaves, and the dirpixels in each interval are observed in a nonuniform distribution. After tree-pruning, the full binary tree becomes an NSBT with eight leaves shown in Fig. 6(c), where leaves are corresponding to nonuniform intervals but with basically equivalent dirpixels [see Fig. 6(d)].

In summary, all the edge detection methods including both pixel-based and phase-based are efficient in the overcomplete wavelet domain. Therefore, as the first step of dual multiresolution transform, orientation distribution estimation result can be obtained by edge detection in the overcomplete wavelet LH, HL and HH subbands of the video frame.

B. Multiresolution Analysis

Scale multiresolution serves as the first multiresolution in the dual multiresolution transform by using the wavelet decomposition. The scale multiresolution framework for 2-D wavelets is extended from 1-D scaling and wavelet functions (φ and ψ) [17]: $\varphi_j^{(2)}(\mathbf{t}) = \varphi_j(t_1)\varphi_j(t_2)$, $\psi_j^{(2),1}(\mathbf{t}) = \varphi_j(t_1)\psi_j(t_2)$, $\psi_j^{(2),2}(\mathbf{t}) = \psi_j(t_1)\varphi_j(t_2)$, $\psi_j^{(2),3}(\mathbf{t}) = \psi_j(t_1)\psi_j(t_2)$ where the family of $\{\varphi_{j,\mathbf{k}}^{(2)}(\mathbf{t}) = 2^{-j}\varphi^{(2)}(2^{-j}\mathbf{t}-\mathbf{k})\}$ and $\{\psi_{j,\mathbf{k}}^{(2),i}(\mathbf{t}) = 2^{-j}\psi^{(2),i}(2^{-j}\mathbf{t}-\mathbf{k})\}_{i=1,2,3}$ form an orthonormal basis of $\mathcal{V}_j^2$ and $\mathcal{W}_j^{2,i}$ at the scale 2^j, respectively. Three orthogonal subspaces $\mathcal{W}_j^{2,1} = \mathcal{V}_j \otimes \mathcal{W}_j$, $\mathcal{W}_j^{2,2} = \mathcal{W}_j \otimes \mathcal{V}_j$ and $\mathcal{W}_j^{2,3} = \mathcal{W}_j \otimes \mathcal{W}_j$ construct the detail space $\mathcal{W}_j^2$ by $\mathcal{W}_j^2 = \oplus_{i=1}^3 \mathcal{W}_j^{2,i}$, which is connected to the approximation space $\mathcal{V}_j^2 = \mathcal{V}_j \otimes \mathcal{V}_j$ as the complementary for the next scale: $\mathcal{V}_{j+1}^2 = \mathcal{V}_j^2 \oplus \mathcal{W}_j^2$. From the dyadic property of wavelet basis functions, basis of the jth scale approximation space can be split into $(j-1)$th scale approximation space and detail subspaces by filtering with quadrature mirror filters.

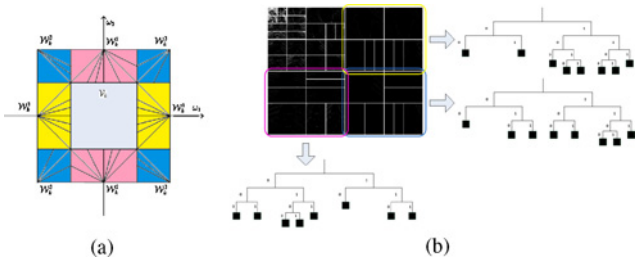

Fig. 7. Decomposition in the frequency and pixel domain. (a) Directional subspaces. (b) Frame of decomposition.

As the second multiresolution vehicle, orientation multiresolution is designed to make each subband after the dual multiresolution transform contains nearly equivalent amount of dirpixels within the subband bound. Equivalently, a narrow directional region with dense directional spectrum information deserves the same-sized subband of a wide directional regions with sparse directional spectrum information. Moreover, since the scale factor of different wavelet highpass subbands is two, it is reasonable to halve the number of orientation resolutions from fine to coarser scales. Next, we apply NUDFB to the detail multiresolution subspaces $\mathcal{W}_k^s$ by employing partition operators (the superscript "2" is omitted since all the following discussion are based on the 2-D case, and properties of the partition operator can be seen in Appendix A).

Proposition 1: Suppose we divide $\mathcal{W}_k^s$ into L subspaces with Q different orientation multiresolutions ($\{r_1, r_2, \cdots, r_Q\}$), and partition operator is applied to $\mathcal{W}_k^s$ for $r_{min} = min\{r_1, r_2, \cdots, r_Q\}$ times iteratively according to (20), that is

$$\begin{aligned}\mathcal{W}_k^s &= \left(\delta^{d_1}\mathcal{D}_{k,p_1}^{(r_{min})}\right)\bigoplus\cdots\bigoplus\left(\delta^{d_Q}\mathcal{D}_{k,p_Q}^{(r_{min})}\right)\\ &= \left(\bigoplus_{p_{d_1}=2^{d_1}p_1}^{2^{d_1}p_1+2^{d_1}-1}\mathcal{D}_{k,p_{d_1}}^{(r_{min}+d_1)}\right)\bigoplus\left(\bigoplus_{p_{d_2}=2^{d_2}p_2}^{2^{d_2}p_2+2^{d_2}-1}\mathcal{D}_{k,p_{d_2}}^{(r_{min}+d_2)}\right)\\ &\quad\bigoplus\cdots\bigoplus\left(\bigoplus_{p_{d_Q}=2^{d_Q}p_Q}^{2^{d_Q}p_Q+2^{d_Q}-1}\mathcal{D}_{k,p_{d_Q}}^{(r_{min}+d_Q)}\right)\end{aligned} \tag{5}$$

where $0 \le p_i \le 2^{r_{min}}-1$, $d_i = r_i - r_{min}$, $i = 1, 2, \cdots, Q$, $\sum_{i=1}^{Q} 2^{d_i} = L$, and $\{\lambda_{k,\mathbf{n},p_i}^{(r_{min})}\}$, the basis of $\mathcal{D}_{k,p_i}^{(r_{min})}$, can be divided into the family of $\{\lambda_{k,\mathbf{n},p_{d_i}}^{(r_{min}+d_i)}\}$ which forms the basis of the subspace $\mathcal{D}_{k,p_{d_i}}^{(r_{min}+d_i)}$. Proposition III-B can be illustrated in Fig. 7(a).

Thus, for any particular wavelet highpass subspace $\mathcal{W}_k^s$, there exists an NSBT structure containing some of its basis $\{\lambda_{k,\mathbf{n},p}^{(l)}\}_{\mathbf{n}\in\mathbb{Z}^2, 0\le p\le 2^l-1}$ to project itself into several nonuniform directional subspaces, meaning that the orientation multiresolution is achieved. Such a dual multiresolution decomposition example is provided in Fig. 7(b).

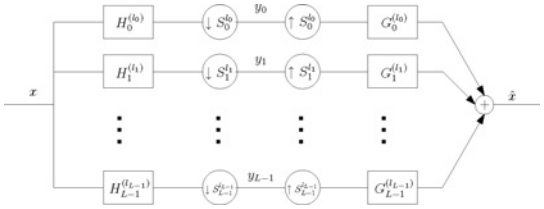

Fig. 8. Multi-channel view of NUDFB that has L channels with equivalent filters and sampling lattices.

