

ADAPTIVE WAVELET TRANSFORM FOR IMAGE COMPRESSION VIA DIRECTIONAL QUINCUNX LIFTING

Chuo-Ling Chang, Arian Maleki, and Bernd Girod

Information Systems Laboratory, Department of Electrical Engineering, Stanford University
{chuoling,arianm,bgirod}@Stanford.EDU

ABSTRACT

We propose a novel adaptive wavelet transform that exploits local image properties for image compression. It combines wavelet filters adaptive to edge orientations with quincunx subsampling to form a 2-D nonseparable transform through lifting. Filter selections are efficiently represented. Significant improvement on both subjective and objective quality over the conventional separable transform is observed. In addition, unlike previous adaptive transforms, the symmetry in quincunx subsampling enables even quality for image features along different directions and the compression performance is insensitive to image orientation.

1. INTRODUCTION

The 2-D discrete wavelet transform (DWT) is the most important new image compression technique of the last decade [1] [2] [3]. Conventionally, the 2-D DWT is carried out as a separable transform by cascading two 1-D transforms in the vertical and horizontal direction. Such a separable transform cannot efficiently represent edges in the image not aligned in these two directions since it distributes the energy of these edges into several subbands.

Lifting is a procedure to design wavelet transforms that are ensured to achieve perfect reconstruction [4]. Using lifting, Claypoole et al. developed transforms that locally adapt the length of the wavelet filters in the prediction step of lifting such that filtering is not performed across edges [5]. However, no significant improvement on objective quality measurement over the conventional 2-D DWT was reported, although subjective improvement was observed. Similarly, Taubman proposed to locally adapt the filtering direction in the prediction step to edge orientations, and some objective quality improvement was reported [6]. Note that both approaches do not explicitly signal the filter selections to the decoder even in the presence of quantization noise. In [5], a mechanism is introduced to exactly recover the selections at the decoder, given that the encoder has certain prior knowledge about the quantization noise. In [6], the filter selection process is designed to be robust against quantization noise so that it can be reliably repeated at the decoder.

More recently, approaches that adaptively select filtering directions via lifting, similar to [6], have again been proposed [7][8], but they choose to explicitly signal the selections to the decoder. These approaches have demonstrated significant subjective and objective quality improvement on images rich of textures, mostly due to the efficient representation and coding of the filter selections. However, although they are designed based on the frame-

work of separable transforms, i.e., two cascaded processes aiming at different sets of directions, adaptivity has rendered them into nonseparable transforms. Hence, the order of the two processes does affect the decorrelation ability. Therefore, having a fixed order, they typically favor certain directions, resulting in uneven quality for different image features and the compression performance is sensitive to image orientation.

We propose to combine directional lifting with quincunx subsampling to achieve an adaptive wavelet transform for image compression. Quincunx subsampling has been adopted with wavelet transform for several applications [9] [10]. Gouze et al. proposed a method to design quincunx wavelet transforms using lifting that adapts the wavelet filters globally to image statistics [10]. However, it usually involves complicated 2-D filters and the compression performance on typical images sampled on orthogonal grids is inferior to the separable 2-D DWT. The proposed adaptive transform using directional quincunx lifting involves only simple filters and achieves compression performance comparable to the previous adaptive transforms in [7] and [8], with a significant gain over the separable 2-D DWT for images with rich textures. Furthermore, owing to the symmetry in the quincunx grids, the transform is insensitive to edge directions and image orientation.

In the remainder of this paper, we describe various components of the proposed directional quincunx lifting scheme in Section 2. Experimental results are reported in Section 3, demonstrating the objective and subjective improvement over other adaptive transforms and the conventional transform.

2. DIRECTIONAL QUINCUNX LIFTING

2.1. Discrete Wavelet Transform with Quincunx Lifting

Let $\mathbf{X} = \{X[l], l \in \Pi\}$ denote an image on a 2-D orthogonal sampling grid $\Pi = \{(m, n)^T \in \mathbb{Z}^2 | 0 \leq m \leq M - 1, 0 \leq n \leq N - 1\}$. To apply the DWT on \mathbf{X} , we first split the image into two fields, $\mathbf{X}_e = \{X[l_e], l_e \in \Pi_e\}$ and $\mathbf{X}_o = \{X[l_o], l_o \in \Pi_o\}$, with quincunx subsampling defined as $\Pi_e = \{(m, n)^T \in \Pi | m + n \text{ even}\}$ and $\Pi_o = \{(m, n)^T \in \Pi | m + n \text{ odd}\}$, as illustrated in Fig. 1(a).

