Adaptive video coding using mixed-domain filter banks having optimal-shaped subbands

Bo Liu and L. T. Bruton
Department of Electrical and Computer Engineering
The University of Calgary
Calgary, Alberta, Canada T2N 1N4
Tel:(403)220-4881 Email:bruton@enel.ucalgary.ca

ABSTRACT

This contribution presents a mixed-domain filter bank. This filter bank can realize arbitrary (and thereby optimal) spectral decomposition and therefore is very suitable for the efficient subband coding of video signals. A coding scheme using the proposed filter bank is considered, which takes advantage of the spectral distribution of the signal and is especially selective to motion.

I. INTRODUCTION

In recent years, multidimensional (MD) filter banks have been successfully used in subband coding of still images as well as video signals [1, 2]. The basic idea of subband coding is to split the frequency band of the signal into a number of subbands and then to code each subband signal separately, taking advantage of the distribution of the frequency spectrum of the signal and of the perceptual properties of the human visual system (HVS) to achieve high coding efficiency. Most of the previously reported work on subband coding of images and video signals has employed separable filters for computational and design simplicity. Filter banks using non-separable filters achieve more flexibility, such as directional capability [3, 4, 5]. However, a filter bank can only realize certain regular spectral decompositions because of the limitation on the region of support (ROS) of the passbands. In order to optimally make use of the distribution of the signal spectrum, as well as to take advantage of the HVS, filter bank systems which are able to realize efficient spectral decompositions are highly desirable.

Filter banks which are especially selective to motion (i.e. which decompose the signal into subbands covering the spectrum of motions and subbands excluding these motions) are considered to be efficient for signal decomposition and thus for signal coding. This is because video signals can usually be decomposed into the static background and moving objects. If the motions of objects in video are fairly smooth, as is usually the case in practice, these moving objects can be approximately considered as piecewisely linear trajectory (LT) signals, which are fully described by a respective linear motion at a constant velocity [6]. Further [6], the 3D frequency spectrum of an ideal LT signal is confined to a frequency-domain plane, which contains the origin, having its normal along the direction of the corresponding trajectory. The shape of the 3D frequency spectrum of short-temporal blocks of the piecewisely LT signal approximates a 3D frequency-domain plane, having its normal along the direction of the corresponding linear trajectory in the block. The static background is a special case of the LT signal and its 3D frequency spectrum lies on the plane where $\omega_t = 0$. The frequency response of the HVS can be modeled as a 3D filter bank, where each subband is selective to a motion with a given velocity [7].

It is difficult to realize the above mentioned motion-selective decompositions with conventional 3D filter banks due to the limitation on the ROS of the 3D passbands. In this contribution, instead of the conventional 3D spatiotemporal domain analysis/synthesis filtering, we employ the mixed-domain (mix-D) approach [8] to develop a novel 3D mix-D filter bank structure, where we assume that the 3D mixed-domain is composed of the 2D spatial transform domain k_1-k_2 (corresponding to the spatial domain n_1-n_2) and the 1D temporal domain n_3. The proposed filter bank has the advantage that it can realize *arbitrary* spectral decompositions by a set of *properly designed* 1D linear difference equation (LDE) filter banks. These 3D mix-D filter banks satisfy the perfect reconstruction (PR) conditions under arbitrary spectral decompositions. Thus the design and implementation of mix-D filter banks turns out to be a 1D approach, which is easier than the conventional MD approach. The mix-D filter bank is especially well-suited for applications of video coding, because video signals are naturally bounded in the spatial domain and well-established transform coding methods, such as discrete cosine transform (DCT) coding, can be conveniently used. At the same time, these signals are unbounded along the temporal axis, therefore LDE filters are most suitable for processing along this direction. A coding scheme using the proposed filter bank is also considered, where the signal is first decomposed into mix-D subband signals. An adaptive energy-based method is then used to adaptively select the energy-dominated subbands containing the highest signal energies. This method combines transform coding and subband coding methods, leading to efficient data compression.