The above results inspire us to consider a multi-channel view that resorts to NUDFB with L subbands as L parallel channels with equivalent filters and diagonal sampling lattices. Because of the partition operator, the leaf nodes of NSBT do not share the same sampling density, as in the p-th channel, which is $det(S_p^{(l_p)}) = 2^{l_p}$, $p = 0, 1, \cdots, L-1$. A typical multi-channel view of NUDFB is demonstrated in Fig. 8.

Let $R_p^{(l_p)}$ denote the support region associated to the analysis and synthesis filters $H_p^{(l_p)}(\omega)$ and $G_p^{(l_p)}(\omega)$. Owing to the property of DFB [21], $R_p^{(l_p)}$ tiles the 2-D frequency plane. Consequently, such a multi-channel view is complete.

Proposition 2: Suppose that the filter bank in Fig. 8 is perfect reconstructable. Then any 2-D signal in $L^2(\mathbb{Z}^2)$ can be uniquely represented as

$$x[\mathbf{n}] = \sum_{p=0}^{L-1} \sum_{\mathbf{m}\in\mathbb{Z}^2} y_p[\mathbf{m}] g_p^{(l_p)}[\mathbf{n} - S_p^{(l_p)}\mathbf{m}] \tag{6}$$

where

$$y_p[\mathbf{m}] = \langle x[\mathbf{n}], h_p^{(l_p)}[S_p^{(l_p)}\mathbf{m} - \mathbf{n}] \rangle. \tag{7}$$

Therefore, the family of $\{g_p^{(l_p)}[\mathbf{n} - S_p^{(l_p)}\mathbf{m}]\}_{0\le p<L,\mathbf{m}\in\mathbb{Z}^2}$ and $\{h_p^{(l_p)}[S_p^{(l_p)}\mathbf{m} - \mathbf{n}]\}_{0\le p<L,\mathbf{m}\in\mathbb{Z}^2}$ are called the dual basis for all the discrete signals in $L^2(\mathbb{Z}^2)$ where p denotes the direction and $\mathbf{m}$ denotes the position index, respectively.

Proposition 3: If we substitute $x[\mathbf{n}] = g_p^{(l_p)}[\mathbf{n} - S_p^{(l_p)}\mathbf{m}]$ in (6), and call to a remembrance of the uniqueness of representation, we will get such a biorthogonal relationship between these two basis as

$$\langle g_p^{(l_p)}[\mathbf{n} - S_p^{(l_p)}\mathbf{m}], h_{p'}^{(l_{p'})}[S_{p'}^{(l_{p'})}\mathbf{m}' - \mathbf{n}] \rangle = \delta[p - p']\delta[\mathbf{m} - \mathbf{m}']. \tag{8}$$

Fig. 9(a) and (b) shows an example of the frequency and time response of NUDFB shown in Fig. 3. A "23 − 45" biorthogonal filter bank designed by [22] is used in the DFB stage. From these figures, we can see that the basis still keeps the characteristic of anisotropy.

C. NUDFB Design

There are two ways to design a 2-D directional filter bank: [23] introduces the original DFB construction by using the diamond-shaped filters to process the pre-modulated source signals and employs complex tree expanding rules to rearrange the split subbands, while [24] only uses the fan filter (shift-modulated version of diamond-shaped filter) and tactfully

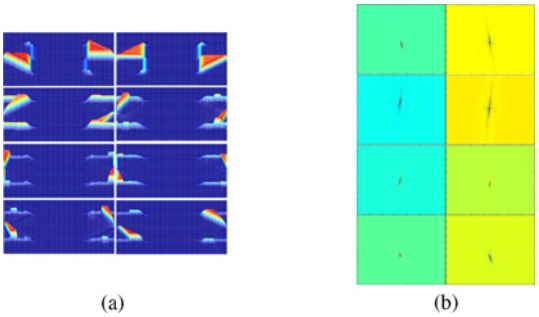

Fig. 9. Impulse response of NUDFB. (a) Frequency domain. (b) Time domain.

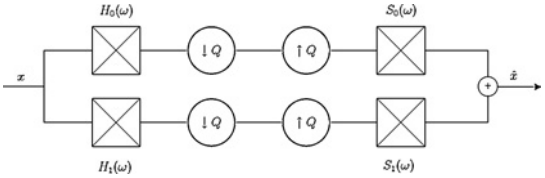

Fig. 10. General 2-D nonseparable filter bank.

decomposes the two-determinant sampling matrix into the Smith form to establish quincunx filter banks (QFB) with a symmetric binary tree (SBT) structure, which simplifies the construction process of DFB. Such a basic structure of 2-D nonseparable filter bank can be seen in Fig. 10.

Here we consider DFB with a single depth level as the basic element of NUDFB. For those regions which require better orientation resolutions, deeper levels of decompositions are spanned under the parent node, which provides sparser sampling lattices and support regions, and finally an NSBT structure will be expanded. The DFBs with SBT and NSBT structures are conceptually similar, with the main difference that the nodes in the NSBT case may not have the same number of offsprings as what happens in the SBT case during the construction of the filter banks.

In the general 2-D nonseparable filter bank design process, we need the quincunx sub-lattices with determinant of two [25] to satisfy the requirement of critical sampling, such as $Q_0 = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$, $Q_1 = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. Decompose Q_0 and Q_1 into the Smith form, we may get $Q_0 = R_1 D_0 R_2 = R_2 D_1 R_1$ and $Q_1 = R_0 D_0 R_3 = R_3 D_1 R_0$, where the unimodular matrices $R_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $R_1 = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$, $R_2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, $R_3 = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$ and the diagonal matrices $D_0 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$, $D_1 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$.

For the first and second levels of decomposition, Q_0 and Q_1 are used as the sampling lattice respectively. Since $Q_0 Q_1 = 2I$, the overall 2-D sampling density after these two levels of decompositions are critical. From the third level of decomposition, a pair of unimodular sampling lattices $R_i\,(i = 0, 1, 2, 3)$ are cascaded before and after the QFB to provide finer direction resolution. Since the operation of sampling and filtering can be swapped by multirate noble identities [25], we can obtain the overall sampling lattice as

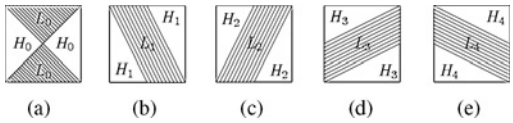

Fig. 11. Frequency supporting regions of equivalent maximally decimated filters of NUDFB with the fan filters. (a) Fan filter with sampling lattice Q_0 and Q_1 in the first and second levels. (b)–(d) Parallelogram filters with sampling lattice P_1 to P_4 in the third level.

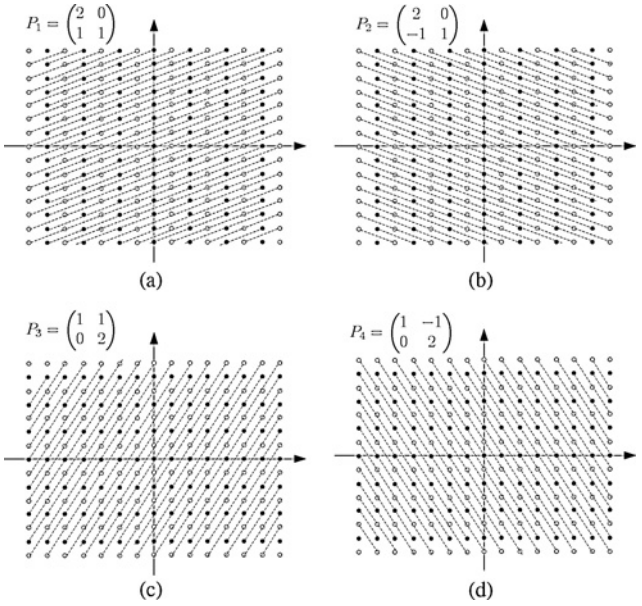

Fig. 12. Equivalent directional sampling lattice of (a) P_1, (b) P_2, (c) P_3, and (d) P_4.

