Scalable Direction Representation for Image Compression with Direction-Adaptive Discrete Wavelet Transform

Tao Xu, Chuo-Ling Chang, and Bernd Girod

Information Systems Laboratory
Department of Electrical Engineering, Stanford University
Stanford, CA 94305 U.S.A.

ABSTRACT

The direction-adaptive discrete wavelet transform (DA-DWT) locally adapts the filtering direction to the geometric flow in the image. DA-DWT image coders have been shown to achieve a rate-distortion performance superior to non-adaptive wavelet coders. However, since the direction information must always be signalled regardless of total bit-rate, performance at very low bit-rates might be worse. In this paper, we propose two scalable direction representations: the layered scheme which is similar to the scalable motion vector representation in scalable video coding and the level-unit scheme which provides finer granularity upon the layered scheme. Experimental results indicate that we can achieve the desirable performance at both low and high bit rates with our proposed level-unit scheme. Significant improvement in image quality (about 3-5 dB) is observed at very low bit rate, relative to non-scalable coding of the direction information.

Keywords: Adaptive signal processing, Scalable image coding, Wavelet transform

1. INTRODUCTION

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

Lifting is a procedure to design DWTs that are guaranteed to achieve perfect reconstruction.[4] Using lifting, Claypoole et al. developed transforms which locally adapt the length of the wavelet filters in the prediction step of the lifting structure such that the neighborhoods used for prediction never cross an edge.[5] However, no significant improvement on objective quality measure over the conventional 2-D DWT was reported, although subjective improvement was observed. Similarly, Taubman proposed locally adapting the filtering direction to edge orientations in the prediction step, and some objective quality improvement was reported.[6] Note that both approaches do not explicitly signal the filter selections to the decoder. In [5], an update-first mechanism was introduced to ensure that the encoder and decoder were choosing predictors based on the same quantized data. In [6], the filter selection process was designed to be robust against quantization noise so that it could be reliably repeated at the decoder.

Recently, approaches that adaptively select the filtering direction via lifting, similar to [6], have again been proposed,[7–10] but they choose to explicitly signal the direction selections to the decoder. These approaches have shown significant subjective and objective quality improvements on images with rich textures. In this paper, we refer to the adaptive transforms in these approaches in general as the direction-adaptive discrete wavelet transform (DA-DWT). The direction information is always coded losslessly as side information. Therefore, at low bit rate, the direction information can consume an undue proportion of the total bit rate and such high fidelity direction representation is of little benefit for reconstructed images. It is desirable to have an appropriate scalable direction representation so that the amount of the direction overhead can be adjusted adaptively according to the

Further author information: E-mail: {taox, chuoling, bgirod}@stanford.edu, Telephone: +1 650 724 3647, Fax: +1 650 724 3648.

total bit rate. A similar issue of the scalable motion vector representation for scalable video coding is discussed in [11] [12].

In Section 2, we first discuss the non-scalable direction representation and propose a more reliable predictor for the direction predictive coding in DA-DWT. In Section 3, we present two scalable direction representations: the layered scheme which is similar to the scalable motion vector representation in scalable video coding and the level-unit scheme which provides finer granularity upon the layered scheme. We observed that the additional distortion caused by using a coarser direction representation at the decoder is almost independent of the rate of the wavelet coefficients in a wide range. Based on this observation, we further deduce an algorithm to allocate the rate for the scalable direction representation in Section 4. Experimental results are reported in Section 5, demonstrating the objective and subjective quality improvements of our proposed level-unit scheme at low bit rate.

2. DIRECTION-ADAPTIVE DISCRETE WAVELET TRANSFORM FOR IMAGE COMPRESSION

The direction-adaptive discrete wavelet transform (DA-DWT) was proposed to locally adapt the filtering direction to the geometric flow in images.[7–10] This approach is able to eliminate most of the large wavelet coefficients within the high-pass subbands. In other words, the DA-DWT can achieve better energy compaction than the conventional DWT. Therefore, although we need to spend a certain number of bits explicitly signaling the direction selections in the DA-DWT, the more efficient transform leads to significant subjective and objective quality improvements on images with rich textures.

2.1. Direction Selection in DA-DWT

To reduce the overhead needed to signal the direction selections, the directions may be determined blockwise in the DA-DWT.[7–10] And to further increase the efficiency of the prediction step in the lifting structure, each block may be further partitioned into sub-blocks, for example, 2×1, 1×2, 2×2, 4×1, 1×4, 4×2, 2×4, or 4×4 sub-blocks.[9] In the following discussion, we refer to these different block partitions as the "mode" of the blocks.

