J. Vis. Commun. Image R. 23 (2012) 1128–1143

Contents lists available at SciVerse ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

A reduced-complexity image coding scheme using decision-directed wavelet-based contourlet transform

Chao-Hsiung Hung, Hsueh-Ming Hang *

Department of Electronics Engineering, National Chiao-Tung University, Hsinchu, Taiwan

ARTICLE INFO

Article history:
Received 1 November 2011
Accepted 18 June 2012
Available online 30 June 2012

Keywords:
Contourlet transform
Wavelet-based contourlet transform
Bit-plane coding
Directional filter bank
Directional transform
Wavelet transform
Image coding
Computational complexity reduction
Adaptive Directional Transform

ABSTRACT

Recently the wavelet-based contourlet transform (WBCT) is adopted for image coding because it matches better image textures of different orientations. However, its computational complexity is very high. In this paper, we propose three tools to enhance the WBCT coding scheme, in particular, on reducing its computational complexity. First, we propose short-length 2-D filters for directional transform. Second, the directional transform is applied to only a few selected subbands and the selection is done by a mean-shift-based decision procedure. Third, we fine-tune the context tables used by the arithmetic coder in WBCT coding to improve coding efficiency and to reduce computation. Simulations show that, at comparable coded image quality, the proposed scheme saves over 92% computing time of the original WBCT scheme. Comparing to the conventional 2-D wavelet coding schemes, it produces clearly better subjective image quality.

1. Introduction

Wavelet-based image coding method becomes a popular image compression topic in recent years [1,2]. For example, it was adopted by JPEG2000 [2] as an international image coding standard. Typical wavelet-based image coding scheme consists of three stages: two-dimensional discrete wavelet transform (2-D DWT), quantization, and arithmetic coding [2]. A digital image is first transformed by 2-D DWT to produce a set of transform coefficients. After quantization, these coefficients are compressed to a binary stream by an entropy coding tool.

However, the 2-D DWT is inefficient in representing the edge signals that are not aligned with the vertical or the horizontal axes [3]. Many 2-D directional transforms have thus been developed to solve this problem [5–8], [13–19]. Among them, the *wavelet-based contourlet transform* (WBCT) [19] technique has the critical-sampling property, consumes comparatively less computational power, and requires no side information for decoding. Therefore, we focus on WBCT in this study.

The arithmetic coding methods [9,10,21,22,33–35] are commonly adopted to compress the transformed/quantized coefficients. Particularly, the *embedded block coding with optimized truncation* (EBCOT) [9] technique was adopted by JPEG2000. It was originally designed for intra-subband coding. In this study, we adopt ESCOT (*3-D embedded subband coding with optimized truncation*) [33], which is an extension of EBCOT to inter-frame video coding, because our future work is targeting at video coding.

Combining WBCT and ESCOT, a WBCT image coding scheme can achieve a better coding performance than a regular 2-D DWT image coding scheme. However, there are a few issues in the existing WBCT coding schemes. They need a large amount of computations because the existing WBCT directional filters have a large support. And, we found that for a specific picture, some WBCT frequency subbands do not need further directional transform. Furthermore, the context table in ESCOT needs adjustment to match the characteristics of quantized WBCT coefficients.

To solve these issues, we propose three tools in this paper to enhance the WBCT image coding scheme. First, we suggest a set of short-length 2-D directional filters [38] and verify their performance. Second, we design a mean-shift-based decision scheme to dynamically select the proper subbands for directional transform [39]. Third, we re-design the context tables of ESCOT to match the data directionality. With these tools, our proposed scheme reduces 92% or higher the computational complexity of the original WBCT image coding scheme at similar visual quality [38].

The rest of this paper is organized as follows. Literature reviews and detailed problem statements are given in Section 2 and Section 3 describes the use of short-length 2-D filters. Section 4 presents a mean-shift-based decision algorithm for choosing the proper subbands for directional transform. Section 5 illustrates new entropy-coding context tables that are optimized for compressing the

* Corresponding author. Fax: +886 3 5731791.
E-mail address: hmhang@mail.nctu.edu.tw (H.-M. Hang).

1047-3203/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jvcir.2012.06.008

filtered coefficients. Experimental results are shown in Section 6. Finally, Section 7 concludes this paper.

2. Literature reviews and problem statements

Wavelet-based image coding systems typically consist of transform, quantization, and entropy coding. In this section, we briefly review the evolvement of directional transforms from 2-D DWT to WBCT. Then, we summarize the operations of ESCOT. When we put these two elements together, they form a conventional WBCT scheme.

2.1. Literature review: directional transform

2-D DWT is the tensor product of two *one-dimensional discrete wavelet transforms* (1-D DWT), and the Daubechies 9-7 wavelet filter is often in use. 1-D DWT can represent the piecewise smooth 1-D signals by a few coefficients [2]. But the outputs of 2-D DWT would contain many small coefficients for 2-D edges when these edges are not aligned with the vertical or the horizontal axes [3]. If we quantize these coefficients to zero, the coded image shows Gibbs artifacts along the edges [4]. To solve this problem, 2-D multiresolution transforms such as the *spatial-domain multiresolution directional transform* (SMDT) and the *frequency-domain multiresolution directional transform* (FMDT) [20] have been developed to improve the directionality of 2-D DWT. SMDT uses the adaptive directional 1-D DWT to align with the image texture direction [5–8]. In this way, SMDT can pack the signal energy into a few subbands. But it needs a huge amount of computation to decide a suitable direction and it also requires extra side information for carrying the direction information. On the other hand, FMDT uses a set of pre-selected 2-D filters to perform multiresolution directional decomposition [11–17]. Each filter corresponds to a basis function with specific spatial direction and resolution. FMDT can represent 2-D directional texture patterns by relatively few large coefficients. It needs less computational power and requires no side information for decoding. Therefore, we focus on the FMDT approach in this study.

2-D DWT and Laplacian pyramid (LP) [21] are two frequently used multiresolution transforms. Fig. 1(a) shows the filter bank structure of a 2-D DWT. After transform, it outputs four subband signals – HL (the horizontal high-pass and vertical low-pass subband signal), LH (the horizontal low-pass and vertical high-pass subband signal), HH (the horizontal high-pass and vertical high-pass subband signal), and LL (the horizontal low-pass and vertical low-pass subband signal). $G_1(z)$~$G_4(z)$ are the filters with specific pass bands and their output frequency partitions are given in Fig. 1(b). D_2 represents the decimation matrix, and $D_2 = 2I_2$, where I_2 is an identity matrix. 2-D DWT is a critical-sampling transform that keeps the same amount of data after one transform.

In contrast, Fig. 1(c) shows a typical LP wavelet system that does not satisfy the critical-sampling condition. It decomposes the input into one low-pass subband signal, LL, and one high-pass subband signal, H. $F_1(z)$ is the corresponding synthesis filter for the analysis filter $G_1(z)$. Fig. 1(d) shows the frequency partition of these two subbands. When the synthesized subband signal LL is subtracted from the original input, it produces the high-pass subband signal H. Without down-sampling, subband H is free from frequency scrambling. In this case, the LP system behaves as an over-sampling transform and it increases 25% data size after the transform.

The subband signal LL in LP (Fig. 1(c)) is identical to the subband signal LL in 2-D DWT when their $G_1(z)$ and D_2 are the same. That is, these two LL signals occupy the same frequency partition as in Fig. 1(b) and Fig. 1(d), respectively. In a multi-level 2-D DWT, the subband signal LL produced by the first 2-D DWT is further processed by the sub-sequent 2-D DWT's. Likewise, in a multi-level LP, the LL subband signal may be further processed by sub-sequent LP.