$P_1 = R_0Q_0 = D_0R_2 = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$, $P_2 = R_1Q_1 = D_0R_3 = \begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix}$, $P_3 = R_2Q_1 = D_1R_0 = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$, and $P_4 = R_3Q_0 = D_1R_1 = \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix}$. All the matrices $P_i\,(i = 1, 2, 3, 4)$ and Q_0 have the determinant of 2, which constitute the sampling lattices of the maximally decimated filter banks with the fan filters, and their supporting regions are shown in Fig. 11. The filter banks with lattice $P_i\,(i = 1, 2, 3, 4)$ which is shown in Fig. 12 are called parallelogram filters.

A multilevel NUDFB can be considered as a cascading structure of the quincunx or parallelogram filters, while each basic element and its associated orientation range can be abstracted to a leaf node of NSBT. From a parent node to its two child nodes, either "0" or "1" is indexed so that every branch in the NSBT can be uniquely labeled by a binary sequence. A deeper leaf node of the NSBT corresponds to a directional filter with finer orientation resolution, and the entire path from the root to the node is labeled as a longer binary sequence. Initially, the NSBT is constructed as a full binary tree with each leaf node covering a well-distributed orientation range, and then a balance algorithm described in Algorithm 1 is designed to prune the binary tree to maintain nearly equal number of dirpixels within the orientation range of each leaf node. After obtaining the NSBT structure of the NUDFB, we can refer to the binary index of each branch to decompose the video frame by cascading the quincunx or parallelogram filters and get the final subbands.

Fig. 3 gives a typical example of the analysis part of NUDFB. Assuming that the result of the orientation distribution estimation in the overcomplete wavelet domain shows the need of dividing the whole wavelet subspace into eight directional frequency subbands, according to the criterion that each subband contains nearly equivalent amount of directional pixels, the first two levels of the filter banks can divide the frequency domain into four coarse directional subbands, and another four finer directional subbands are elaborated by the parallelogram filter banks of H_i and L_i, $i = 1, 2$.

Every NUDFB is fulfilled through the topology structure of NSBT where each node possesses two 2-D nonseparable filters as its children. If we can perfectly reconstruct every branch in the NSBT, the whole filter bank can be perfectly reconstructed. Because only five different directional filters are used, all the binary filter banks must be designed to guarantee that the whole filter bank is perfectly reconstructed [24].

From the supporting regions of the five filters, we have

$$\begin{cases} H_0(\omega) = L_0(\omega - 2\pi Q_0^{-T}\mathbf{k}) \\ H_i(\omega) = L_i(\omega - 2\pi P_i^{-T}\mathbf{k}) \end{cases}$$

where $i = 1, 2, 3, 4$. The polyphase representation of multidimensional filter banks gives the simple conclusion on perfect reconstruction, for example, type I polyphase form of $L_0(\omega)$ and $H_0(\omega)$ are as follows:

$$\begin{cases} L_{0,0}(\omega) = \sum_{\mathbf{k}\in\mathcal{N}(Q_0^T)} L_0(Q_0^{-T}(\omega - 2\pi\mathbf{k})) \\ L_{0,1}(\omega) = \sum_{\mathbf{k}\in\mathcal{N}(Q_0^T)} e^{j(Q_0^T(\omega-2\pi\mathbf{k}))^T\mathbf{k}} L_0(Q_0^{-T}(\omega - 2\pi\mathbf{k})) \\ H_{0,0}(\omega) = \sum_{\mathbf{k}\in\mathcal{N}(Q_0^T)} H_0(Q_0^{-T}(\omega - 2\pi\mathbf{k})) \\ H_{0,1}(\omega) = \sum_{\mathbf{k}\in\mathcal{N}(Q_0^T)} e^{j(Q_0^T(\omega-2\pi\mathbf{k}))^T\mathbf{k}} H_0(Q_0^{-T}(\omega - 2\pi\mathbf{k})). \end{cases} \tag{9}$$

We can infer that $H_{0,0}(\omega) = L_{0,0}(\omega)$ and $H_{0,1}(\omega) = -L_{0,1}(\omega)$ since

$$\begin{cases} H_{0,0}(\omega) = \sum_{\mathbf{k}\in\mathcal{N}(Q_0^T)} L_0(Q_0^{-T}(\omega - 2\pi\mathbf{k}) - 2\pi Q_0^{-T}\omega) \\ \qquad = L_{0,0}(\omega) \\ H_{0,1}(\omega) = \sum_{\mathbf{k}\in\mathcal{N}(Q_0^T)} e^{j(Q_0^T(\omega-2\pi\mathbf{k}))^T\mathbf{k}} e^{-j2\pi Q_0^{-T}\mathbf{k}} \\ \qquad L_0(Q_0^{-T}(\omega - 2\pi\mathbf{k}) - 2\pi Q_0^{-T}\mathbf{k}) \\ \qquad = -\sum_{\mathbf{k}\in\mathcal{N}(Q_0^T)} e^{j(Q_0^T(\omega-2\pi\mathbf{k}))^T\mathbf{k}} L_0(Q_0^{-T}(\omega - 2\pi\mathbf{k})) \\ \qquad = -L_{0,1}(\omega). \end{cases} \tag{10}$$

In conclusion, all the type I polyphase forms of the analysis filters L_i and $H_i\,(i = 0, 1, 2, 3, 4)$ follow the constraint that

$$\begin{cases} L_i(\omega) = L_{i,0}(\omega) + e^{-j\omega^T \mathbf{k}} L_{i,1}(\omega) \\ H_i(\omega) = H_{i,0}(\omega) + e^{-j\omega^T \mathbf{k}} H_{i,1}(\omega) \\ \qquad = L_{i,0}(\omega) - e^{-j\omega^T \mathbf{k}} L_{i,1}(\omega). \end{cases} \tag{11}$$

From the multi-dimensional filter bank theory [25], the synthesis filter for a perfect reconstruction system must have the same spectrum range with the corresponding analysis filter. Likewise, the type II polyphase decomposition of the synthesis filters F_i and G_i $(i = 0, 1, 2, 3, 4)$ follows the constraint that

$$\begin{cases} F_i(\omega) = e^{-j\omega^T \mathbf{k}} F_{i,0}(\omega) + F_{i,1}(\omega) \\ G_i(\omega) = e^{-j\omega^T \mathbf{k}} F_{i,0}(\omega) - F_{i,1}(\omega). \end{cases} \tag{12}$$

The perfect reconstruction in the polyphase domain can be achieved if and only if the following conditions are satisfied:

$$L_i(\omega) F_i(\omega)^T = e^{j\omega^T \mathbf{l}} I \tag{13}$$

where $\mathbf{l}$ is an arbitrary vector, and

$$\begin{cases} L_i(\omega) = \begin{pmatrix} L_{i,0}(\omega) & L_{i,1}(\omega) \\ L_{i,0}(\omega) & -L_{i,1}(\omega) \end{pmatrix} \\ F_i(\omega) = \begin{pmatrix} F_{i,0}(\omega) & F_{i,1}(\omega) \\ F_{i,0}(\omega) & -F_{i,1}(\omega) \end{pmatrix} \end{cases} \tag{14}$$

for $i = 0, 1, 2, 3, 4$; substitute all these definitions into (13), we have

$$\begin{cases} L_{i,0}(\omega) F_{i,0}(\omega) + L_{i,1}(\omega) F_{i,1}(\omega) = e^{j\omega^T \mathbf{l}} = c \\ L_{i,0}(\omega) F_{i,0}(\omega) - L_{i,1}(\omega) F_{i,1}(\omega) = 0. \end{cases} \tag{15}$$