The optimal direction(s) and mode of each block are selected by minimizing a Lagrangian cost function as:

$$J = \mathrm{SAD} + \lambda R \tag{1}$$

where SAD is the sum of the magnitude of the wavelet coefficients in a block of the high-pass subband, R denotes the number of bits spent on the overhead for the direction and mode information. This overhead R is taken into account through a pre-defined Lagrangian multiplier λ, which depends on the reconstruction quality of interest, similar to the rate-constrained motion estimation in video coding.[13]

2.2. Non-scalable Direction Representation in DA-DWT

For the mode information, we encode the 1×1 mode by run-length coding since this mode occurs frequently and the others are coded with variable-length coding. The direction selection of each sub-block (note that a block with the 1×1 mode is also referred to as a sub-block) is predicted from the directions of the sub-blocks in the causal neighborhood and the residual is coded with variable-length coding. Predictive coding of the directions can be performed by the comparison predictor as:[7]

$$\begin{aligned} P &= \mathrm{comparison}(a, b, c) \\ &= \begin{cases} b & |a-b| > |a-c| \\ c & |a-b| \leq |a-c| \end{cases} \end{aligned} \tag{2}$$

where P is the prediction value for the current sub-block, a, b, and c are the directions of the top-left, left and top neighboring sub-blocks respectively as shown in Fig. 1. If $|a-b| > |a-c|$, i.e. the directions in this neighborhood are more horizontally correlated than vertically correlated, it is reasonable to choose the horizontal neighbor as the prediction value for the current sub-block; otherwise the vertical neighbor is used as the prediction value. So the comparison predictor is performed on a sub-block by sub-block basis.

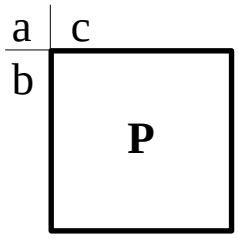


Figure 1. Illustration of the comparison predictor

Based on the comparison predictor, we propose the "vote" predictor. We introduce the concept of the "sub-block unit" which is the smallest possible sub-block. In other words, it is the sub-block corresponding to the finest block mode (e.g. the 4×4 mode in [9]). Therefore, every sub-block is composed of one or more sub-block units. While the comparison predictor is performed for each sub-block, as mentioned earlier, the vote predictor is performed for each sub-block unit. We first predict every sub-block unit within the current sub-block individually (note that these predictions may have different values), and then vote to decide the final prediction of the current sub-block, i.e. choose the most frequent prediction value.

Suppose the current sub-block in Fig. 1 is composed of MN sub-block units which are indexed as (i,j) and the corresponding predictions are denoted as $P_{(i,j)}$, $i = 1, \ldots, M$, $j = 1, \ldots, N$. The directions of the neighboring sub-block units are a, b_1, $\ldots$, b_M, c_1, $\ldots$, c_N* as shown in Fig. 2. We express the vote prediction process in (3) and (4)

$$P_{(i,j)} = \text{comparison}(\text{neighbor}(i,j)) \tag{3}$$

$$P = \text{vote}(P_{(i,j)} | i = 1, \ldots, M, j = 1, \ldots, N) \tag{4}$$

where the neighbor() operation is to obtain the directions of the causal top-left, left and top neighboring sub-block units respectively. For example, the result of neighbor(1,2) is $(c_1, P_{(1,1)}, c_2)$. Note here, we use the prediction $P_{(1,1)}$ because the actual direction of the (1,1) sub-block unit is unavailable until all $P_{(i,j)}$'s are obtained. This also ensures the decoder is able to replicate the same prediction process as the encoder.

To sum up, the essential difference between the comparison and vote predictor is that the former only uses a, b_1 and c_1 to predict, but the latter uses a, b_1, $\ldots$, b_M, c_1, $\ldots$, c_N to predict as shown in Fig. 2. So we actually introduce more neighbors to yield more reliable prediction for the current sub-block via the vote process. Our experimental results show that the vote predictor performs slightly better than the comparison predictor. Fig. 3 provides an example of these two different predictors. We can see that the comparison predictor predicts direction 0 as 5, whereas the vote predictor is able to correct this mistake.

3. SCALABLE DIRECTION REPRESENTATION FOR DA-DWT

In this section, we propose two schemes to provide a scalable direction representation. With the scalable direction representation, the finest direction representation is always employed at the encoder to carry out wavelet analysis of the DA-DWT, but the decoder may receive only a coarser direction representation for wavelet synthesis. We refer to this as direction mismatch.