In the several variations of FMDT, *contourlet transform* (CT) [3][11] and *wavelet-based contourlet transform* (WBCT) [17] adopt the *directional filter banks* (DFB) structure [22] as shown in Fig. 2(a), where four 2-D filters and four decimation matrices are illustrated. These four 2-D filters decompose the input signal to four directional subbands. Each subband has a specific directional pass band. These 2-D filters, $A_1(z)$~$A_4(z)$, are fan filters and their corresponding output frequency partitions are drawn in Fig. 2(b). The decimation matrices rotate and down-sample the signals along specific directions. The first stage of CT uses LP to produce the subband signals, LL and H. It further uses the DFB in Fig. 2(a) to decompose the subband signal H into four directional subband signals; Fig. 2(c) shows the frequency partition of a typical CT. Because LP increases the data size, it is less preferred in the compression scenario. Therefore, another structure, WBCT, was proposed. It uses 2-D DWT to first generate four subbands, LL, HL, LH and HH. It further decomposes each of the three higher subband signals, HL, LH, and HH, by the DFB in Fig. 2(a). Fig. 2(d) shows the frequency partition produced by WBCT, wherein the twelve directional subbands are labeled from 1 to 12. WBCT has the critical-sampling property and it maintains the same data size. Thus, it is more desirable for compression purpose. Therefore, in this study we focus on the WBCT structure.

2.2. Literature review: arithmetic coding

After transform and quantization, we use arithmetic coding to compress the produced coefficients. Arithmetic coding algorithms [9,10,18,19,26–28] provide rather good compression efficiency by considering three types of correlations among the coefficients.

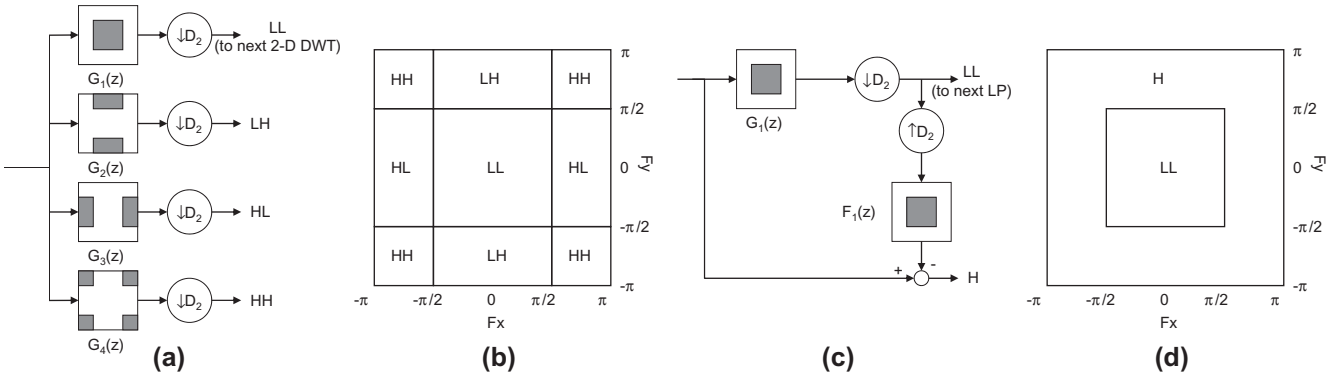

Fig. 1. (a) Filter bank structure of a 2-D DWT. (b) Frequency partitions produced by 2-D DWT. (c) Analysis filter structure of a LP. (d) Frequency partitions produced by LP.

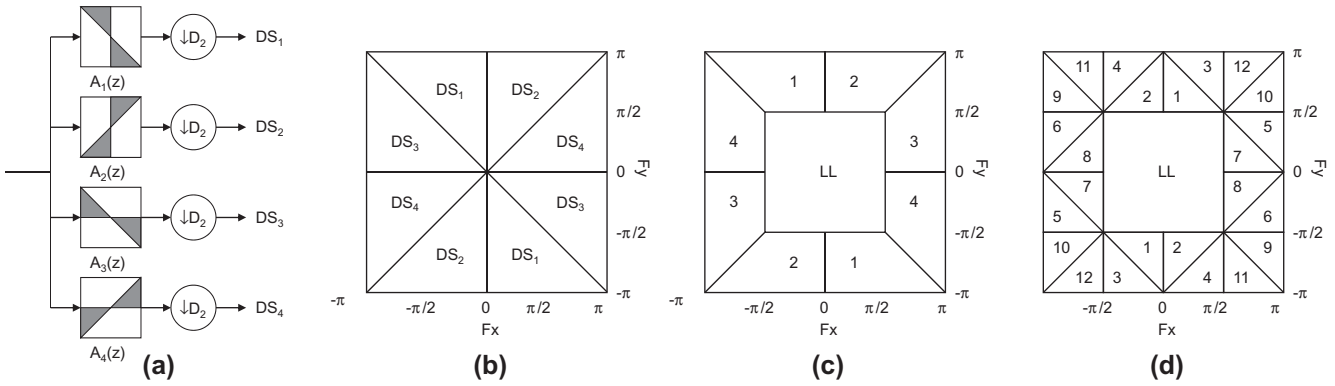

Fig. 2. (a) A four directional DFB structure. (b) Frequency partitions produced by the DFB in (a). (c) Frequency partition produced by CT. (d) Frequency partition produced by WBCT.

First, the inter-subband coding methods, such as the *set partitioning in hierarchical tree* (SPIHT) method [18] and the *embedded zerotrees of wavelet transform* (EZW) method [19], mitigate the inter-band correlations in a tree structure. Second, the intra-subband coding methods partition the coefficients in one subband to several non-overlapped coding blocks and handle only the correlations among the neighbors in one coding block (the intra-subband correlations). Examples in this category are the *embedded block coding with optimized truncation* (EBCOT) method [9], the *3-D embedded subband coding with optimized truncation* (ESCOT) method [26], and the *tarp-filter-based system that classifies coefficients to achieve embedding* (TCE) method [10]. Third, the mixed inter-subband and intra-subband coding methods cover both the inter-subband and intra-subband correlations. Examples are the *embedded conditional entropy coding of wavelet coefficients* (ECECOW) method [27] and the *embedded coding using zeroblocks of wavelet coefficients and context modeling* (EZBC) method [28]. To save computing power, for single image compression, we use the intra-subband coding methods in this study.

Among the intra-subband coding methods, EBCOT is popular for image coding and is adopted by JPEG2000; ESCOT is popular for video coding and was adopted by the wavelet video coding reference software in the MPEG scalable coding standard development. EBCOT's context model includes eight neighbors (3×3 square) of a to-be-processed coefficient. ESCOT extends this context model by considering all spatial-temporal neighbors in a $3 \times 3 \times 3$ cube. We adopt ESCOT to ease our future development in video coding.

2.3. Problem statement

We explore three issues of the current WBCT image coding scheme in this study: filter replacement, subband skipping, and context modification. The first target is to reduce the computation of filtering. In a typical WBCT scheme, the input image is first processed by a 2-D DWT and four subbands are generated: LL^1, HL^1, LH^1, and HH^1. Then, we filter the LL^1 subband signal again by 2-D DWT to obtain LL^2, HL^2, LH^2, and HH^2. Likewise, we recursively apply 2-D DWT to the LL^i subband, and produce LL^{i+1}, HL^{i+1}, LH^{i+1}, and HH^{i+1}, wherein '*i*' represents the 2-D DWT iterations. Also, we can use DFB in Fig. 2(a) to decompose HL^i, LH^i, and HH^i at the chosen levels.