Without loss of generality, assuming the constant $c = 2$, we finally obtain the condition of perfect reconstruction of one pair of siblings of the filter bank with a binary tree structure as

$$\begin{cases} F_{i,0}(\omega) = \dfrac{1}{L_{i,0}(\omega)} \\ F_{i,1}(\omega) = \dfrac{1}{L_{i,1}(\omega)}. \end{cases} \tag{16}$$

Algorithm 1 The NSBT Balancing Algorithm

Input: Video frame $\mathbf{x}$, Decomposition level n
Output: NSBT structure with balanced number of dirpixels
A. Design the NSBT structured decomposition path:
Decompose the input image in overcomplete wavelet domain;
for *processing all three overcomplete wavelet subbands* **do**
  Obtain the orientation of each pixel (dirpixel) $PC(\mathbf{x})$ via phase congruency method;
end
Set current processing scale $i \longleftarrow 3$;
Set current decomposition level $m \longleftarrow 2$;
for $i = 3; i \leq n; i++$ **do**
  for *processing three wavelet subbands in i-th scale* **do**
    Set number of subbands $num \longleftarrow 2^{m+2}$;
    Establish a full binary tree with num leaf nodes in i-th scale, index the leaf node from $k = 0, \cdots, num - 1$ by binary;
    Divide the orientation range of $[-\pi/2, \pi/2]$ into num pieces, k-th leaf node on the full binary tree cumulates the number of dirpixels in the orientation range of $[-\pi/2 + k(\pi/num), -\pi/2 + (k+1)(\pi/num)]$;
    while $num > 2^m$ **do**
      Look through all of the leaf nodes, find two adjacent leaf nodes with least number of dirpixels, prune the tree by deleting these two nodes and leaving their parent node as a new leaf node with truncated binary index;
      $num \longleftarrow num - 1$;
    end
    $m \longleftarrow m + 1$;
  end
end
for $i = 1; i \leq 2; i++$ **do**
  Establish a full binary tree with 2^i leaf nodes in i-th scale without pruning.
end

IV. Experimental Results

A. Nonlinear Approximation (NLA)

Video coding always introduces lossy and quantization noise to the coefficients of spatial transformed 2-D signals, while most coding schemes erase the relatively small coefficients, and preserve most significant coefficients with quantization noise. Hence, we select M-most significant coefficients in the transform domain, and do the inverse spatial transform to obtain the reconstructed image, and observe what kind of edge and texture information that NUDFB efficiently captures.

Fig. 13 gives the PSNR results of NLA versus M retained coefficients tested on a sampled frame of several video test sequences by the dual multiresolution transform and compared with other transforms, e.g., HWD and DWT. For the test sequences *Coastguard* of CIF resolution (352×288) and *Barbara* of size 512×512, we decompose them into 4 scale levels in all of the transforms and $\{1, 2, 4, 8\}$ directional subbands from the coarsest to the finest scale for the directional transforms, e.g., dual multiresolution transform and HWD. For 4CIF sequence *City* and *Harbor*, we decompose them into 5 scales and $\{1, 2, 4, 8, 16\}$ directional subbands from the coarsest to the finest scale. Besides, for other test sequences such as *Flower*, *Tempete*, *Walk*, and *Crew*, the dual multiresolution transform provides comparable result to that of HWD transform and wavelets.

To show the visual results of NLA, we select $M = 4096$ most significant coefficients from the dual multiresolution

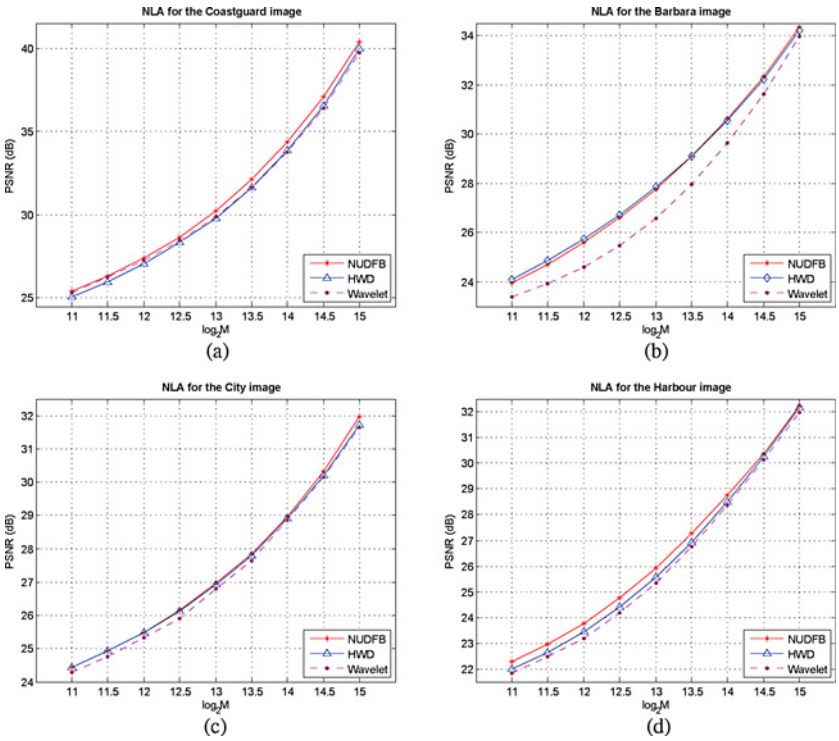

Fig. 13. Examples of the NLA PSNR results. (a) NLA results for *Coastguard* image. (b) NLA results for *Barbara* image. (c) NLA results for *City* image. (d) NLA results for *Harbor* image.

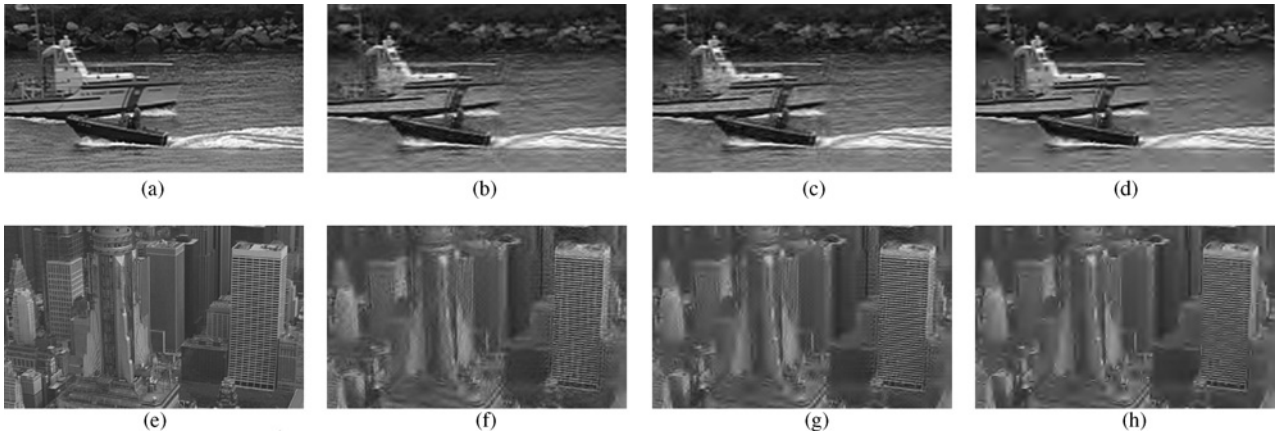

Fig. 14. NLA reconstruction results for one frame of *Coastguard* and *City* with $M = 4096$ most significant coefficients in the transform domain. (a) Original *Coastguard* frame. (b) Dual multiresolution transform: PSNR = 27.40 dB. (c) HWD transform: PSNR = 27.05 dB. (d) DWT: PSNR = 27.28 dB. (e) Original *City* frame. (f) Dual multiresolution transform: PSNR = 25.49 dB. (g) HWD transform: PSNR = 25.48 dB. (h) DWT: PSNR = 25.34 dB.