Denote the low-pass subband sampled on Π_e by $\mathbf{L} = \{L[l_e], l_e \in \Pi_e\}$ and the high-pass subband on Π_o by $\mathbf{H} = \{H[l_o], l_o \in \Pi_o\}$, wavelet analysis on \mathbf{X} using lifting with quincunx subsampling can generally be expressed as

$$H[l_o] = X[l_o] - P_{l_o}(\mathbf{X}_e), \quad \forall l_o \in \Pi_o \quad (1a)$$

$$L[l_e] = X[l_e] + U_{l_e}(\mathbf{H}), \quad \forall l_e \in \Pi_e \quad (1b)$$

where $P_{l_o}(\cdot)$ and $U_{l_e}(\cdot)$ are functions of the sample values in the input field with a scalar output.

¹This work was supported, in part, by STMicroelectronics, and, in part, by the Max Planck Center for Visual Computing and Communication.

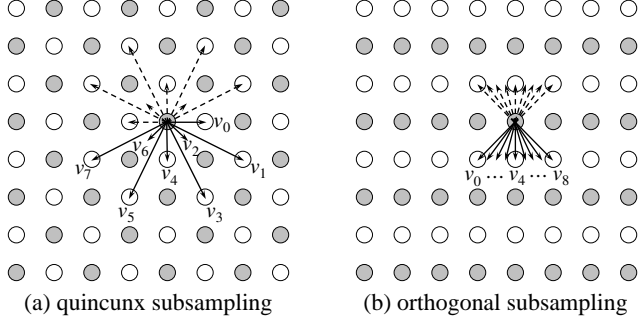


Fig. 1. Prediction directions for lifting on the image: white dots denote Π_e and gray dots denote Π_o .

2.2. Adaptive Direction Selection

In general, for each location in the the odd field, l_o , P_{l_o} in the prediction step is chosen to predict $X[l_o]$ from samples in \mathbf{X}_e such that the energy of the residual, $H[l_o]$, is minimized. We define 8 predictor candidates from which P_{l_o} can be adaptively selected:

$$P_i(\mathbf{X}_e, l_o) = \frac{1}{2}(X[l_o - v_i] + X[l_o + v_i]), \forall i = 0, \dots, 7 \quad (2)$$

where $v_0 = [0, 1]^T$, $v_1 = [1, 2]^T$, $v_2 = [0.5, 0.5]^T$, $v_3 = [2, 1]^T$, $v_4 = [1, 0]^T$, $v_5 = [2, -1]^T$, $v_6 = [0.5, -0.5]^T$, and $v_7 = [1, -2]^T$. Each P_i corresponds to applying wavelet analysis along a particular direction, as illustrated in Fig. 1(a). Note that we apply symmetric extension at image boundaries to account for the sample values at the out-of-bound locations not defined in \mathbf{X} . In addition, sample values at non-integer locations, such as $X[l_o + v_2]$, are not defined in \mathbf{X} . We obtain such values by interpolating from the existing samples colinear with the non-integer location in the direction perpendicular to the direction of wavelet analysis. For example, using a simple 4-tap linear interpolation filter, we define $X[l_o + v_2] = \frac{1}{4}(-X[l_o + v_5] + 3X[l_o + v_4] + 3X[l_o + v_0] - 1X[l_o - v_7])$ (refer to Fig. 1(a)) and similarly for other values at non-integer locations. For comparison, the 9 prediction directions for directional lifting with orthogonal subsampling as proposed in [8] are shown in Fig. 1(b). Note that the configuration in Fig. 1(b) requires interpolation not only at half-resolution locations but also at quarter-resolution locations.