*b_i $(i = 1, \ldots, M)$ may have different values, so may c_j $(j = 1, \ldots, N)$.

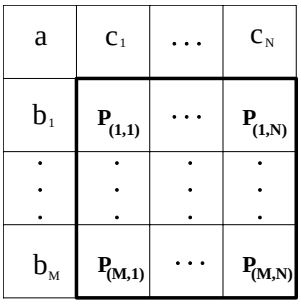
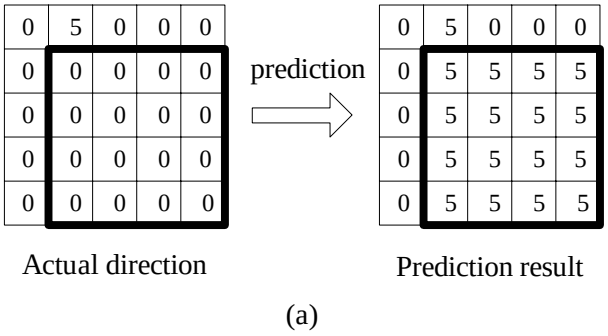


Figure 2. Illustration of the vote predictor



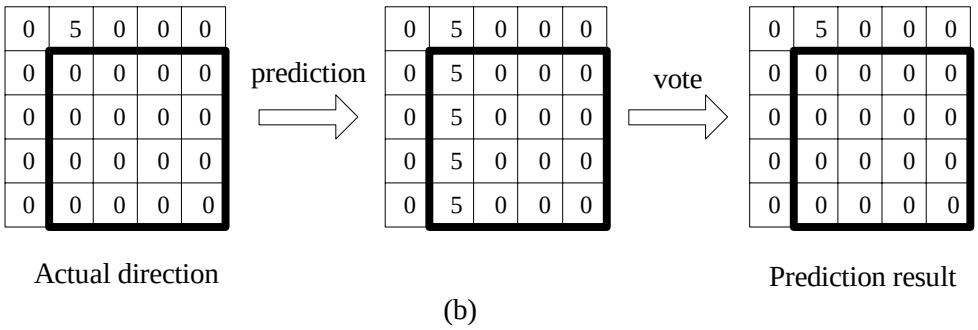


Figure 3. Examples of (a) the comparison predictor and (b) the vote predictor. For simplicity, we consider the prediction of the current block (i.e. the mode 1×1) which is represented by the large bold block, each small square block denotes the sub-block unit, the numbers represent the directions or predictions of the corresponding sub-block units.

3.1. Layered Scheme

Similar to the scalable motion vector representation for scalable 3D wavelet video coding,[12] we can represent the direction information with multiple quality layers (one base layer and several enhancement layers). A coarse direction representation can be reconstructed from the base layer and it can be successively refined by the subsequent enhancement layers. The base layer corresponds to the low bit-rate region and is generated using a relatively large λ in (1), while the enhancement layers correspond to the higher bit-rate region and are generated using a set of relatively small λ.

As shown in **Fig. 4 (a)**, assuming K quality layers are used, which correspond to the Lagrangian multipliers $\lambda_1 > \lambda_2 > \cdots > \lambda_K$, to represent the direction information. We obtain the coarsest direction representation D_1

by minimizing the Lagrangian cost function (1) with λ_1. The overhead rate R in (1) is computed regarding to the prediction residual $L_1 = D_1 - P_1$, where P_1 is the prediction of D_1 from the causal neighborhood. Note that we only need to transmit the prediction residual L_1 as the base layer to the decoder. At the ith ($i = 2, 3, \ldots, K$) enhancement layer, we may choose the previous direction representation D_{i-1} as the prediction of the current direction representation D_i, i.e. $P_i = D_{i-1}$. We refer to such a predictor as an "inter-layer" predictor. The inter-layer predictor is able to exploit the large amount of redundancy between successive direction representations and hence fully utilize the information we have already sent in the embedded bit-stream. The ith-coarsest direction representation D_i is obtained by minimizing the Lagrangian cost function (1) with λ_i and P_i. Similarly, we only transmit the prediction residual $L_i = D_i - P_i = D_i - D_{i-1}$ as the ith enhancement layer to the decoder ($i = 2, 3, \ldots, K$). In addition, we can use the inter-layer predictor for the mode information at the enhancement layers to exploit the redundancy between successive mode representations.