transform, HWD and wavelet domain of the *Coastguard* and *City* sequences and get their reconstructed approximation in Fig. 14. It can be seen that the proposed dual multiresolution transform produces less artifacts than HWD. Like the HWD transform [10], we can observe that dual multiresolution transform has better capability of capturing the curving edges, directional textures and other details with the same amount of significant coefficients comparing with DWT.

B. Comparison With WSVC for Scalability

We compare our proposed DMSVC scheme with the latest WSVC scheme under the platform of MSRA 3-D wavelet video coder *VidWav* [26], and test the combined scale and time scalability in the experiments. The reference software is configured to multiplex five layers with different spatial and time scalabilities into one bitstream. The video frames in one GOP are temporally decomposed into five temporal subbands, each temporal subband is further spatially decomposed by dual multiresolution transform with NUDFB into a group of subbands according to the orientation distribution estimation. In order to show the performance differences between NUDFB and UDFB in spatial decomposition, we also incorporate the HWD into the SVC framework to develop the HWDSVC scheme as a reference. All schemes provide 3 scales of decomposition for CIF sequences and 4 scales for 4CIF sequences. From the coarsest to the finest scale, we

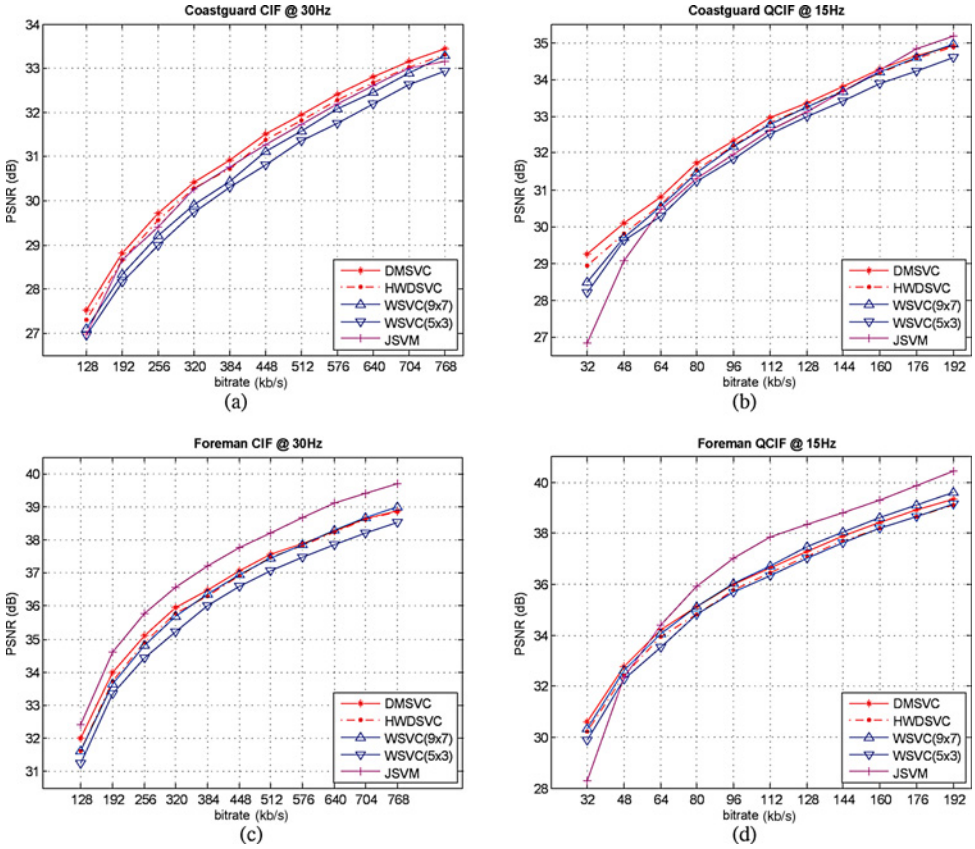

Fig. 15. Scalability performance comparison by PSNR on CIF sequences. (a) *Coastguard* CIF sequence at frame rate of 30 Hz. (b) *Coastguard* QCIF sequence at frame rate of 15 Hz. (c) *Foreman* CIF sequence at frame rate of 30 Hz. (d) *Foreman* QCIF sequence at frame rate of 15 Hz.

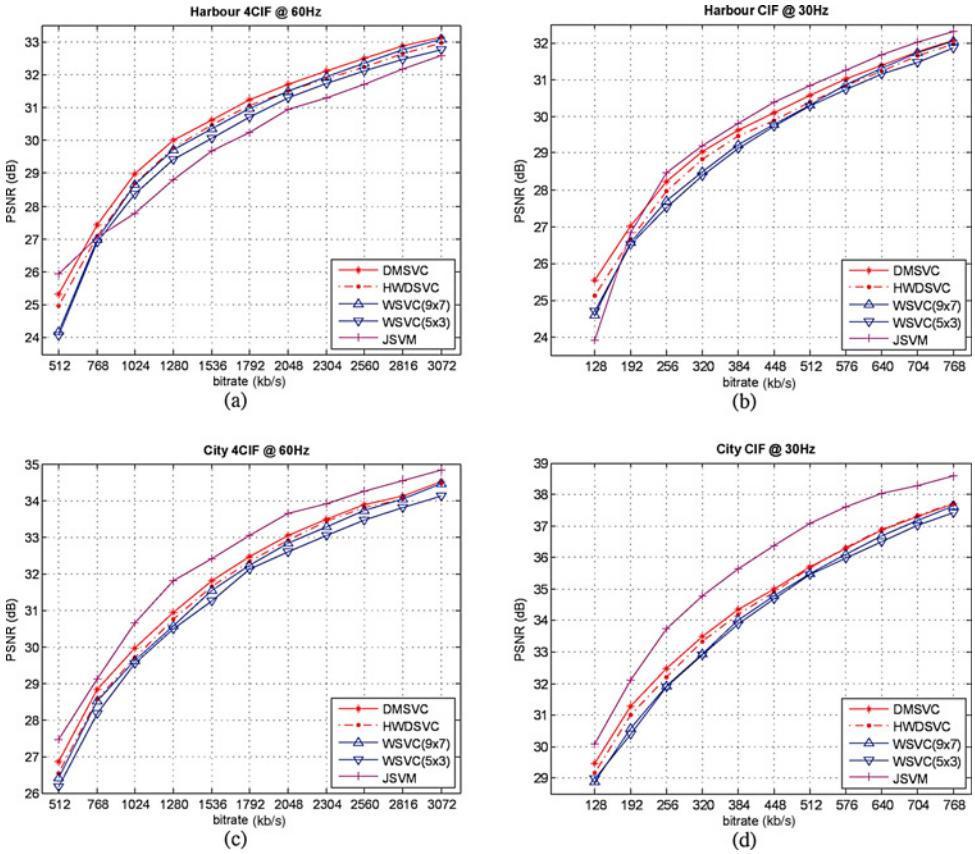

Fig. 16. Scalability performance comparison by PSNR on 4CIF sequences. (a) *Harbor* 4CIF sequence at frame rate of 60 Hz. (b) *Harbor* CIF sequence at frame rate of 30 Hz. (c) *City* 4CIF sequence at frame rate of 60 Hz. (d) *City* CIF sequence at frame rate of 30 Hz.