II. 3D MIX-D FILTERING

Consider the MD mix-D filtering technique [8]. We summarize here the method for the 3D case (with slightly different notation than [8]). The discrete Fourier transform (DFT) is employed here because of the direct correspondence between the transform domain and the frequency-domain energy content of the signal. However, other transforms can be used [9]. The 3D mix-D filtering operation is as follows:

1. The 2D DFT is applied to the 3D input signal $\mathbf{x}(n_3) \equiv [x_{n_1,n_2}(n_3)]$ over the two spatial varibles, n_1 and n_2, resulting in a complex 3D mix-D sequence $\mathbf{X}(n_3) \equiv [X_{k_1,k_2}(n_3)]$.

2. Each complex sequence $X_{k_1,k_2}(n_3)$ is filtered by a 1D LDE filter over the temporal variable n_3 leading to $\mathbf{Y}(n_3) \equiv [Y_{k_1,k_2}(n_3)]$. The frequency response of each 1D LDE filter is determined by the required 3D frequency response at each 2-tuple (k_1, k_2).

3. The complex output sequences of the 1D LDE filters are inversely transformed to give the required 3D filtered output $\mathbf{y}(n_3) \equiv [y_{n_1,n_2}(n_3)]$.

III. A 3D MIX-D FILTER BANK

Using the above mix-D approach, we propose the mix-D filter bank structure, as shown in Fig.1. The process is summarized as follows:

1. The input signal matrix $\mathbf{x}(n_3) \equiv [x_{n_1,n_2}(n_3)]$ is first DFT-transformed over the two spatial variables, n_1 and n_2, resulting in a complex 3D mix-D sequence $\mathbf{X}(n_3) \equiv [X_{k_1,k_2}(n_3)]$.

2. 1D LDE *analysis filtering* is applied on each sequence $X_{k_1,k_2}(n_3)$ over the temporal variable n_3. This Analysis Filter Bank splits its input into the temporal frequency ($\omega_3 - axis$) subbands $\mathbf{X1}_{k_1,k_2}(n_3)$. The elements of the vector $\mathbf{X1}_{k_1,k_2}(n_3)$ are the ω_3-axis subband signals at each 2-tuple (k_1, k_2).

3. Equivalent frequency domain analysis filtering, which is simply a mapping (or data reordering) operation $\mathbf{R}$, is used to partition $[\mathbf{X1}_{k_1,k_2}(n_3)]$ into a set of 3D mix-D subband vectors $\mathbf{X2}_i(n_3)$, $i = 0, 1, \ldots, N_s - 1$ and N_s is the number of 3D subbands. The overall 3D decompositions are determined by those 1D decompositions as well as the reordering operation, which leads to simple implementations. By adjusting the decompositions of the 1D filter banks, we can realize arbitrary 3D spectral decompositions which conventional filter banks cannot always realize.

4. At the Synthesis Section, the reversed procedure is carried out. Equivalent frequency domain synthesis filtering, which is the inverse mapping $\mathbf{R}^{-1}$, is employed to the 3D subband vectors $\hat{\mathbf{X}}\mathbf{2}_i(n_3)$, $i = 0, 1, \ldots, N_s - 1$, to recover the ω_3-axis subband signals $\hat{\mathbf{X}}\mathbf{1}_{k_1,k_2}(n_3)$.

5. The ω_3-axis subband signals at each of 2-tuples (k_1, k_2) are combined in the synthesis filters to form the reconstructed 3D mix-D signal $\hat{\mathbf{X}}(n_3) \equiv [\hat{X}_{k_1,k_2}(n_3)]$. The 1D LDE analysis filters used in step "2" and the corresponding synthesis filters form 1D analysis/synthesis filter banks. In order to realize arbitrary decompositions, the 1D *uniform-band* filter banks cannot always meet the requirements and 1D *nonuniform-band* filter banks [10, 11] are generally required.