LL^1 and its split subband signals (LL^i, HL^i, LH^i, and HH^i, where $i > 1$) contain the low and mid frequency components in the sensitive range of human visual system. When we apply the DFB in Fig. 2(a) on these subbands and quantize their transform coefficients, the ringing effects may appear on the smooth image regions. Thus, we tend to represent these coarse subband signals by 2-D DWT bases [16]. On the other hand, we apply the directional transform to HL^1, LH^1, and HH^1 to match their directional textures. But some of these subbands may be inappropriate for directional transform.

In Fig. 2(a), fan filters are constructed by using the hourglass filters. Typical hourglass 2-D filters are in the form of 23×23 and 45×45 matrices [23]. When we convolute them to generate $A_1(z) \sim A_4(z)$ in Fig. 2(a), the sizes of the fan filters are 45×45, 67×67, or 89×89. Because the fan filter sizes are large, the filtering process costs a huge amount of computation and may produce many small coefficients. Therefore, it is desirable to use shorter filters.

2-D DWT and ESCOT match quite well. The transformed/quantized coefficients typically have stronger energy along the horizontal and the vertical directions because these are the major directions of filers. This horizontal and the vertical direction properties are considered in the ESCOT design and thus the 2-D DWT+ESCOT scheme provides good compression performance. However, the major direction of transform coefficients produced by WBCT is decided by the directional filter of that subband. Applying the original ESCOT context tables to code these coefficients is inappropriate. Therefore, we need to construct new ESCOT context tables for WBCT subbands.

3. Short-length 2-D filters

To reduce computational load of the current WBCT, we design new short-length 2-D filters (SLF). The design procedure is as follows. We first choose an appropriate 1-D filter, up-sample it, and map it to a 2-D filter.

We begin our design from a 1-D type-II linear phase finite impulse response filter [23,25]. Eq. (1) is a 1-D prototype filter $\beta(z)$, wherein the coefficients $\{v_k\}$ satisfy (2) so that $\beta(e^{j0}) = 1$. When $N_1 = 1$ (short filter), $\beta(z)$ has a wide transition band. To keep a good balance between the transition band width and the filter length, we select $N_1 = 2$, and thus, $v_1 = 0.5916$ and $v_2 = -0.0982$. Fig. 3 (a) and (b) show the magnitude and the phase responses of $\beta(z)$.

We up-sample $\beta(z)$ by 2 and get $\beta(z^2)$. Fig. 3 (c) and (d) show the magnitude and the phase responses of $\beta(z^2)$. In Fig. 3 (d), $\beta(z^2)$ contains a phase discontinuity of π at frequency 0.5π. Because of this phase discontinuity, the left-side and the right-side amplitudes in Fig. 3 (c) have different signs.

$$\beta(z) = \sum_{k=1}^{N_1} v_k \cdot (z^{-N_1+k} + z^{-N_1-k+1}) \tag{1}$$

$$\sum_{k=1}^{N_1} v_k = 0.5 \tag{2}$$

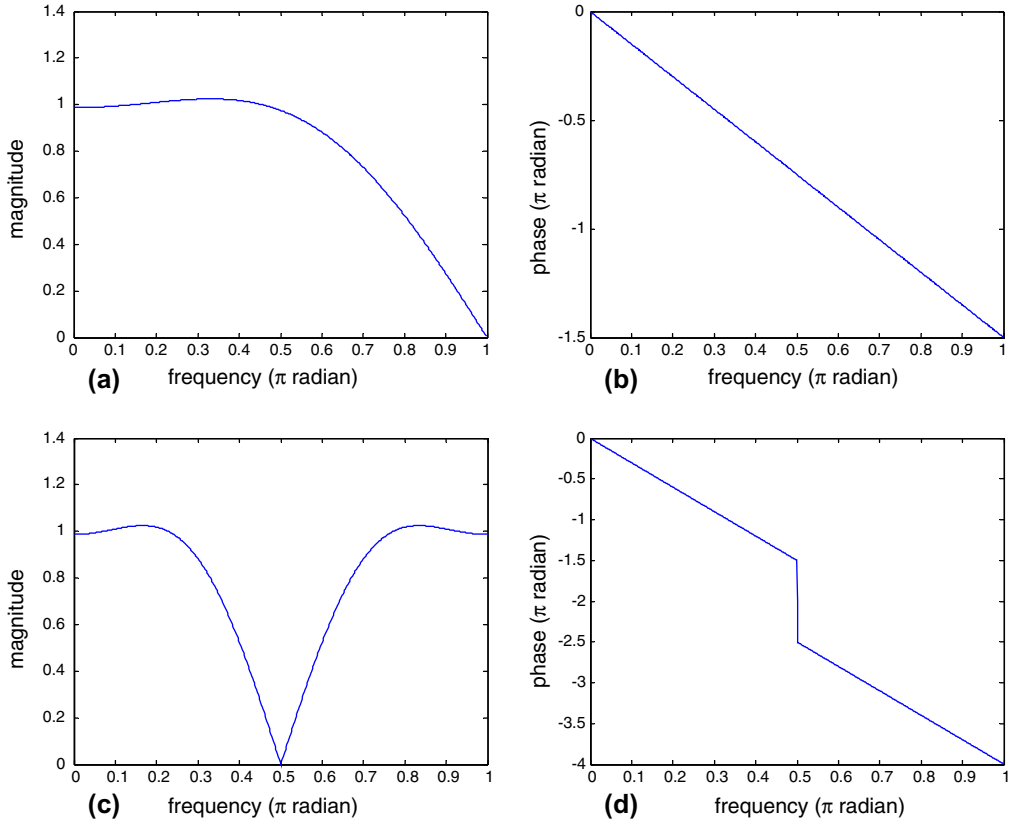

Fig. 3. (a) Magnitude response of $\beta(z)$. (b) Phase response of $\beta(z)$. (c) Magnitude response of $\beta(z^2)$. (d) Phase response of $\beta(z^2)$.

We then map $\beta(z^2)$ to a 2-D filter [24]. From $\beta(z^2)$, we derive the quadrant filters and rotate them by 45 degrees to construct the hourglass filters [12]. In Fig. 4, the symbol z_h denotes the horizontal frequency, and z_v denotes the vertical one. In Fig. 4 (a), we shift $\beta(z^2)$ by 0.5π along the negative frequency axis and the shifted $\beta(z^2)$ in horizontal direction is denoted by $\alpha(z_h{}^2)$. Similarly, the shifted $\beta(z^2)$ in vertical direction is denoted by $\alpha(z_v{}^2)$ in Fig. 4 (b). In Fig. 4 (c), we multiply $\alpha(z_h{}^2)$ and $\alpha(z_v{}^2)$ together to obtain a quadrant filter $\alpha(z_h, z_v)$. Accordingly, the four acquired quadrant filters are defined by (3)–(6) [12]. We rotate these quadrant filters by (7) to obtain the hourglass filters. In (7), an hourglass filter $A'(\omega)$ is obtained from a quadrant filter $A(\omega)$ [3], wherein Q_0 and Q_1 are the quincunx sampling matrices specified by (8).