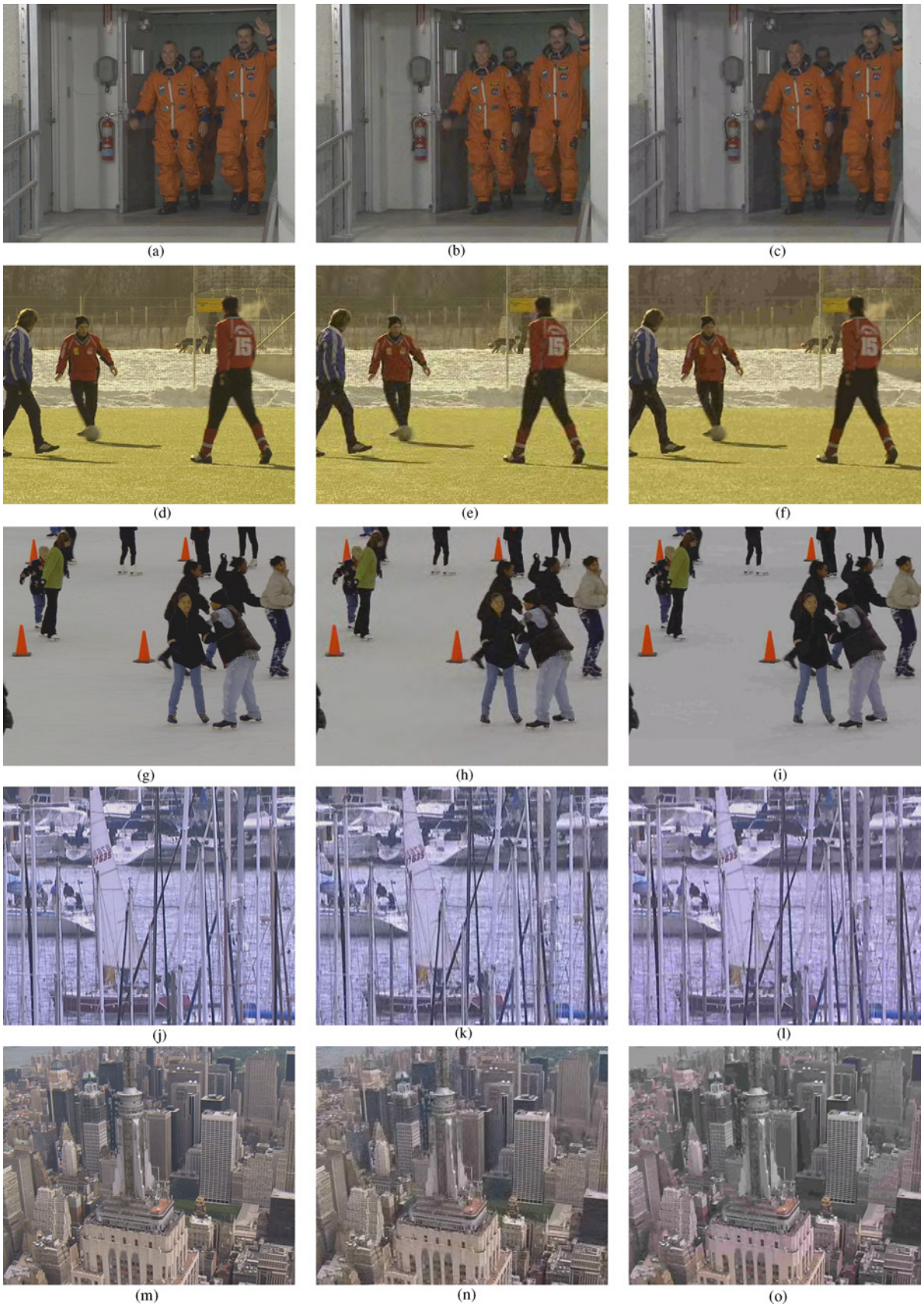

Fig. 17. Visual results under 1024 kb/s. From left to right and from top to bottom are the original and reconstructed 4CIF frame of *Crew*, *Soccer*, *Ice*, *Harbor*, and *City* by DMSVC and WSVC, respectively. (a) Original 4CIF frame of *Crew*. (b) PSNR = 30.09 dB. (c) PSNR = 29.52 dB. (d) Original 4CIF frame of *Soccer*. (e) PSNR = 29.15 dB. (f) PSNR = 28.36 dB. (g) Original 4CIF frame of *Ice*. (h) PSNR = 37.02 dB. (i) PSNR = 36.59 dB. (j) Original 4CIF frame of *Harbor*. (k) PSNR = 30.07 dB. (l) PSNR = 29.63 dB. (m) Original 4CIF frame of *City*. (n) PSNR = 34.76 dB. (o) PSNR = 34.47 dB.

split {1, 4, 8} directional subbands in CIF case, and {1, 2, 4, 8} directional subbands in 4CIF case for DMSVC and HWDSVC.

Figs. 15 and 16 give the rate-distortion performance for the combined spatial and temporal scalability between DMSVC, HWDSVC and WSVC on CIF and 4CIF sequences, respectively. The proposed SVC frameworks with the directional decomposition, e.g., DMSVC and HWDSVC, bring out promising results for the sequences with significant directional textures, and the performance of DMSVC is higher than that of HWDSVC since NUDFB brings less artifact than UDFB. Specifically, when the *Harbor* sequence is coded at the bit rate of 512 kb/s, DMSVC provides up to 1.2 dB PSNR improvement over the WSVC scheme. For other sequences such as *Coastguard*, *Foreman*, and *City*, DMSVC also shows comparable performance to WSVC.

The visual effects of 4CIF counterparts coded at 1024 kb/s are shown in Fig. 17. Since the dual multiresolution transform has a stronger ability on capturing curve smooth discontinuities and smaller error decay order on NLA, more details on textures have been preserved after reconstruction.

C. Comparison With H.264/SVC for Scalability

We also provide the performance comparison between DMSVC and the H.264/SVC scheme in the aforementioned figures, and adopt *JSVM6.5* from [27] as the reference software and the Palma CE conditions [28] as the configuration parameters. Specifically, the GOP size for temporal scalability is set to support five temporal layers, while keeping three spatial layers for 4CIF sequences and two spatial layers for CIF sequences. For each specific spatial layer, the following parameters for temporal decomposition have been enabled: closed-loop prediction structure of SVC, adaptive QP selection, the mode of intra-macroblock and inter-layer prediction. Besides, the pixel range of motion search is set 64 as the proposed DMSVC. It can be obviously seen that, with the increase of video resolution and the abundance of complex texture structure, DMSVC shows an increasingly approximate objective performance to H.264/SVC. For the sequences full of directional information, e.g., edges and textures, sketch lines, and contours, perceptual quality index such as SSIM has illustrated a better effect than H.264/SVC, especially in the low bitrate range. It sufficiently justifies the reconstructed video frames of the proposed DMSVC have better structure similarity and visual quality than other SVC schemes.