6. $\hat{\mathbf{X}}(n_3) \equiv [\hat{X}_{k_1,k_2}(n_3)]$ is finally inverse-transformed to give the reconstructed signal $\hat{\mathbf{x}}(n_3) \equiv [\hat{x}_{n_1,n_2}(n_3)]$.

IV. PR CONDITIONS

Let $\mathbf{x}_z(z_3) \equiv \mathcal{Z}\{\mathbf{x}(n_3)\}$, $\mathbf{X}_z(z_3) \equiv [X_{z(k_1,k_2)}(z_3)] \equiv \mathcal{Z}\{[X_{k_1,k_2}(n_3)]\}$, $\mathbf{X1}_{z(k_1,k_2)}(z_3) \equiv \mathcal{Z}\{\mathbf{X1}_{k_1,k_2}(n_3)\}$ and $\mathbf{X2}_{zi}(z_3) \equiv \mathcal{Z}\{\mathbf{X2}_i(n_3)\}$, where $\mathcal{Z}\{\cdot\}$ represents the z-transform. The corresponding signals in the synthesis section use similar notations. From the diagram shown in Fig.1, we obtain

$$\mathbf{X}_z(z_3) = \mathbf{W}_{N_1}\mathbf{x}_z(z_3)\mathbf{W}_{N_2}, \tag{1}$$

where $\mathbf{W}_{N_1}$ and $\mathbf{W}_{N_2}$ are the DFT matrices, i.e.

$$\mathbf{W}_{N_i} = \left[W_{N_i}^{nm}\right], \; W_{N_i}^{nm} = e^{-j\frac{2\pi nm}{N_i}}. \tag{2}$$

Since $\mathbf{R}$ and $\mathbf{R}^{-1}$, shown as part of Fig.1, are a pair of mutually invertible linear mapping operations, if it is assumed that

$$\hat{\mathbf{X}}\mathbf{2}_{zi}(z_3) = \mathbf{X2}_{zi}(z_3), \quad i = 0, 1, \ldots, N_s - 1, \tag{3}$$

then

$$\hat{\mathbf{X}}\mathbf{1}_{z(k_1,k_2)}(z_3) = \mathbf{X1}_{z(k_1,k_2)}(z_3), \tag{4}$$

and the reconstructed 3D mix-D signal $\hat{\mathbf{X}}_z(z_3)$ can be written as

$$\hat{\mathbf{X}}_z(z_3) = \left[\hat{X}_{z(k_1,k_2)}(z_3)\right], \; 0 < k_1 < N_1, \; 0 < k_2 < N_2, \tag{5}$$

where

$$\hat{X}_{z(k_1,k_2)}(z_3) = \sum_{i=0}^{L-1}\Big(\frac{1}{p_i q_i}\sum_{m=0}^{p_i-1}\sum_{n=0}^{q_i-1}\big(F_i(W_{p_i}^{m} z_3^{\frac{1}{p_i}})\cdot$$

$$\cdot\, H_i(W_{p_i}^{m}W_{q_i}^{n} z_3^{\frac{1}{p_i}})X_{z(k_1,k_2)}(W_{q_i}^{np_i} z_3)\big)\Big), \tag{6}$$

where the subscript, (k_1, k_2) of $L_{(k_1,k_2)}$, $p_{(k_1,k_2)_i}$, $q_{(k_1,k_2)_i}$, $H_{(k_1,k_2)_i}$ and $F_{(k_1,k_2)_i}$, is omitted. L_{k_1,k_2} is the number of subbands.