$$H_0(z_h, z_v) = (1 + \alpha(z_h, z_v))/\sqrt{2} \tag{3}$$

$$H_1(z_h, z_v) = z_0(\sqrt{2} - (\sqrt{2}H_0(z_h, z_v) - 1)H_0(z_h, z_v)) \tag{4}$$

$$F_0(z_h, z_v) = -z_h^{-1}H_1(-z_h, z_v) \tag{5}$$

$$F_1(z_h, z_v) = z_h^{-1}H_0(-z_h, z_v) \tag{6}$$

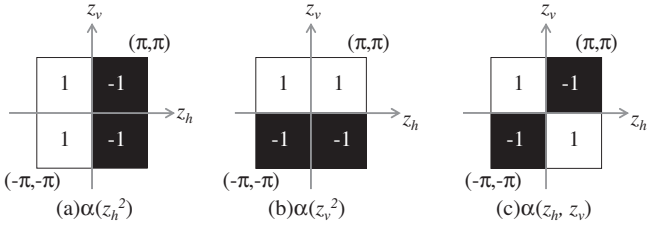

Fig. 4. Derivation of quadrant filters.

$$A'(\omega) = A(Q_0^{-T}\omega) = A(\frac{1}{2}Q_1^T\omega) = A(\frac{1}{2}Q_0\omega) \tag{7}$$

$$Q_0 = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, Q_1 = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \tag{8}$$

Fig. 5 shows a cascaded DFB structure [3]. The left half, $H'_0(z_h, z_v)$ and $H'_1(z_h, z_v)$, are the analysis filters, and the right half, $F'_o(z_h, z_v)$ and $F'_1(z_h, z_v)$, are the corresponding synthesis filters. The signals $DS_1 \sim DS_4$ are identical to those in Fig. 2(a) and their frequency partitions are in Fig. 2(b). This two-level analysis DFB structure consists of hourglass filters and quincunx sampling matrices. We rotate the quadrant filter $H_0(z_h, z_v)$ in (3) to obtain the hourglass filter $H'_0(z_h, z_v)$. $H'_1(z_h, z_v)$, $F'_o(z_h, z_v)$ and $F'_1(z_h, z_v)$ are designed similarly.

The sizes of our proposed 2-D hourglass short-length filters (SLF) are 7×7 and 13×13. They are much smaller than the sizes (23×23 and 45×45) of their corresponding long-length filters (LLF) [23]. Fig. 6 shows the magnitude responses of SLF and LLF. Although the transition band of SLF seems wider than that of the LLF, SLF matches the image local variation well due to its small size.

Table 1 shows the impacts of SLF and LLF on the DFB computational complexities. We compare two DFB implementations, direct structure and ladder structure, on the non-zero SLF/LLF coefficients. S is the size of input image. The numbers of multiplications and additions are proportional to S. The runtime is measured by running Matlab r2008b on a PC with Intel Core 2 Quad Q9400 CPU. The numbers of multiplications and additions include both convolution and down-sampling operations. When the sizes of the hourglass filters are 23×23, 45×45, 7×7 and 13×13, the numbers of nonzero coefficients are 145, 649, 17 and 65, respectively. For both the direct and the ladder structures, the SLF-based

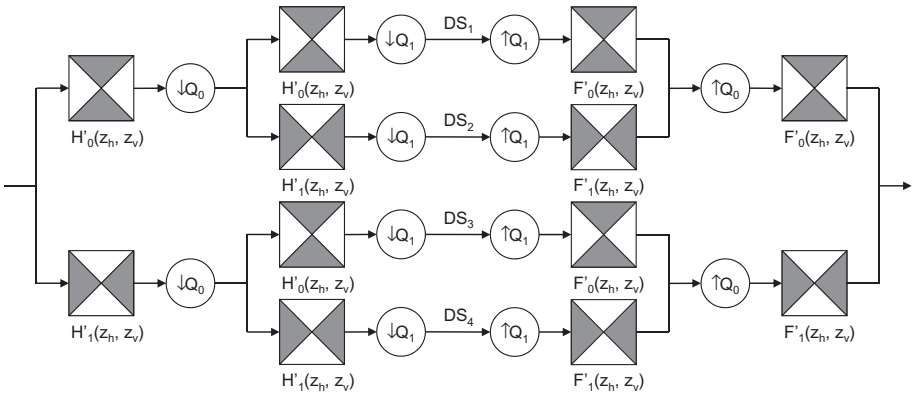

Fig. 5. A four-channel cascaded DFB.

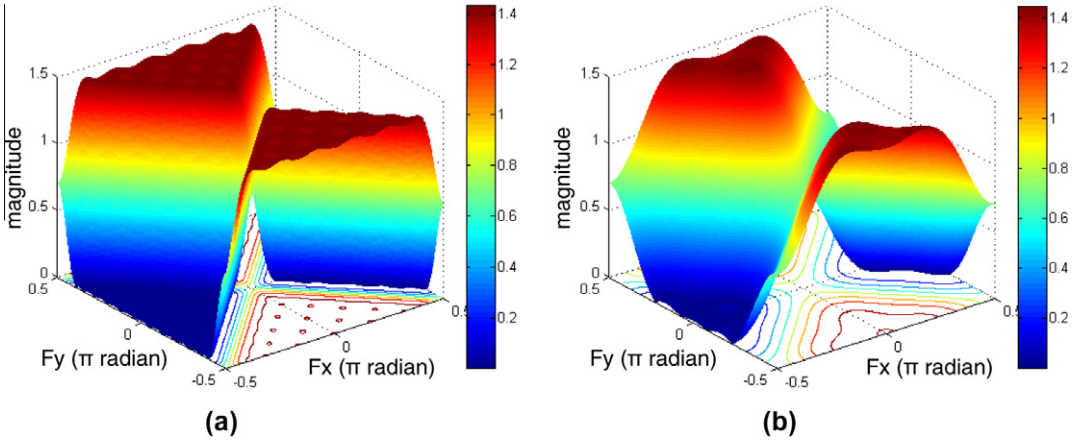

Fig. 6. (a) LLF, whose size = 23 × 23 [29]. (b) SLF, whose size = 7 × 7.

Table 1
The computational complexity and run time measured on the non-zero filter coefficients.

	LLF		SLF	
	Direct structure	Ladder structure	Direct structure	Ladder structure
Number of Multiplications	4S(145 + 649 + 2) = 3124S	4S(144 + 2) = 584S	4S(17 + 65 + 2) = 336S	4S(16 + 2) = 68S
Number of Additions	4S(145 + 649 + 2) = 3124S	4S(144 + 2) = 584S	4S(17 + 65 + 2) = 336S	4S(16 + 2) = 68S
S = 512 × 512	43.656 s	14.938 s	9.078 s	3.953 s
S = 256 × 256	10.906 s	3.813 s	1.797 s	0.854 s
S = 128 × 128	2.859 s	0.953 s	0.438 s	0.219 s

DFB takes approximately only 10% multiplications and additions of those of the LLF-based DFB. In the runtime profile, the SLF-based DFB saves roughly 80% computation time in both structures. The performance gap between our theoretical estimates (multiplications and additions) and experimental measurements (runtime) are largely due to data transfer (disk access).

4. Mean-shift-based decision on subband selection

In the WBCT image coding scheme, we apply the directional transform to the LH^1, HL^1, and HH^1 subbands. Yet, only the subband signal with significant energy in that direction would benefit from the directional transform. We thus try to locate the subbands with this property. Essentially, we identify the energy peaks and find their locations.

Mean shift technique is adopted to locate the energy peaks in the frequency spectrum. Mean shift is an iterative, nonparametric estimator of the peak location [32][33]; it finds a path to local maximum [34]. Let $\{x_i\}_{i=1\ldots n}$ be an arbitrary n-point data set in the d-dimensional Euclidean space R^d. First, we calculate the mean shift vector $m(x)$ by (9), wherein x is the center of current window, h is the window radius, and $K(x)$ is the flat kernel defined by (10). Then, we update the center by setting $m(x) + x$ as the center of the next window. We repeat this process until $m(x)$ converges to 0.

$$m(x) = \frac{\sum_{i=1}^{n} x_i K\left(\frac{x - x_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)} - x \tag{9}$$

$$K(x)\begin{cases} 1, & \text{if } \|x\| \leqslant 1 \\ 0, & \text{if } \|x\| > 1 \end{cases} \tag{10}$$

Fig. 7 shows our proposed mean-shift-based decision process for selecting the subbands. To illustrate the decision flow, we use a 512 × 512 pixel, 256 gray-level image as the input.