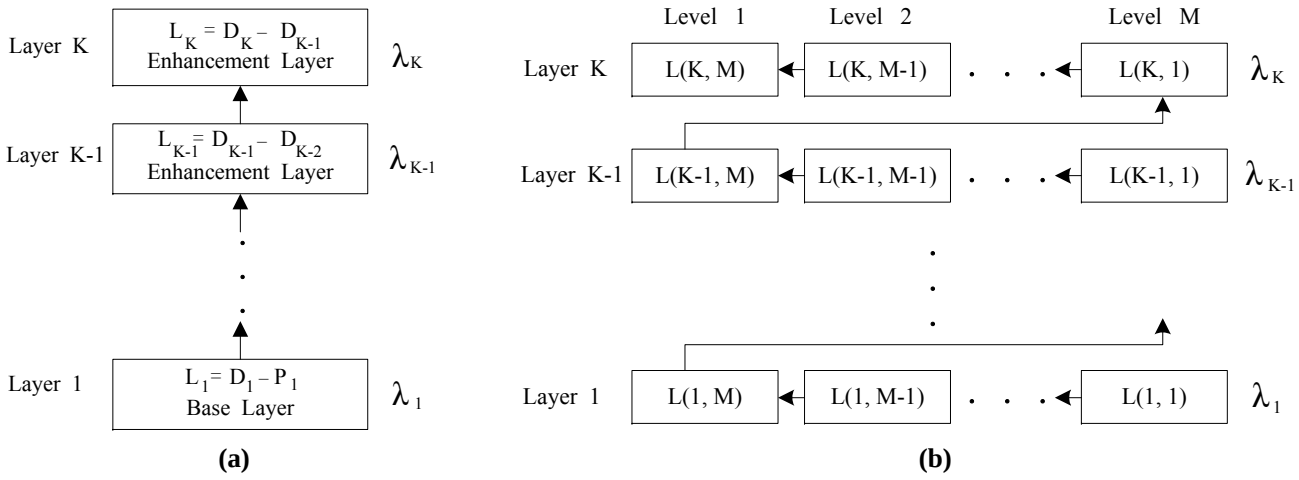


Figure 4. Two proposed scalable direction representations: (a) layered (b) level-unit

3.2. Level-Unit Scheme

In the layered scalable scheme as shown in **Fig. 4 (a)**, the number of natural truncation points of the embedded bit-stream of the direction information is only K, i.e. each layer corresponds to one natural truncation point. To enhance the scalability, we propose a "level-unit" scheme that provides many more natural truncation points by decomposing each quality layer into the level units as shown in **Fig. 4 (b)**. Note that we refer to the process of decomposing the image into the 4 subbands, i.e. **LL**, **LH**, **HL**, **HL** as one level of the DA-DWT (e.g. Level 1 in **Fig. 4 (b)**). Typically, multiple levels of the transform are performed by iteratively applying this process on the resulting **LL** subband. Each additional level of the transform requires an additional set of direction representation.

Suppose the M-level DA-DWT is performed on the input image. Denote the jth level unit within the ith layer as $L(i, M+1-j)^{\dagger}$, where $i = 1, 2, \ldots, K$, $j = 1, 2, \ldots, M$. The relationship between the quality layer and the corresponding level-units is $L_i = \bigcup_{j=1}^{M} L(i, j)$, $i = 1, 2, \ldots, K$. As the arrows shown in **Fig. 4 (b)**, our proposed level-unit scheme adaptively transmits the direction information in the level-unit-wise fashion rather than only layer-wise. The natural truncation point corresponding to the level unit $L(i, j)$ is denoted as $T(i, j)$ and the direction information from $L(1, 1)$ through $L(i, j)$ in the transmission order is referred as the truncated direction representation $C(i, j)$ with the truncation point $T(i, j)$, $i = 1, 2, \ldots, K$, $j = 1, 2, \ldots, M$. The number of natural truncation points for the level-unit scalable scheme is MK, much larger than that of the layered scalable scheme.

[†]Note that we intentionally reverse the order of the second coordinate with the level index as shown in **Fig. 4 (b)**.