We choose to minimize the absolute difference rather than the energy of $H[l_o]$ to reduce the complexity, and the overhead for selecting each candidate is also taken into account in direction selection through a pre-defined Lagrangian multiplier λ . To reduce the overhead needed to signal the predictor selection for carrying out the inverse transform, we adapt the predictor in a block-wise fashion, rather than pixel-wise. The original grid Π is evenly partitioned into K non-overlapping blocks, each denoted by B_k , $k = 0, \dots, K - 1$. For each B_k , we select the best direction d_k by

$$d_k = \arg \min_i \sum_{l_o \in \Pi_o \cap B_k} |X[l_o] - P_i(\mathbf{X}_e, l_o)| + \lambda R_{k,i} \quad (3)$$

where $R_{k,i}$ denote the number of bits spent on the overhead for selecting $d_k = i$ for B_k . For the update step, we similarly define the 8 candidates for U_{l_e} as

$$U_i(\mathbf{H}, l_e) = \frac{1}{4}(H[l_e - v_i] + H[l_e + v_i]), \quad i = 0, \dots, 7 \quad (4)$$

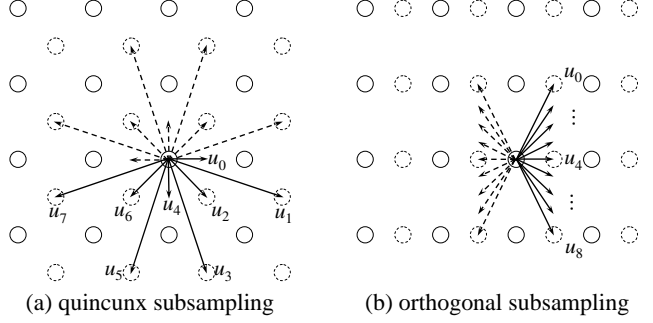


Fig. 2. Prediction directions for lifting on the low-pass subband: dashed dots denote Π_{L_e} and solid dots denote Π_{L_o} .

For samples in B_k , we apply the update step along the same direction d_k selected in the prediction step. As a result, for each block B_k (1) can be rewritten as

$$H[l_o] = X[l_o] - P_{d_k}(\mathbf{X}_e, l_o), \quad \forall l_o \in \Pi_o \cap B_k \quad (5a)$$

$$L[l_e] = X[l_e] + U_{d_k}(\mathbf{H}, l_e), \quad \forall l_e \in \Pi_e \cap B_k \quad (5b)$$

2.3. Wavelet Pyramid Generation

Each of \mathbf{L} and \mathbf{H} consists of half samples as in the original image \mathbf{X} . In order to obtain a structure similar to that from one level of the conventional separable 2-D DWT, i.e., four subbands each having a quarter of samples as in \mathbf{X} , we further apply another wavelet transform on \mathbf{L} to generate subbands \mathbf{LL} and \mathbf{LH} , and on \mathbf{H} to generate \mathbf{HL} and \mathbf{HH} .

The low-pass band \mathbf{L} is split into two fields, $\mathbf{L}_e = \{L[l_{L_e}], l_{L_e} \in \Pi_{L_e}\}$ and $\mathbf{L}_o = \{L[l_{L_o}], l_{L_o} \in \Pi_{L_o}\}$, where $\Pi_{L_e} = \{(m, n)^T \in \Pi_e | m \text{ even}\}$ and $\Pi_{L_o} = \{(m, n)^T \in \Pi_e | m \text{ odd}\}$ as illustrated in Fig. 2(a); \mathbf{H} is similarly split into $\mathbf{H}_e = \{H[l_{H_e}], l_{H_e} \in \Pi_{H_e}\}$ and $\mathbf{H}_o = \{H[l_{H_o}], l_{H_o} \in \Pi_{H_o}\}$, where $\Pi_{H_e} = \{(m, n)^T \in \Pi_o | m \text{ even}\}$ and $\Pi_{H_o} = \{(m, n)^T \in \Pi_o | m \text{ odd}\}$. With lifting, the predictor candidates for \mathbf{L} and \mathbf{H} are defined as

$$P_i^L(\mathbf{L}_e, l_{L_o}) = \frac{1}{2}(L[l_{L_o} - u_i] + L[l_{L_o} + u_i]) \quad (6)$$

$$P_i^H(\mathbf{H}_e, l_{H_o}) = \frac{1}{2}(H[l_{H_o} - u_i] + H[l_{H_o} + u_i]), \quad i = 0, \dots, 7$$

where $u_0 = [0, 1]^T$, $u_1 = [1, 3]^T$, $u_2 = [1, 1]^T$, $u_3 = [3, 1]^T$, $u_4 = [1, 0]^T$, $u_5 = [3, -1]^T$, $u_6 = [1, -1]^T$, and $u_7 = [1, -3]^T$ as illustrated in Fig. 2(a). The 9 directions for orthogonal subsampling proposed in [8] are shown in 2(b) for comparison.