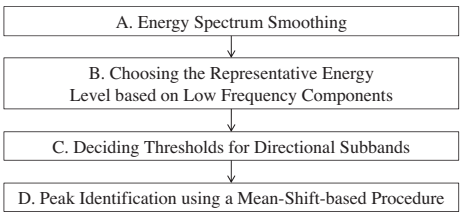

Fig. 7. The flowchart of the proposed mean-shift-based decision algorithm.

4.1. Energy spectrum smoothing

We calculate the input image frequency spectrum by the *2-D discrete Fourier transform* (2-D DFT). The frequency spectrum comprises 512×512 *discrete frequency components* (DFC). The DFC is generally a complex number with the form in (11) and their energy levels are in form of (12). Herein, (x,y) represents the coordinate pair of a DFC, $1 \leqslant x \leqslant 512$, and $1 \leqslant y \leqslant 512$.

$$m(x,y) = a(x,y) + b(x,y)i \tag{11}$$

$$c(x,y) = (a(x,y))^2 + (b(x,y))^2 \tag{12}$$

In Fig. 8, we copy the left-most column to the right-most column border and copy the up-most row to the bottom-most row border in order to get a symmetric energy spectrum. The zero frequency DFC is at (257,257).

Fig. 9 (a) shows the energy spectrum $c(x,y)$ of the input image Pepper, wherein the energy levels are in $\log_{10}$ scale, i.e., $\log_{10}(c(x,y))$. It contains many small peaks. These small peaks may cause misjudgment on cluster identification. Therefore, we use a smoothing operator (defined in Fig. 9 (c)) to reduce small peaks [31]. Fig. 9 (b) shows the smoothed energy spectrum. The large energy peaks typically stand out after smoothing.

4.2. Choosing the representative energy level based on the low frequency components

Fig. 9(b) shows natural images contain strong low frequency components. We choose it as the basis for calculating the threshold value for identifying energy peaks. Fig. 10(a) shows the subband signals generated by WBCT and Fig. 10(b) shows the DFC coordinates in the upper half subband LH 4-0. The gray area is called the *low frequency zone*, and the white area is the *high frequency zone*. Because the upper half subband is symmetric to its lower half, we only look at the DFC in the upper half of LH^1. The upper half of LH^1 is the region bounded by $129 \leqslant x \leqslant 385$ and $1 \leqslant y \leqslant 129$. Along each column x of LH^1, we calculate the mean $\mu(x)$ and the variance $\sigma(x)$ of the DFC by (13) and (14). We find that the DFC magnitudes in the center three columns ($256 \leqslant x \leqslant 258$) usually have large means and small variances. Similar property holds for HL^1. Therefore, we set the width of low frequency zone in LH^1 and HL^1 to 3 when the input image size is 512×512.

$$\mu(x) = \sum_{y=1}^{129} \frac{\log_{10} c(x,y)}{129}, \quad 129 \leqslant x \leqslant 385 \tag{13}$$

$$\sigma(x) = \sqrt{\sum_{y=1}^{129} \frac{(\log_{10} c(x,y))^2}{129} - \left(\sum_{y=1}^{129} \frac{\log_{10} c(x,y)}{129}\right)^2}, \quad 129 \leqslant x \leqslant 385 \tag{14}$$

To detect the peaks, we calculate the *representative* energy levels of the low frequency components. Eq. (15) computes the DFC mean of the LH^1 low frequency zone, and (16) computes that of the HL^1 low frequency zone. With these DFC means, we define the representative energy level LH_L for LH^1 by (17), and HL_L for HL^1 by (18). Essentially, we like to select a threshold that identifies the peaks with "significant" energy. In (17), when the average energy level of low frequency components in HL subband is at least four times larger than that in the LH subband, we use the former as the threshold; otherwise, the latter. The parameter "$\log_{10}(4)$" denotes the case that the large energy is at least 4 times of small ones. Correspondingly, the absolute magnitude of the large energy is at least twice of that of the small energy because the energy is the square of the absolute value. In this case, the difference in bit plan coding is significant.

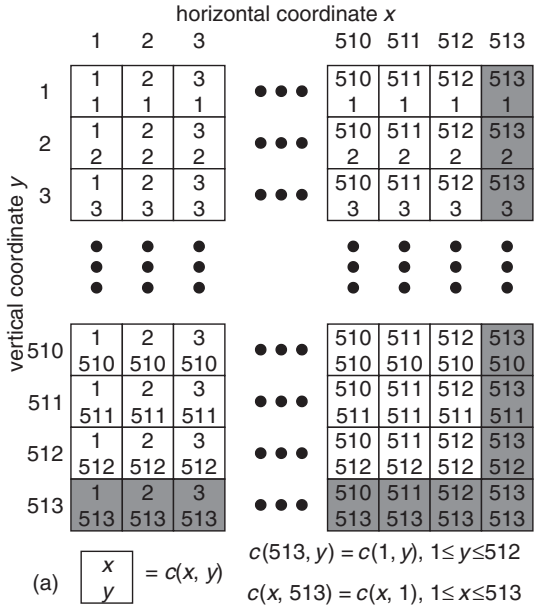

Fig. 8. The coordinates of energy coefficients $c(x, y)$. The padded data are in gray background.

$$LH\_\mu = \frac{\sum_{y=1}^{129}\sum_{x=256}^{258} \log_{10} c(x,y)}{3 \times 129} \tag{15}$$

$$HL\_\mu = \frac{\sum_{y=256}^{258}\sum_{x=1}^{129} \log_{10}(4) c(x,y)}{3 \times 129} \tag{16}$$

$$\text{if}((HL\_\mu - LH\_\mu) < \log_{10}(4)) LH\_L = LH\_\mu, \quad \text{else } LH\_L = HL\_\mu \tag{17}$$

$$\text{if}((LH\_\mu - HL\_\mu) < \log_{10}(4)) HL\_L = HL\_\mu, \quad \text{else } HL\_L = LH\_\mu \tag{18}$$

4.3. Deciding thresholds for directional subbands

A directional subband sometimes contains stronger energy level than the low frequency components. We consider this situation and adjust threshold in this step. We try to determine a peak detection threshold for every WBCT subband. Take the subband LH 4-0 as an example. We only look at the upper half of LH 4-0 because the DFCs in the upper half of LH 4-0 are symmetric to those in the lower half of LH 4-0. In the upper half of LH 4-0, we first consider only the DFC outside the low frequency zone. We calculate the mean LH_4-0_μ and the variance LH_4-0_σ outside the low frequency zone in LH 4-0, i.e., the $c(x,y)$ of white area in Fig. 10 (b), and construct a Gaussian distribution using the calculated mean and variance. In Fig. 11, each Gaussian distribution approximates its corresponding energy histogram well. Thus, the peak detection threshold for LH 4-0 is set by (19). The parameter b in