Since the transmission unit is the level unit, the decoder may receive only a fraction of one layer. Therefore, we must develop a mechanism to construct a complete decoding direction representation at the decoder from the received embedded bit-stream. In general, there are two cases: one is when only the base layer or part of the base layer (i.e. $C(1,j)$, $j=1,2,\ldots,M$) is received at the decoder, and the other is when more than the base layer (i.e. $C(i,j)$, $i>1$) is received at the decoder. In the former case, we propose the "hierarchical upsampling" method to estimate the remaining part of the coarsest direction representation D_1. Because of the inherently hierarchical multi-resolution representation of DWT and the similarity between the directions at different levels, we can upsample the directions at the higher level (i.e. the lower resolution) to serve as the estimate of the directions at the lower level(s) which are not transmitted[‡]. However, in the latter case, we select the most accurate direction representation available at each level as the corresponding direction representation. For example, in the former case, if the decoder only receives $C(1,1)$, we upsample the coarsest direction representation D_1 at Level M to estimate the directions at the remaining levels to form a complete decoding direction representation. In the latter case, if the decoder receives $C(2,1)$, the decoding direction representation consists of the second-coarsest direction representation D_2 at level M and the coarsest direction representation D_1 at the other levels.

4. OPTIMAL TRUNCATION POINT ALGORITHM

With the scalable direction representation, we may choose different truncation points for decoding for different total bit-rates. The remaining problem is how to decide the optimal truncation point for a given total bit rate. Similar to using scalable motion vectors for scalable 3D wavelet video coding,[12] we observed that the distortion introduced by direction mismatch is almost independent of the rate of the wavelet coefficients in a wide range. Based on this observation, we deduce an optimal truncation point algorithm.

We refer to the bit rate required to explicitly transmit the truncated direction representation $C(i,j)$ as $R_{dir}(i,j)$ and the distortion introduced by direction mismatch while the wavelet coefficients are losslessly coded is denoted as $D_{dir}(R_{dir}(i,j))$, $i=1,2,\ldots,K$, $j=1,2,\ldots,M$. Note that $D_{dir}(R_{dir}(K,M))=0$ since the lifting structure guarantees perfect reconstruction. Similarly, the distortion introduced by the wavelet coefficient quantization with the finest direction representation is denoted as $D_{coeff}(R_{coeff})$, where R_{coeff} is the bit rate required for lossy encoding of the wavelet coefficients.

Through experiments, we observe that the distortion caused by direction mismatch is almost independent of the rate of the wavelet coefficients R_{coeff} in a wide range. Let R_{total} represent the total bit-rate. Based on this observation, we can estimate the total distortion $D_{i,j}(R_{total})$ which is decoded by $C(i,j)$ as:

$$\begin{aligned} D_{i,j}(R_{total}) &= D_{coeff}(R_{coeff}) + D_{dir}(R_{dir}(i,j)) \\ &= D_{coeff}(R_{total} - R_{dir}(i,j)) + D_{dir}(R_{dir}(i,j)) \end{aligned} \tag{5}$$

where $i=1,2,\ldots,K$, $j=1,2,\ldots,M$. Assuming we construct $K=3$ quality layers, which correspond to $\lambda_1=150$, $\lambda_2=75$, and $\lambda_3=13$, and $M=3$ levels for the direction information, **Fig. 5** illustrates the relationship between R_{coeff} and the total distortion D for the test image *Spoke* when three different truncated direction representations $C(3,3)$, $C(2,3)$, and $C(1,3)$ are received at the decoder respectively. The distortion values estimated according to Equation (5) are also shown in **Fig. 5**. We can see that the estimation fits nicely with the actual distortion. **Table 1** provides the average deviation of the distortion estimation for other test images.

According to this observation, we can easily find the optimal truncation point $T(i_{opt},j_{opt})$ for any target total bit rate R_{total} by Equation (6):

$$(i_{opt},j_{opt}) = \arg\min_{i,j}\{D_{coeff}(R_{total} - R_{dir}(i,j)) + D_{dir}(R_{dir}(i,j))\} \tag{6}$$

where $i=1,2,\ldots,K$, $j=1,2,\ldots,M$.

[‡]The same block size at different levels is implied here.