The reconstructed signal $\hat{\mathbf{x}}_z(z_3)$ is then obtained as

$$\hat{\mathbf{x}}_z(z_3) = \mathbf{W}_{N_1}^{-1}\hat{\mathbf{X}}_z(z_3)\mathbf{W}_{N_2}^{-1}. \tag{7}$$

If two conditions are met: first, the set of 1D LDE analysis/synthesis filter banks meet their PR conditions respectively and secondly, the overall delays of these 1D LDE filter banks are equal, i.e.

$$T_{z(k_1,k_2)}(z_3) = \frac{\hat{X}_{z(k_1,k_2)}(z_3)}{X_{z(k_1,k_2)}(z_3)} = z_3^{-n_o}, \tag{8}$$

otherwise delay operations are required, then

$$\hat{\mathbf{X}}_z(z_3) = z_3^{-n_o}\mathbf{X}_z(z_3), \tag{9}$$

and therefore,

$$\hat{\mathbf{x}}_z(z_3) = \mathbf{W}_{N_1}^{-1}\hat{\mathbf{X}}_z(z_3)\mathbf{W}_{N_2}^{-1} = z_3^{-n_o}\mathbf{x}_z(z_3). \tag{10}$$

From (10), it is clear that the output of the 3D filter bank is the delayed version of the input and thus perfect reconstruction is achieved. The above two conditions guarantee the perfect reconstruction of the proposed 3D mix-D filter bank. Therefore, in order to design a 3D mix-D filter bank, we only need to design a set of 1D PR LDE filter banks (having *nonuniform-bands* in general). The details of the design of *nonuniform-band* filter banks have been reported in earlier work [10, 11].

V. PROPOSED CODING SCHEME

Consider a video coding scheme using the proposed filter bank. Its flow graph is shown in Fig.2, where spatial 8×8 block processing is employed in order to achieve high throughput and so that each block typically consists of one or two different types of motions, making the tracking of the motions easier. After subband filtering (**H**), a time-varying adaptive energy-based selection of the subbands is used. Since in real-time coding applications, the velocities of motions are usually unknown, and also vary with the time, it is necessary to estimate the location of the planar spectrum corresponding to the motion. A simple energy-based estimation is used, which is based on the observation that if the 3D planar spectrum, corresponding to a motion, crosses a subband, then a certain amount of energy must be contained in this subband, otherwise the energy in the subband is assumed zero. In the practical implementation, the number of distinct motions and the corresponding relative thresholds in each subband are first estimated according to the characteristics of the video signal. An energy estimation procedure is then applied to each subband and the highest energy subbands are selected. The signal is then quantized and entropy-encoded. This coding scheme combines transform coding and subband coding methods.

A. Mix-D Filter Bank Using the DCT

The DCT is often preferred for coding applications, because it is a real-valued transform and also because of its superior energy-compaction property. It can be shown that the mix-D filter bank using the DCT has the same structure as the one using the DFT except that the spectral characteristics of the 1D LDE analysis and synthesis filters are different. Since the spectrum of the DCT does not correspond to the frequency-domain energy distribution, the design of the filter banks using the DCT is usually more complicated than in the DFT case. The relationship between the DFT and DCT is used to determine the required frequency response of the subband filters. However, for some special cases, for example if the impulse response of the filter is real and symmetrical, the relationship between its DCT and DFT is very simple and the N-point DCT is equal to its 2N-point DFT with a constant gain and a phase shift in its defined range [12]. For LT signals, the ROS of the spectrum is the same for both the DCT and DFT cases; that is, they lie in planes. Static backgrounds are a special case of LT signals. Therefore a similar strategy to that used in the DFT case) can be used for the spectral decompositions in the DCT case.

B. Coding Example

In this example, a full motion video (News Reporter) has been coded using the proposed coding scheme. An adaptive selection of subbands is employed based on the number of motions and relative energy level in each subband. The original video is represented at 8 bits/pix and the encoded video has the compressed representation of 0.25 bits/pix while retaining subjectively good quality.

VI. CONCLUSION

In this contribution, a mix-D 3D filter bank is presented. The proposed filter bank can realize arbitrary spectral decompositions by a set of properly designed 1D PR LDE filter banks, leading to straightforward design and implementation. A coding scheme using the mix-D filter bank is proposed and an example shows the potential for using an adaptive mix-D filter bank coding scheme.