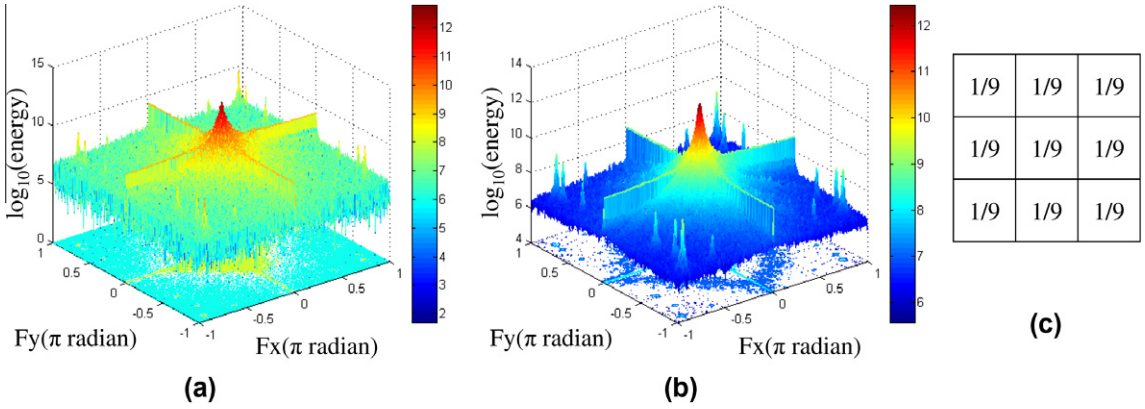

Fig. 9. (a) Energy spectrum of image Pepper. (b) Smoothed energy spectrum of image Pepper. (c) Smoothing operator.

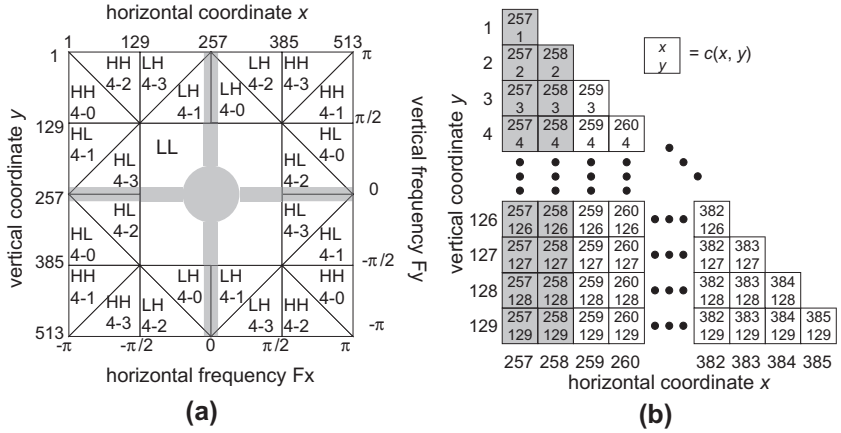

Fig. 10. (a) The subband frequency domain partition produced by WBCT. (b) The DFC coordinates in the upper half subband LH 4-0. The gray area in (a) and (b) is the low frequency zone.

(19) is chosen to be 0.7 because we like to eliminate the 75% DFC candidates. Together with the representative energy level LH_L defined earlier, 25% or fewer DFC candidates may be identified as energy peaks. We repeat similar procedures on LH 4-1∼LH 4-3, and HL 4-0∼HL 4-3.

Generally, the transmission priority of HH^1 is lower than the other subbands due to its lower information contents. Because of its low energy, we use the thresholds of its neighboring subbands to identify the energy peaks in HH^1. For example, we set the threshold HH_4-0_T of HH 4-0 by (20) using the parameters of HL 4-1.

$$\mathrm{LH\_4-0\_T} = \mathrm{Max}(\mathrm{LH\_L}, \mathrm{LH\_4-0}\_\mu + b \times \mathrm{LH\_4-0}\_\sigma) \tag{19}$$

$$\mathrm{HH\_4-0\_T} = \mathrm{Max}(\mathrm{HL\_L}, \mathrm{HL\_4-1}\_\mu + b \times \mathrm{HL\_4-1}\_\sigma) \tag{20}$$

4.4. Peak identification using a mean-shift-based procedure

We like to identify a directional subband that has significant energy by examining the discrete frequency components (DFC) of an image. This typically is caused by periodic texture patterns. And its corresponding DFC pattern is a cluster of DFCs with high energy. Thus, an energy peak in this paper is defined as a cluster of coefficients ($c(x,y)$ in a neighborhood) whose energy level is larger than the threshold. It has two properties: the energy level is high and these high-energy DFC coefficients are clustered in a small neighborhood. We use an image cluster identification scheme, Mean-Shift technique, to allocate them.

1) When a $c(x,y)$ within a directional subband and outside the low frequency zone is greater than the threshold of that subband, its location (x,y) is set to be the center of a search window. We then calculate its mass center coordinates $(x_{\mathrm{mass}},y_{\mathrm{mass}})$ by (21). The window size is chosen to be 11×11, or, roughly, its radius $r = 5$, because a small radius often leads to too many small peaks and a large radius sometimes misses peaks. In the search procedure, we extend the coefficients outside the subband boundary by periodic extension.

$$(x_{mass},y_{mass}) = \left(\frac{\sum_{m=x-5}^{x+5}\sum_{n=y-5}^{y+5} m \cdot c(m,n)}{\sum_{m=x-5}^{x+5}\sum_{n=y-5}^{y+5} c(m,n)}, \frac{\sum_{m=x-5}^{x+5}\sum_{n=y-5}^{y+5} n \cdot c(m,n)}{\sum_{m=x-5}^{x+5}\sum_{n=y-5}^{y+5} c(m,n)} \right) \tag{21}$$

We round x_{mass} and y_{mass} to the nearest integers and set the rounded (x_{mass},y_{mass}) as the center of next search window. Then, we use (21) again to update the mass center. We repeat this procedure until the rounded (x_{mass},y_{mass}) converges. Thus, a *peak candidate* is identified.

2) The number of the peak candidates is recorded by a table $d(x,y)$. The initial values of all entries of $d(x,y)$ are 0. When we identify a DFC at (x,y) as an energy peak candidate, we increase $d(x,y)$ by 1. When the table value of a specific location (x,y) is greater than 10 and it is also the largest $d(x,y)$ within a 3×3 window, the DFC located at (x,y) is judged as an *energy peak*. When one subband contains one or more energy peaks in the high frequency zone, it is considered to be suitable for directional decomposition.