We select the best direction for each block in \mathbf{L} and \mathbf{H} in the same way as for \mathbf{X} , except that each of \mathbf{L} and \mathbf{H} is divided into only $\frac{K}{2}$ blocks. Note that Π_{L_e} , Π_{L_o} , Π_{H_e} , and Π_{H_o} are again 2-D orthogonal grids, therefore they can be easily rearranged into the original grid Π with each subband occupying a quadrant of Π . We normalize the sample values in \mathbf{LL} by a factor of 2 and those in \mathbf{HH} by $\frac{1}{2}$ in order to keep the transform close to orthonormal. The whole process discussed so far constitutes one level of the proposed nonseparable and adaptive wavelet transform. This process can be iteratively applied on the \mathbf{LL} subband to generate a wavelet pyramid, as would have resulted from multiple levels of the conventional separable 2-D DWT. For each additional level of the transform, the block size remains the same so that the number of blocks is reduced by a factor of 4, but the Lagrangian multiplier λ is doubled to match with the normalization in \mathbf{LL} .

2.4. Adaptive Block-Mode Selection

To further increase the efficiency of the prediction step, each block B_k can be evenly partitioned into 4^t sub-blocks, $t = 0, \dots, T-1$, denoted as $B_{k,j}^t$, $j = 0, \dots, 4^t - 1$. The best direction is selected for each sub-block $B_{k,j}^t$ as described in (3), and denoted as $d_{k,j}^t$. Each sub-block size corresponds to a block-mode, and the best block-mode for block B_k is denoted by b_k . For each $B_{k,j}^t$, we denote the sum of absolute difference (SAD) in the high-pass samples predicted with $d_{k,j}^t$ by $SAD_{k,j}^t$, the overhead for selecting $d_{k,j}^t$ by $R_{k,j}^t$ and that for selecting $b_k = t$ by \bar{R}_t , then b_k is determined by

$$b_k = \arg \min_t \sum_{j=0}^{4^t-1} SAD_{k,j}^t + \lambda \left(\sum_{j=0}^{4^t-1} R_{k,j}^t + \bar{R}_t \right) \quad (7)$$

This is similar to the partition mode in [8] and the variable block-size motion compensation in H.264 [11]. In our current implementation, we consider 3 block-modes starting from 16×16 blocks.

2.5. Skip Mode Selection

For each wavelet decomposition, in addition to the adaptive transform using directional quincunx lifting, a non-adaptive quincunx wavelet transform is also carried out for a special skip mode. Using the definition in (2) and (6), the fixed predictors of the non-adaptive transforms are defined as $P_{skip} = \frac{1}{2}(P_2 + P_6)$, $P_{skip}^L = \frac{1}{2}(P_0^L + P_4^L)$ and $P_{skip}^H = \frac{1}{2}(P_0^H + P_4^H)$. The corresponding filters for the update steps are similarly defined.

After each adaptive transform, the Lagrangian costs from (7) are summed over all blocks and compared with the SAD of the high-pass samples (prediction residual) resulting from the non-adaptive transform. If the latter is smaller, the skip mode is issued and the non-adaptive transform that does not require any block-mode and direction overhead is applied. The skip mode selection is signalled with only 1 bit for each wavelet decomposition.

2.6. Block-Mode and Direction Coding

To encode the block-mode, we empirically observed that an appropriate λ value typically results in dominant occurrence of mode 0. Therefore, we encode a run of mode 0 by symbol 0 followed by the binary representation of the run-length; each occurrence of mode 1 and 2 is simply encoded by symbol 10 and 11 respectively.

To encode the direction in each sub-block with the selected block-mode, it is predicted from the direction selection at the nearest sub-block in the causally neighboring blocks, similar to that described in [8]. Thanks to the cyclic property of the proposed directions, the prediction residuals can be suitably mapped into 8 symbols and encoded by variable length coding (VLC). The VLC table can be either pre-determined from training image sets or optimized for the image by iterating between forming the VLC table and re-applying the adaptive transform until the table converges.

In our current implementation, for each level of the transform, the maximum run-length for mode 0 is optimally chosen and signalled along with the mode selections. The VLC table is adapted to each image, usually taking about 3 iterations to converge, and signalled with 24 bits. In addition, we take the average rate to encode an occurrence of mode 0 in the previous iteration as the \bar{R}_0 in (7) to refine the rate calculation.