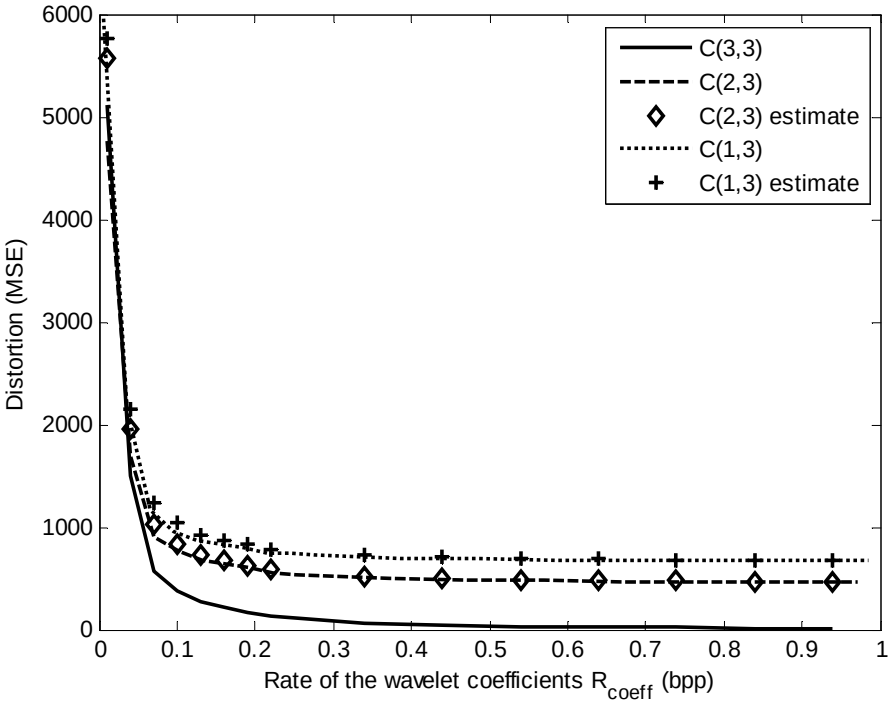


Figure 5. The rate of the wavelet coefficients and distortion curves for image *Spoke*

Table 1. Average deviation of distortion estimation (unit: %)

Truncated direction representation	$C(2,3)$	$C(2,2)$	$C(2,1)$	$C(1,3)$	$C(1,2)$	$C(1,1)$
Spoke	2.59	3.05	3.16	2.89	3.74	3.85
Lena	3.34	3.47	3.55	3.06	4.76	4.76
MRI	3.54	3.88	3.48	3.09	3.63	4.66
Barbara	6.21	6.95	6.71	5.79	6.85	5.54
Goldhill	5.79	5.92	6.03	5.85	7.04	7.27
Boats	3.95	4.06	4.16	3.58	3.97	5.03
Harbour	7.21	8.13	8.54	8.01	8.56	9.13
Man	5.14	6.07	6.74	5.02	6.65	7.05
Smandril	8.17	9.64	10.13	8.38	9.71	10.22
Cameraman	4.16	5.03	5.15	4.23	5.14	5.31

5. EXPERIMENTAL RESULTS

In the experiments, we compare three different cases: (1) conventional DWT, (2) DA-DWT with the non-scalable direction representation, (3) DA-DWT with the level-unit scalable direction representation. The DA-DWT is implemented based on the approach described in [9], but without the bandeletization procedure. For all cases, we adopt the $(6,6)$ interpolating wavelet.[9,14] The wavelet transform coefficients are encoded by the TCE embedded bit-plane coder,[15] which has been shown to achieve performance comparable to JPEG 2000.

For the non-scalable direction representation, we encode and decode at $\lambda = 13$. For the scalable direction

representation, we construct $K = 2$ quality layers with $\lambda_1 = 150$ for the base layer, $\lambda_2 = 13$ for the enhancement layer and $M = 3$ levels. Fig. 6 shows the compression performance for the 256×256 image *Spoke* and the 512×512 image *Barbara*. Fig. 7 shows the reconstructed images of *Spoke* and *Barbara* in these three different cases at different total bit rates. We can see that conventional DWT performs well at the low bit rates but poorly at the high bit rates. In contrast, the DA-DWT with the non-scalable direction representation performs well at the high bit rates but poorly at the low bit rates because a significant amount of bit-rates is devoted to transmit the direction information. With our proposed level-unit scalable direction representation, we can achieve the desirable performance at both low and high bit rates. Significant improvements on both subjective and objective quality (around 3-5 dB) are observed at low bit rates, relative to non-scalable coding of the direction information.