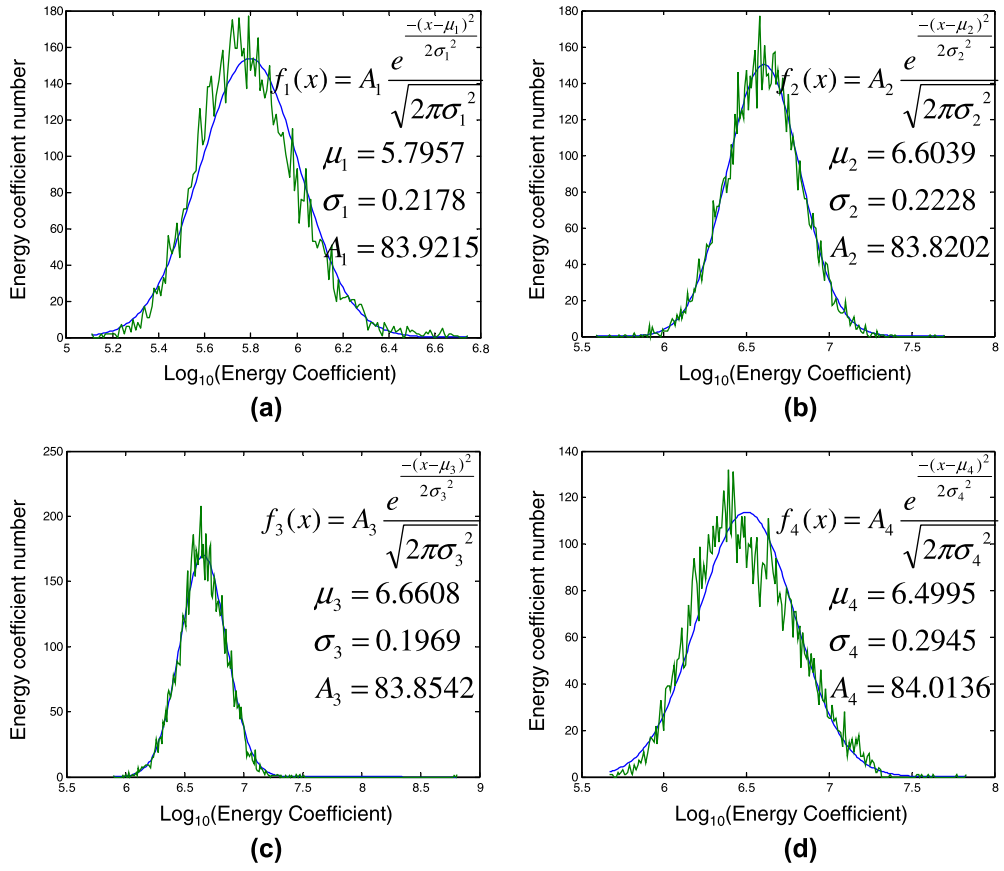

Fig. 11. The DFC energy histograms of some directional subbands. Each histogram is approximated by a Gaussian distribution. The directional subbands and the corresponding images are (a) LH 4-0 of Boat, (b) HL 4-3 of Lena, (c) LH 4-3 of Pepper, and (d) HL 4-0 of Fingerprint.

Table 2 shows some representative test images and their band-decomposition decision results for each subband. All images are images of 256 gray levels, and their sizes are 512×512 pixels. For each subband, the "(x,y)" column denotes the max energy peak location, and the "suitable for DT" column denotes the decision result. As Table 2 shows, the directional transform is inadequate for all subbands of the test image *Lena*; some subbands of *Barbara*, *Fingerprint*, *Pepper*, *Boat*, and *Couple* are suitable for directional transform, and all subbands of *Elaine* benefit from the directional transform. Fig. 12 shows the identified peaks by red dots. We fail to identify some peaks for two reasons. First, some peaks contain energy lower than the threshold. Second, when a peak is near the low frequency zone, clusters identified by the mean mean-shift scheme are occasionally in the low frequency zone. Fig. 19 shows a portion of test images *Barbara* and *Elaine*. They contain periodic signals. Identifying these signals in the spatial domain is hard. These periodic signals are corresponding to energy peaks in the frequency domain and thus we perform peak identification in frequency domain.

4.5. Computational Complexity

We now look at the computational complexity issue of our decision algorithm. We examine the amount of multiplications and additions for the steps in Fig. 7. We assume that the input image size is $S = W \times H$. Herein, W is the width of the input image and H is the height. We also assume that W and H are all power of 2 and we can implement the 2-D DFT in the radix-2 fast Fourier transform (FFT) structure.

1) In Step A of Fig. 7, we apply 2-D DFT to the input image, obtain its energy spectrum, and then apply a smoothing filter to the spectrum. The 2-D DFT is implemented by the radix-2 FFT, and thus the required numbers of real-value additions and multiplications are given by (22) and (23), in which ceil(x) means the smallest integer greater than or equal to x. Next, Eq. (12) needs 1 real addition and 2 real multiplications to calculate the energy of a DFC. For the entire image, the required numbers of real additions and real multiplications are in (24) and (25). The smoothing operator in Fig. 9(c) requires 8 real additions and 1 real multiplication

Table 2
Some test images, their max energy peak location in each subband ((x,y)) and the decision result for each subband (suitable for DT).

Wavelet subband	LH[1]		HL[1]		HH[1]	
Test image	(x,y)	Suitable for DT	(x,y)	Suitable for DT	(x,y)	Suitable for DT
Barbara	(213,126)	N	(130,366)	Y	(92,384)	N
Fingerprint	(257,129)	N	(128,200)	Y		N
Lena	(154,121)	N	(122,259)	N	(129,390)	N
Pepper		N		N	(32,24)	Y
Boat		N	(118,259)	Y		N
Couple	(236,96)	N	(109,259)	Y		N
Elaine	(177,123)	Y	(107,212)	Y	(83,32)	Y

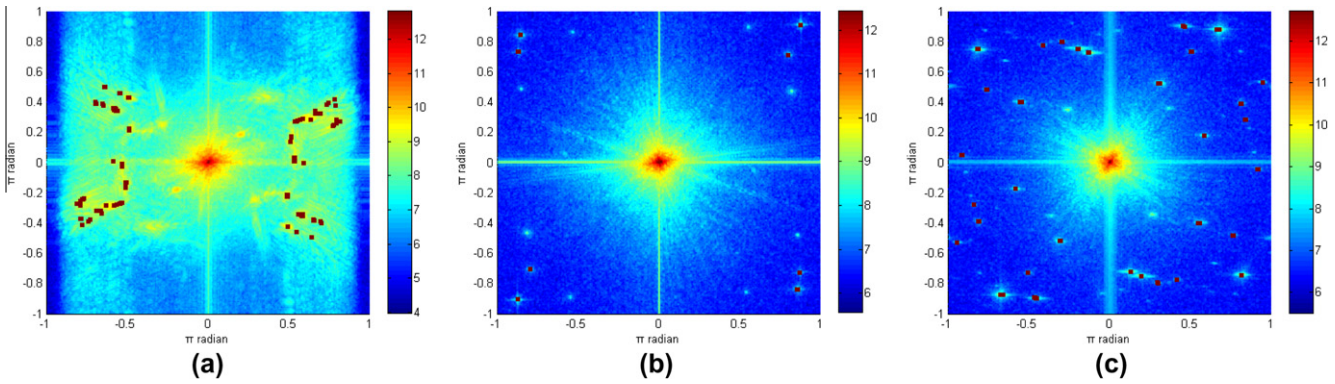

Fig. 12. Energy spectrum of test images (a) *Barbara*, (b) *Pepper*, and (c) *Elaine*. Horizontal axis and vertical axis represent horizontal frequency and vertical frequency, respectively. The energy spectrums are all in $\log_{10}$ scale. The red squares are the locations of the identified energy peaks.

for each $c(x,y)$. Thus, the total numbers of real additions and multiplications are given by (26) and (27). Finally, the overall numbers of real additions and real multiplications in Step A are (28) and (29).

$$N_{real\_addition\_in\_DFT} = 2 \times N_{complex\_multiplication\_in\_DFT} + 2 \times N_{complex\_addition\_in\_DFT} = 3 \times W \times H \times (\text{ceil}(\log_2 W) + \text{ceil}(\log_2 H)) \quad (22)$$

$$N_{real\_multiplication\_in\_DFT} = 4 \times N_{complex\_multiplication\_in\_DFT} = 2 \times W \times H \times (\text{ceil}(\log_2 W) + \text{ceil}(\log_2 H)) \quad (23)$$

$$N_{real\_addition\_in\_calculating\_power} = W \times H \quad (24)$$

$$N_{real\_multiplication\_in\_calculating\_power} = 2 \times W \times H \quad (25)$$

$$N_{real\_addition\_in\_smoothing\_spectrum} = 8 \times W \times H \quad (26)$$

$$N_{real\_multiplication\_in\_smoothing\_spectrum} = 2 \times W \times H \quad (27)$$

$$N_{real\_addition\_in\_stepA} = 3 \times W \times H \times (\text{ceil}(\log_2 W) + \text{ceil}(\log_2 H)) + W \times H + 8 \times W \times H \quad (28)$$

$$N_{real\_multiplication\_in\_stepA} = 2 \times W \times H(\text{ceil}(\log_2 W) + \text{ceil}(\log_2 H)) \times 2 \times W \times H + W \times H \quad (29)$$