3. EXPERIMENTAL RESULTS

We compare the compression performance resulting from the proposed adaptive wavelet transform using directional quincunx lifting (referred as DQL-DWT) to that from the directional orthogonal lifting scheme proposed in [8] (DOL-DWT) with the prediction directions illustrated in Fig. 1(b) and Fig. 2(b). The conventional separable transform using the 5/3 wavelet (SEP-DWT) is also compared, as the support it requires for wavelet filtering is similar to that in the two adaptive schemes. In our current implementation, the wavelet coefficients are encoded by the SPIHT algorithm [2] to generate a scalable bitstream.

The DOL-DWT scheme we implemented shares the same block-modes, skip mode selection, and overhead coding as in DQL-DWT. For DOL-DWT we use a 4-tap interpolation filter derived from truncating the sinc kernel, resulting in the same half-resolution filter as in DQL-DWT (Section 2.2) but a more complicated quarter-resolution filter with non-rational coefficients. The compression performance for the 512×512 image *Barbara* and *Lenna* is shown in Fig. 3 and Fig. 4 respectively. Different qualities (PSNR) are obtained by truncating the scalable bitstreams at different bit-rates. The block-mode and direction selections are losslessly coded and this overhead is included in the plots.

For the original *Barbara*, DOL-DWT and DQL-DWT both outperform SEP-DWT by up to 2.6 dB as shown at the top of Fig. 3. To evaluate the sensitivity of the two adaptive transforms to image orientation, we rotate *Barbara* by 90° and the performance is shown at the bottom of Fig. 3. Compared to the performance for the original *Barbara*, for DOL-DWT the quality drops by up to 0.9

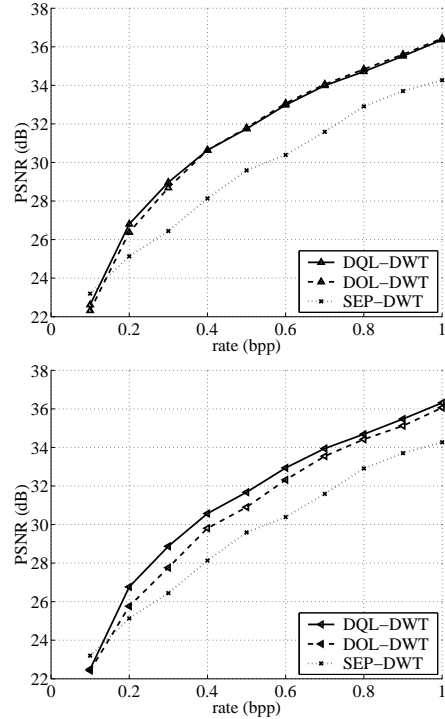


Fig. 3. Compression performance on (top) *Barbara* (bottom) *Barbara* rotated by 90° . Block-mode and direction overhead takes about 0.05 bpp for DQL-DWT and DOL-DWT.

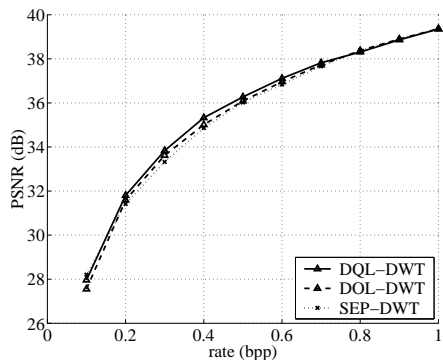


Fig. 4. Compression performance on *Lenna*. Block-mode and direction overhead takes about 0.023 bpp for DQL-DWT and DOL-DWT.

dB, whereas the difference for DQL-DWT is negligible. This can be explained as follows. In the implemented DOL-DWT, wavelet decomposition in the vertical-ish directions (Fig. 1(b)) is carried out before that in the horizontal-ish directions (Fig. 2(b)). The former decomposition is more effective since it is applied on higher resolutions. Hence the transform is more advantageous for images with more vertical-ish edges, for example, the stripes on the cloth in *Barbara*. On the other hand, thanks to the symmetry provided by quincunx subsampling the performance of DQL-DWT is insensitive to image orientation.

For *Lenna*, PSNR improvement from the adaptive transforms is less significant although subjective improvement can still be observed. The maximum PSNR gain is about 0.5 dB for DOL-DWT and 0.3 dB for DQL-DWT. Directional lifting is less effective in *Lenna* since it contains not as many salient features as in *Barbara*. Similar observations on these two images are reported in [7].