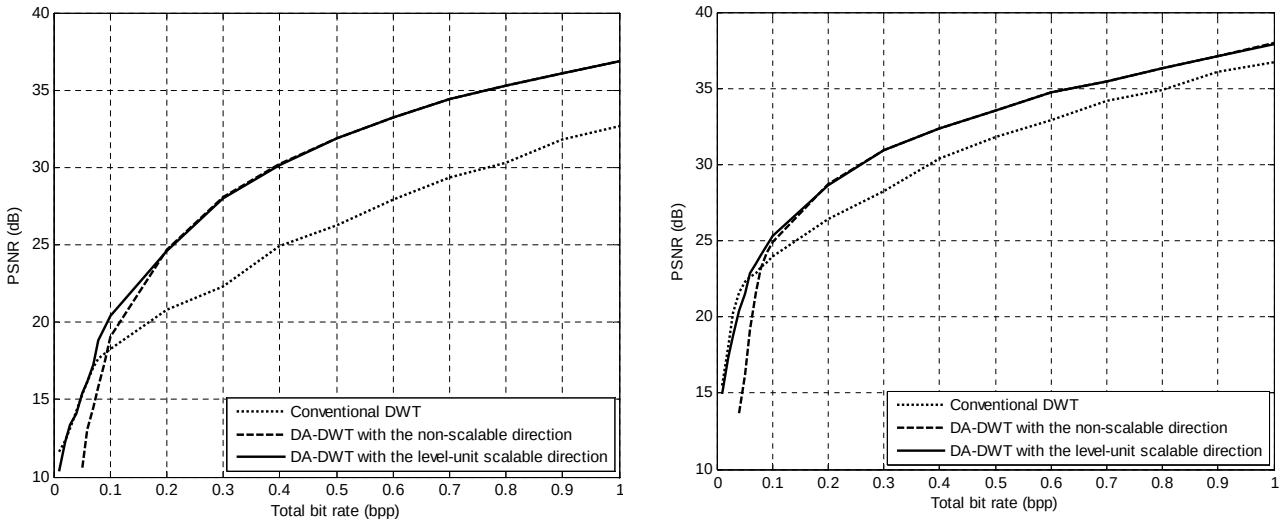


Figure 6. Compression performance for (left) *Spoke* (right) *Barbara*

6. CONCLUSION

In this work, two scalable direction representations are investigated for direction-adaptive discrete wavelet transform(DA-DWT): the layered scheme and the level-unit scheme. In addition, we observed that the distortion caused by using a coarser direction representation at the decoder is almost independent of the rate of the wavelet coefficients in a wide range. For any total bit-rate, we are able to choose the most appropriate direction representation by the proposed optimal truncation point algorithm. Experimental results confirm the effectiveness of the level-unit scheme, which can achieve the desirable performance at both low and high bit-rate regions.

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, Jun. 1996.
3. D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*, Kluwar Academic Publishers, Norwell, 2002.
4. W. Sweldens, "The lifting scheme: A construction of second generation wavelets," *SIAM Journal on Mathematical Analysis* **vol. 29**, 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, 2003.