To a great extent, the H.264/SVC scheme is dependent on the temporal decomposition stage: close-loop hierarchical B-picture rather than open-loop MCTF in WSVC and the proposed DMSVC. The open-loop coder control of MCTF would accumulate the quantization errors [29] and thus reduce the coding efficiency. Moreover, the divergence between DMSVC and H.264/SVC on spatial scalability is either frame-based or macroblock-based. Although DMSVC uses a block-based motion model like H.264/SVC, it can not support the intra-mode because the spatio-temporal decomposition is enabled to put a group of frames together within the coding passes in a global manner. Hence, the single layer DMSVC or WSVC only supports open-loop encoding/decoding without in-loop deblocking filter. It has been out of scope to pursue a sparser spatial decomposition and representation in generic video coding and design an adaptive orientational multiresolution transform and nonuniform directional filterbank beyond the traditional trajectory. However, the appropriate incorporation of local compensation would be investigated in the proposed dual multiresolution transform and the co-located NUDFB design in the future.

V. Conclusion

In order to capture the intrinsic geometric structure of the 2-D video signal and represent it more sparsely, we introduced a dual multiresolution transform with nonuniform directional filter banks into the current SVC framework with fully compatibility. The proposed spatial decomposition can select the anisotropic basis of multiscale and multidirection subspaces adaptively according to the orientation distribution histogram of the video frame and project the frame into these spaces. This paper has made two main contributions.

1) The proposal of nonuniform directional frequency decompositions under arbitrary scales which are fulfilled by a NSBT topology structure with NUDFB design. The proposed NUDFB provides a multiresolution on directions as well as wavelet filters provide a multiresolution on scales, and it is more flexible to statistically utilize the directional information in video frames to achieve a more efficient filter bank partition. In the dual multiresolution transform, the wavelet basis function in each scale is converted to an adaptive set of nonuniform directional basis. The NUDFB is fulfilled by arraying the topology structure of a NSBT, as a symmetric extension from a two channel filter bank. The paraunitary perfect reconstruction is provided through a polyphase identical form of filter bank, in terms of 2-D nonseparable filters from a 1-D prototype.
2) The development of a novel generic scalable video coding framework with the dual (scale and orientational) multiresolution transform, called DMSVC. Each temporal subband through MCTF is further decomposed into multiscale subbands, and the highpass wavelet subspaces are divided into an arbitrary number of directional subspaces in alignment with the orientation distribution via phase congruency to establish NSBT. Comparing with the isolated wavelet basis, our transform provides a greater correlated set of localized and anisotropic basis functions with video frames. The spatio-temporal subband coefficients are coded by a 3-D ESCOT entropy coding algorithm to match the structure of NSBT.

Appendix A
Properties of Orientation Multiresolution Analysis

Proposition 4: Given a kth order highpass wavelet subspace $\mathcal{W}_k^s$ ($s = 1, 2, 3$), it can be divided into 2^l orthogonal directional subspaces $\mathcal{D}_{k,p}^{(l)}$ by using the equivalent synthesis filter banks $G_{k,p}^l$ where $0 \le p < 2^l$ [7], [9]

$$\mathcal{W}_k^s = \bigoplus_{p=0}^{2^l-1} \mathcal{D}_{k,p}^{(l)}. \tag{17}$$

Proof: Let $\{\lambda^{(l)}_{k,\mathbf{n},p}\}_{\mathbf{n}\in\mathbb{Z}^2}$ be the basis for $\mathcal{D}^{(l)}_{k,p}$, and it satisfies

$$\lambda^{(l)}_{k,\mathbf{n},p} = \sum_{m\in\mathbb{Z}^2} d^{(l)}_p[\mathbf{m}-\mathbf{S}^{(l)}_p\mathbf{n}]\psi_{k,m}(t) \tag{18}$$

where $\mathbf{S}^{(l)}_p$ is the overall sampling lattice

$$\mathbf{S}^{(l)}_p = \begin{cases} diag(2^{l-1},2), & \text{if } 0\le p\le 2^{l-1}-1 \\ diag(2,2^{l-1}), & \text{if } 2^{l-1}\le p\le 2^l-1. \end{cases}$$

Since $\{d^{(l)}_p[\mathbf{m}-\mathbf{S}^{(l)}_p\mathbf{n}]_{0\le p\le 2^l-1,n\in\mathcal{Z}^2}\}$ are the coefficients of directional filter $G^l_{k,p}$, such that $\{\lambda^{(l)}_{k,\mathbf{n},p}\}_{\mathbf{n}\in\mathbb{Z}^2}$ is also the orthonormal basis of $\mathcal{W}^s_k$. ■

Proposition 5: Any directional subspaces $\mathcal{D}^{(l)}_{k,p}$ can be divided into two subspaces by using a first-order partition operator δ as follows:

$$\delta\mathcal{D}^{(l)}_{k,p} = \mathcal{D}^{(l+1)}_{k,2p}\oplus\mathcal{D}^{(l+1)}_{k,2p+1}. \tag{19}$$

Proof: From (18), we know that $\{\lambda^{(l)}_{k,\mathbf{n},p}\}_{\mathbf{n}\in\mathbb{Z}^2}$ can be divided into $\{\lambda^{(l+1)}_{k,\mathbf{n},2p}\}_{\mathbf{n}\in\mathbb{Z}^2}$ and $\{\lambda^{(l+1)}_{k,\mathbf{n},2p+1}\}_{\mathbf{n}\in\mathbb{Z}^2}$ with an extra level of filtering by a pair of equivalent quadrature mirror filters $G^{l+1}_{k,2p}$ and $G^{l+1}_{k,2p+1}$, while $\{\lambda^{(l+1)}_{k,\mathbf{n},2p}\}_{\mathbf{n}\in\mathbb{Z}^2}$ and $\{\lambda^{(l+1)}_{k,\mathbf{n},2p+1}\}_{\mathbf{n}\in\mathbb{Z}^2}$ can be spanned into mutually orthogonal subspaces $\mathcal{D}^{(l+1)}_{k,2p}$ and $\mathcal{D}^{(l+1)}_{k,2p+1}$ with finer orientation resolution, which shows the validity of the partition operator. In other words, from the orthonormal basis of a coarser orientation resolution l, we can obtain a set of two orthonormal basis in the finer orientation resolution $l+1$ by using the partition operator once, and it can be iteratively used. Moreover, if we consider the filter bank to be organized in a binary tree structure, the partition operator can be considered as the split operation. ■

Corollary 1: Obviously, an nth order partition operator δ^n can be inferred as

$$\delta^n\mathcal{D}^{(l)}_{k,p} = \bigoplus_{p_n=2^n p}^{2^n p+2^n-1}\mathcal{D}^{(l+n)}_{k,p_n}. \tag{20}$$

REFERENCES

[1] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[2] R. Xiong, J. Xu, F. Wu, and S. Li, "Barbell-lifting based 3-D wavelet coding scheme," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1256–1269, Sep. 2007.

[3] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. New York: Academic, 1998.