REFERENCES

[1] J. W. Woods and S. D. O'Neil, "Subband Coding of Images," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, Oct. 1986.

[2] G. Karlsson and M. Vetterli, "Three Dimensional Subband Coding of Video," *Proc. IEEE Int. Conf. on ASSP*, 1988.

[3] M. Vetterli, "Multi-dimensional Sub-band Coding: Some Theory and Algorithms," *Signal Processing* 6, Feb. 1984.

[4] E. Viscito and J. P. Allebach, "The Analysis and Design of Multidimensional FIR Perfect Reconstruction Filter Banks for Arbitrary Sampling Lattices," *IEEE Trans. on Circuits and Systems*, vol. 38, Jan. 1991.

[5] T. Chen and P. P. Vaigyanathan, "Multidimensional Multirate Filters Derived from One Dimensional Filters," *Electronics Letters*, 31st Jan. 1991.

[6] L. T. Bruton and N. Bartley, "Three-dimensional Image Processing Using the Concept of Network Resonance," *IEEE Trans. on Circuits and Systems*, vol. 32, July 1985.

[7] R. H. Jonsson and R. M. Mersereau, "Efficient Motion-oriented Filter Banks for Video Coding," *Proc. IEEE Int. Symp. on Circuits and Systems*, 1993.

[8] A. A. Choudhury and L. T. Bruton, "Multidimensional Filtering Using Combined Discrete Fourier Transform and Linear Difference Equation Methods," *IEEE Trans. on Circuits and Systems*, vol. 37, Feb. 1990.

[9] K. S. Knudsen and L. T. Bruton, "Mixed Domain Filtering of Multidimensional Signals," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 1, Sept. 1991.

[10] B. Liu and L. T. Bruton, "The Design of Nonuniform-band Maximally Decimated Filter Banks," *Proc. IEEE Int. Symp. on Circuits and Systems*, 1993.

[11] B. Liu and L. T. Bruton, "The Design of N-band Nonuniform-band Maximally Decimated Filter Banks," *Proc. 27th Asilomar Conf. on Signals, Systems and Computers*, 1993.

[12] B. Chitprasert and K. R. Rao, "Discrete Cosine Transform Filtering," *Proc. IEEE Int. Conf. on ASSP*, 1990.