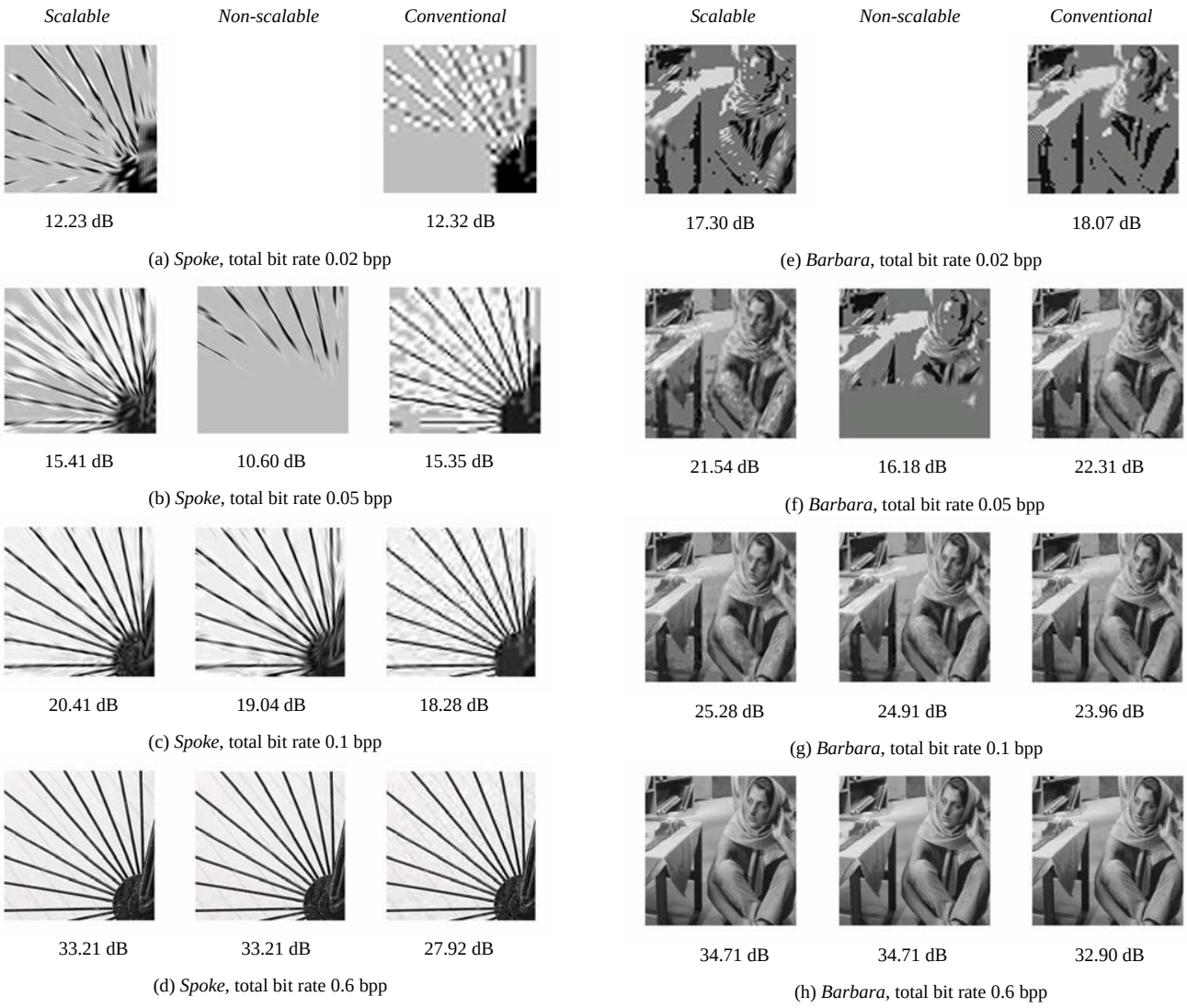


Figure 7. Reconstructed images of *Spoke* and *Barbara* by DA-DWT with the level-unit scalable direction representation (*Scalable*), DA-DWT with the non-scalable direction representation (*Non-scalable*), conventional DWT (*Conventional*) at different total bit rates. The PSNR is listed under each reconstructed image. Note that when the total bit rate is 0.02 bpp ((a) and (e)), there are no reconstructed images for the non-scalable case since the direction overhead is larger than the total bit rate.

6. D. Taubman, "Adaptive, non-separable lifting transforms for image compression," in *Proc. IEEE Int. Conf. on Image Processing, Kobe, Japan*, **vol. 3**, pp. 772–776, Oct. 1999.
7. W. Ding, F. Wu, and S. Li, "Lifting-based wavelet transform with directionally spatial prediction," in *Proc. IEEE Intl. Picture Coding Symposium 2004, San Francisco, CA, USA*, Dec. 2004.
8. C. L. Chang, A. Maleki, and B. Girod, "Adaptive wavelet transform for image compression via directional quincunx lifting," in *Proc. IEEE Workshop on Multimedia Signal Processing, Shanghai, China*, Oct. 2005.
9. C. L. Chang and B. Girod, "Direction-adaptive discrete wavelet transform via directional lifting and bandeletization," in *Proc. IEEE Intl. Conference on Image Processing, Atlanta, GA, USA*, Oct. 2006.
10. D. Wang, L. Zhang, A. Vincent, and F. Speranza, "Curved wavelet transform for image coding," *IEEE Trans. Image Processing* **vol. 15**, pp. 2413–2421, 2006.

11. A. Secker and D. Taubman, "Highly scalable video compression with scalable motion coding," *IEEE Trans. Image Processing* **vol. 13**, pp. 1029–1041, 2004.
12. F. W. S. L. R. Xiong, J. Xu and Y. Q. Zhang, "Layered motion estimation and coding for fully scalable 3D wavelet video coding," in *Proc. IEEE Intl. Conference on Image Processing, Singapore*, **vol. 4**, pp. 2271–2274, Oct. 2004.
13. 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.
14. I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Anal. Appl.* **vol. 4**, pp. 245–267, 1998.
15. C. Tian and S. S. Hemami, "An embedded image coding system based on tarp filter with classification," in *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing, Montreal, Quebec, Canada*, **vol. 3**, pp. 49–52, May 2004.