[4] W. Ding, F. Wu, X. Wu, S. Li, and H. Li, "Adaptive directional lifting-based wavelet transform for image coding," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 416–427, Feb. 2007.

[5] C. Chang and B. Girod, "Direction-adaptive discrete wavelet transform for image compression," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1289–1302, May 2007.

[6] J.-L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 670–684, Jun. 2002.

[7] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.

[8] Y. Lu and M. N. Do, "CRISP-contourlets: A critically sampled directional multiresolution image representation," in *Proc. 10th SPIE Conf. Wavelet Applicat. Signal Image Process.*, Aug. 2003, pp. 655–665.

[9] R. Eslami and H. Radha, "Wavelet-based contourlet transform and its application to image coding," in *Proc. ICIP*, vol. 5. Oct. 2004, pp. 3189–3192.

[10] R. Eslami and H. Radha, "A new family of nonredundant transforms using hybrid wavelets and directional filter banks," *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 1152–1167, Apr. 2007.

[11] T. Chen, "Nonuniform multirate filter banks: Analysis and design with an $\mathcal{H}_\infty$ performance measure," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 572–582, Mar. 1997.

[12] S. Venkataraman and B. C. Levy, "A comparison of design methods for 2-D FIR orthogonal perfect reconstruction filter banks," *IEEE Trans. Circuits Syst.*, vol. 42, no. 8, pp. 525–536, Aug. 1995.

[13] T. Chen and P. P. Vaidyanathan, "Multidimensional multirate filters and filter banks derived from 1-D filters," *IEEE Trans. Signal Process.*, vol. 41, no. 5, pp. 1749–1765, May 1993.

[14] J. Xu, Z. Xiong, S. Li, and Y. Zhang, "Three-dimensional embedded subband coding with optimized truncation (3-D ESCOT)," *Appl. Computat. Harmonic Anal. Special Issue Wavelet Applicat. Eng.*, vol. 10, pp. 290–315, May 2001.

[15] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1158–1170, Jul. 2000.

[16] L. Luo, F. Wu, S. Li, Z. Xiong, and Z. Zhuang, "Advanced motion threading for 3-D wavelet video coding," *Signal Process. Image Commun.*, vol. 19, no. 7, pp. 601–616, Aug. 2004.

[17] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[18] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Nov. 1986.

[19] P. Kovesi, "Phase congruency detects corners and edges," in *Proc. Int. Conf. DICTA*, 2001, pp. 711–724.

[20] X. Lin, "Scalable video compression via overcomplete motion compensated wavelet coding," *Signal Process. Image Commun. Special Issue Subband/Wavelet Interframe Video Coding*, vol. 19, no. 7, pp. 637–651, Aug. 2004.

[21] M. N. Do, "Directional multiresolution image representations," Ph.D. dissertation, Dept. Commun. Syst., Swiss Federal Instit. Technol., Lausanne, Switzerland, Dec. 2001.

[22] S. M. Phoong, C. W. Kim, P. P. Vaidyanathan, and R. Ansari, "A new class of two-channel biorthogonal filter banks and wavelet bases," *IEEE Trans. Signal Process.*, vol. 43, no. 3, pp. 649–665, Mar. 1995.

[23] R. H. Bamberger and M. J. T. Smith, "A filter bank for the directional decomposition of images: Theory and design," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 882–893, Apr. 1992.

[24] E. J. Candes and D. L. Donoho, "Curvelets: A surprisingly effective nonadaptive representation for objects with edges," in *Curve and Surface Fitting*. Nashville, TN: Vanderbuilt Univ. Press, 1999.

[25] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice Hall Signal Processing, 1993.

[26] R. Xiong, X. Ji, D. Zhang, J. Xu, G. Pau, M. Trocan, and V. Bottreau, *Vidwav Wavelet Video Coding Specifications*, ISO/IEC JTC1/SC29/WG11/M12339, Poznan, Poland, Jul. 2005.

[27] M. Wien and H. Schwarz, *Testing Conditions for SVC Coding Efficiency and JSVM Performance Evaluation*, document JVT-Q205, Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, Poznan, Poland, Jul. 2005.

[28] N. Adami, M. Brescianini, R. Leonardi, and A. Signoroni, "SVC CE1:STool: A native spatially scalable approach to SVC," ISO/IEC JTC1/SC29/WG11, 70th MPEG Meeting, Palma de Mallorca, Spain, Tech. Rep. M11368, Oct. 2004.

[29] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B pictures and MCTF," in *Proc. IEEE ICME*, Jul. 2006, pp. 1929–1932.

Hongkai Xiong (M'01–SM'10) received the Ph.D. degree in communication and information systems from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2003.

Since 2003, he has been with the Department of Electronic Engineering, SJTU where he is currently an Associate Professor. From December 2007 to December 2008, he was with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, as a Research Scholar. He has published over 90 international journal/conference papers. In SJTU, he directs the Intelligent Video Modeling Laboratory and multimedia communication area in the Key Laboratory of the Ministry of Education of China—Intelligent Computing and Intelligent System which is also co-granted by Microsoft Research, Beijing, China. His current research interests include source coding/network information theory, signal processing, computer vision and graphics, and statistical machine learning.

Dr. Xiong was the recipient of the New Century Excellent Talents in University Award in 2009. In 2008, he received the Young Scholar Award of Shanghai Jiao Tong University. He has served on various IEEE conferences as a technical program committee member. He acts as a member of the Technical Committee on Signal Processing of the Shanghai Institute of Electronics.

Lingchen Zhu received the B.S. degree in electronic engineering from Southeast University, Nanjing, China, and the M.S. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2008 and 2011, respectively.

His current research interests include multiscale geometric analysis, sparse coding, and their applications on image and video coding.

Nannan Ma received the B.S. degree in electronic engineering from the Wuhan University of Technology, Wuhan, China, in 2006 and the M.S. degree in communication and information systems from Shanghai Jiao Tong University, Shanghai, China, in 2009.

Currently, she is with Marvell Technology Group, Ltd., Shanghai. Her current research interests include subband coding theory, signal processing, and data compression.

Yuan F. Zheng (F'97) received the M.S. and Ph.D. degrees in electrical engineering from Ohio State University, Columbus, in 1980 and 1984, respectively. His undergraduate education was received at Tsinghua University, Beijing, China in 1970.

From 1984 to 1989, he was with the Department of Electrical and Computer Engineering, Clemson University, Clemson, SC. Since August 1989, he has been with Ohio State University where he is currently a Professor, and was the Chairman of the Department of Electrical and Computer Engineering from 1993 to 2004. From 2004 to 2005, he spent a sabbatical year with Shanghai Jiao Tong University, Shanghai, China, and continued to be involved as the Dean of the School of Electronic, Information and Electrical Engineering until 2008. His current research interests include two aspects. One is wavelet transform for image and video, and object classification and tracking, and the other is robotics which includes robotics for life science applications, multiple-robot coordination, legged walking robots, and service robots.

Dr. Zheng was and is on the editorial boards of five international journals. He received the Presidential Young Investigator Award from Ronald Reagan in 1986, and research awards from the College of Engineering of Ohio State University in 1993, 1997, and 2007. He and his students received the Best Conference and Best Student Paper Award a few times in 2000, 2002, and 2006, and received the Fred Diamond for Best Technical Paper Award from the Air Force Research Laboratory, Rome, NY, in 2006. He was appointed to the International Robotics Assessment Panel by the NSF, NASA, and NIH to assess robotics technologies worldwide in 2004 and 2005.