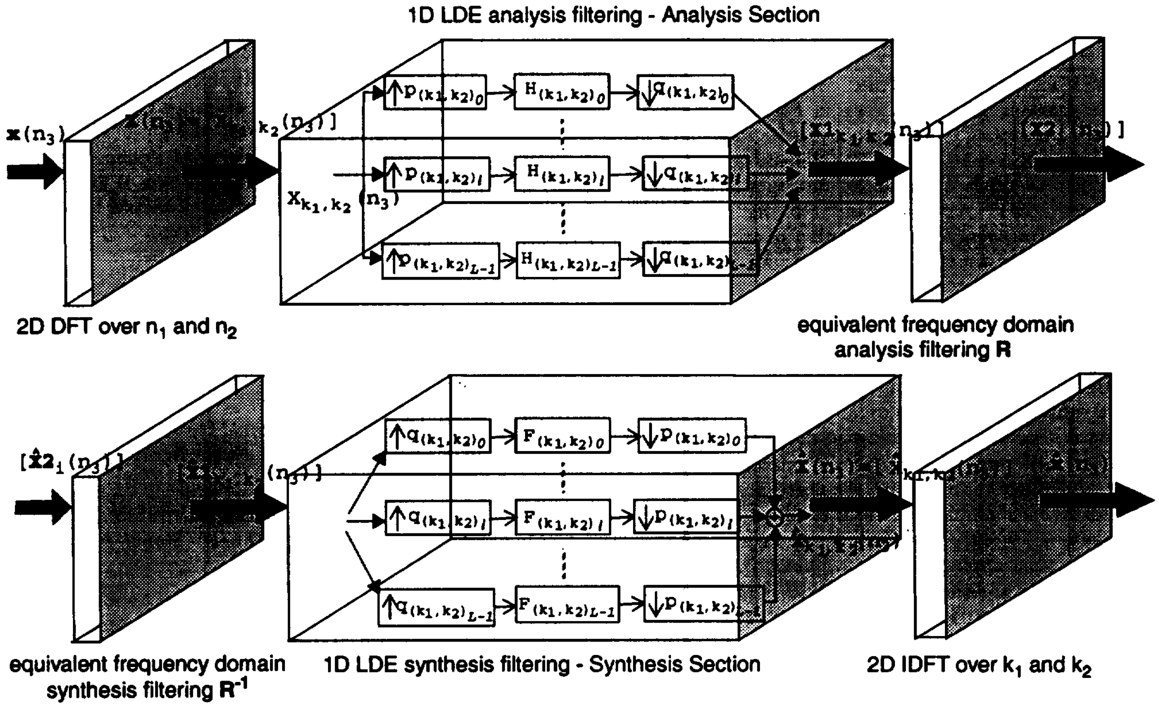

Fig. 1 3D mix-D filter bank

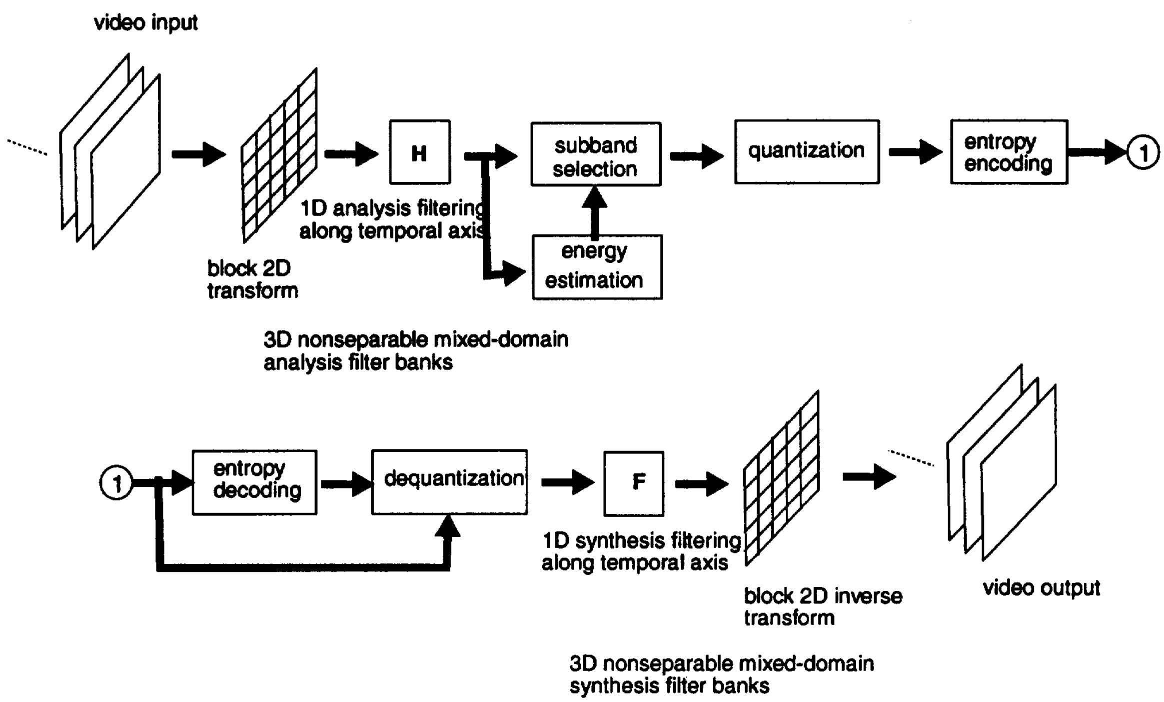

Fig.2 The coding scheme using the proposed mix-D filter bank