Finally, reconstructions of part of the JPEG2000 test image *Bike* encoded at 0.5 bpp using the 3 transforms are shown in Fig. 5 to demonstrate the superior visual quality provided by the proposed DQL-DWT. Note that for DOL-DWT the ringing artifacts are more evident around the horizontal-ish edges than the vertical-ish edges. DQL-DWT outperforms other schemes, both subjectively and objectively, and generally achieve even qualities for edges along various directions. This also suggests the potential of the proposed transform on applications other than natural images, such as graphical illustrations and compound documents.

4. CONCLUSIONS

A novel adaptive nonseparable 2-D wavelet transform for image compression is proposed. It combines directional lifting with quincunx subsampling to exploit local properties of edge orientations in images. Significant improvement on both subjective and objective quality over the conventional separable transform is reported. The uneven reconstruction quality from the previous adaptive transform for image features along different directions is also mitigated with quincunx subsampling. Ongoing research includes adopting more complex predictors and incorporating the proposed transform with JPEG2000-based wavelet coefficient coding.

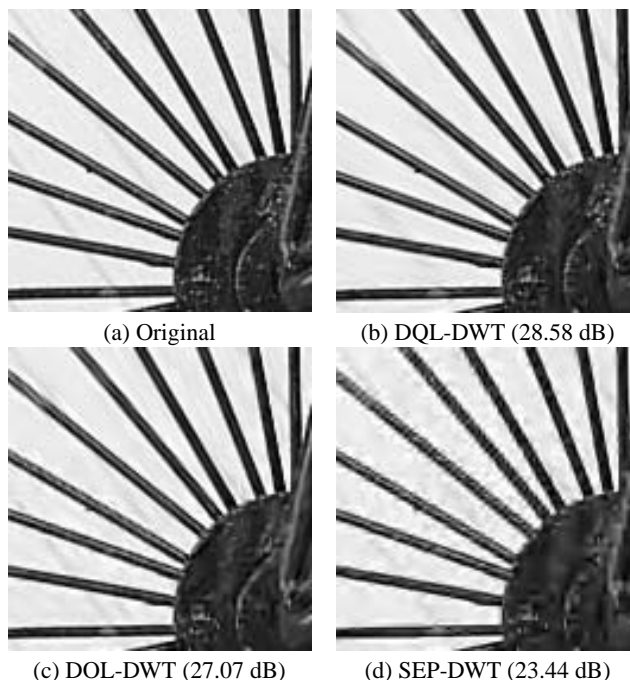


Fig. 5. Part of the JPEG2000 test image *Bike* encoded at 0.5 bpp

5. REFERENCES

- [1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 205–220, Apr. 1992.
- [2] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on Set Partitioning in Hierarchical Trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 243–250, June 1996.
- [3] D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*, Kluwer Academic Publishers, 2002.
- [4] W. Sweldens, "The lifting scheme: A construction of second generation wavelets," *SIAM Journal on Mathematical Analysis*, vol. 29, no. 2, pp. 511–546, 1998.
- [5] R. L. Claypoole, G. M. Davis, W. Sweldens, and R. G. Baraniuk, "Nonlinear wavelet transforms for image coding via lifting," *IEEE Trans. Image Processing*, vol. 12, pp. 1449–1459, Dec. 2003.
- [6] D. Taubman, "Adaptive, non-separable lifting transforms for image compression," in *Proc. IEEE Int. Conf. on Image Processing 1999*, Kobe, Japan, Oct. 1999, vol. 3, pp. 772–776.
- [7] D. Wang, L. Zhang, and A. Vincent, "Improvement of JPEG2000 using curved wavelet transform," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing 2005*, Philadelphia, PA, USA, Mar. 2005, vol. 2, pp. 365–368.
- [8] W. Ding, F. Wu, and S. Li, "Lifting-based wavelet transform with directionally spatial prediction," in *Proc. Picture Coding Symposium 2004*, San Francisco, CA, USA, Dec. 2004.
- [9] M. Barlaud, P. Sole, T. Gaidon, M. Antonini, and P. Mathieu, "Pyramidal lattice vector quantization for multiscale image coding," *IEEE Trans. Image Processing*, vol. 3, pp. 367–381, July 1994.
- [10] A. Gouze, M. Antonini, M. Barlaud, and B. Macq, "Design of signal-adapted multidimensional lifting scheme for lossy coding," *IEEE Trans. Image Processing*, vol. 13, pp. 1589–1603, Dec. 2004.
- [11] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 560–576, 2003.