2) Step B chooses the representative energy levels based on the low frequency zone. Eqs. (15) and (16) calculate the mean of the DFC energy in the low frequency zone. The heights of the low frequency zones in LH^1 and HL^1 are $(\text{ceil}(H/4) + 1)$ and $(\text{ceil}(W/4) + 1)$, respectively. The width is W_{lfz}. Thus, the mean calculation (Step B) needs 2 divisions and $N_{real\_addition\_in\_stepB}$ real additions as shown in (30). We choose $W_{lfz} = 3$ when $S = 512 \times 512$.

$$N_{real\_addition\_in\_stepB} = W_{lfz} \times (\text{ceil}(H/4) + \text{ceil}(W/4) + 2) - 2 \quad (30)$$

3) Step C decides the thresholds for directional subbands. The DFC number in each directional subband is $W \times H/16$, thus the DFC number in each half directional subband is $W \times H/32$. In addition to 2 real divisions, we need $W \times H/32$ real multiplications and $(W \times H/16\text{-}2)$ real additions to calculate the mean and the variance of each half directional subband. LH^1 and HL^1 together have 8 directional subbands in total. The numbers of real additions and real multiplications in step C are, therefore, given by (31) and (32).

$$N_{real\_addition\_in\_stepC} = 8 \times (W \times H/16 - 2) = W \times H/2 - 16 \quad (31)$$

$$N_{real\_multiplication\_in\_stepC} = 8 \times W \times H/32 = W \times H/4 \quad (32)$$

4) Step D identifies the energy peaks. Eq. (21) needs 1 division, 242 multiplications and 480 additions. In total, the numbers of real additions and real multiplications in step D are in (33) and (34), wherein N_{it} is the iteration number. In our experiments, the minimal N_{it} is 11 (test image *Baboon*), the maximal N_{it} is 12487 (test image *Barbara*), and the average N_{it} is 1697.

$$N_{real\_addition\_in\_stepD} = N_{it} \times 480 \quad (33)$$

$$N_{real\_multiplication\_in\_stepD} = N_{it} \times 242 \quad (34)$$

All in all, (35) and (36) give the total number of multiplications and additions in the decision procedure. When $S = W \times H = 512 \times 512$, $W_{lfz} = 3$, $N_{in} = 1697$, the total number of real additions and real multiplications are 17,461,460 and 10,699,826.

Table 3
Computational complexity and run time for the systems with and without decision when LLF is adopted.

	LLF without decision	LLF with decision (fastest)	LLF with decision (slowest)	Ratio (fastes) (%)	Ratio (slowest) (%)
Number of multiplications	114,819,072	10,699,826	125,518,898	9.32	109.32
Number of additions	114,819,072	17,461,460	132,280,521	15.21	115.21
Run time	11.613 s	1.385 s	13.012 s	11.93	112.05

Table 4
Computational complexity and run time for the systems with and without decision when SLF is adopted.

	SLF without decision	SLF with decision (fastest)	SLF with decision (slowest)	Ratio (fastest) (%)	Ratio (slowest)(%)
Number of multiplications	13,369,344	10,699,826	24,069,170	80.03	180.03
Number of additions	13,369,344	17,461,460	30,830,804	130.61	230.61
Run time	2.662 s	1.385 s	4.055 s	52.03	152.33

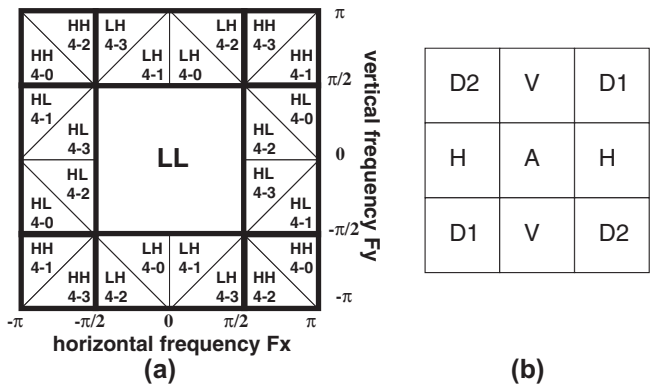

Fig. 13. (a) The directional subbands produced by WBCT. (b) The spatial neighbor directions for coefficient *A*.

$$N_{total\_real\_addition} = W \times H \times (3 \times \text{ceil}(log_2 W) + 3 \times \text{ceil}(log_2 H) + 9 + 1/2) - 16 + W_{lfz} \times (\text{ceil}(H/4) + \text{ceil}(W/4) + 2) - 2 + N_{it} \times 480 \quad (35)$$

$$N_{total\_real\_multiplication} = W \times H \times (2 \times \text{ceil}(\log_2 W) + 2 \times \text{ceil}(log_2 H) + 3 + 1/4) + N_{it} \times 242 \quad (36)$$

Table 3 and Table 4 show the computational complexity and the run time of the entire system with and without decision, wherein the directional filters are LLF and SLF, respectively. With decision, the fastest case occurs when no directional transform is conducted on LH^1, HL^1, and HH^1. And the slowest case occurs when we apply the directional transform to all subbands. In Table 3, the image coding scheme with LLF and decision may save over 84% computational load or 88% run time in the fastest case. In the slowest case, the decision process requires an additional 16% computational load or 13% run time. In Table 4, the image coding scheme with SLF and decision saves about 48% run time in the fastest case and consumes 52% extra run time in the slowest case. On the average, the image coding schemes with decision require less run time.

Table 5
ZC context table for 2-D wavelet subbands.

Wavelet subband	LL LH			HL			HH	
Context	H	V	D1+D2	V	H	D1+D2	H+V	D1+D2
8	2	X	X	2	X	X	X	⩾3
7	1	⩾1	X	1	⩾1	X	⩾1	2
6	1	0	⩾1	1	0	⩾1	0	2
5	1	0	0	1	0	0	⩾2	1
4	0	2	X	0	2	X	1	1
3	0	1	X	0	1	X	0	1
2	0	0	⩾2	0	0	⩾2	⩾2	0
1	0	0	1	0	0	1	1	0
0	0	0	0	0	0	0	0	0

5. New ZC context tables for ESCOT

Arithmetic coding methods encode the transformed/quantized coefficients into a bit-stream. ESCOT is a bit-plane coding method and it uses its neighbors for the *context model*. Let the sequence $x^N = \{x_N, x_{N-1}, \ldots, x_2, x_1\}$ represents one bit-plane of a coefficient block. Because the bit-plane consists of binary symbols, i.e., $x_i \in \{0, 1\}$, the minimum code length of a binary sequence estimated based on the information theory is shown in (37), wherein $P(x_i|x^{i-1})$ is the conditional probability of x_i given $x^{i-1} = \{x_{i-1}, x_{i-2}, \ldots, x_2, x_1\}$. Clearly, x^{i-1} is the subset of x^N. Assuming x^N is a Markov random sequence of some finite order, we then can reduce the size of x^{i-1} down to $\underline{x^{i-1}}$, which is a subsequence of x^{i-1}. This $\underline{x^{i-1}}$ is the context model support [26][27]. Typically, $\underline{x^{i-1}}$ includes the neighbors and the (bit-plane) parents of x_i. Ideally, the optimal context model gives the maximum mutual information [29].

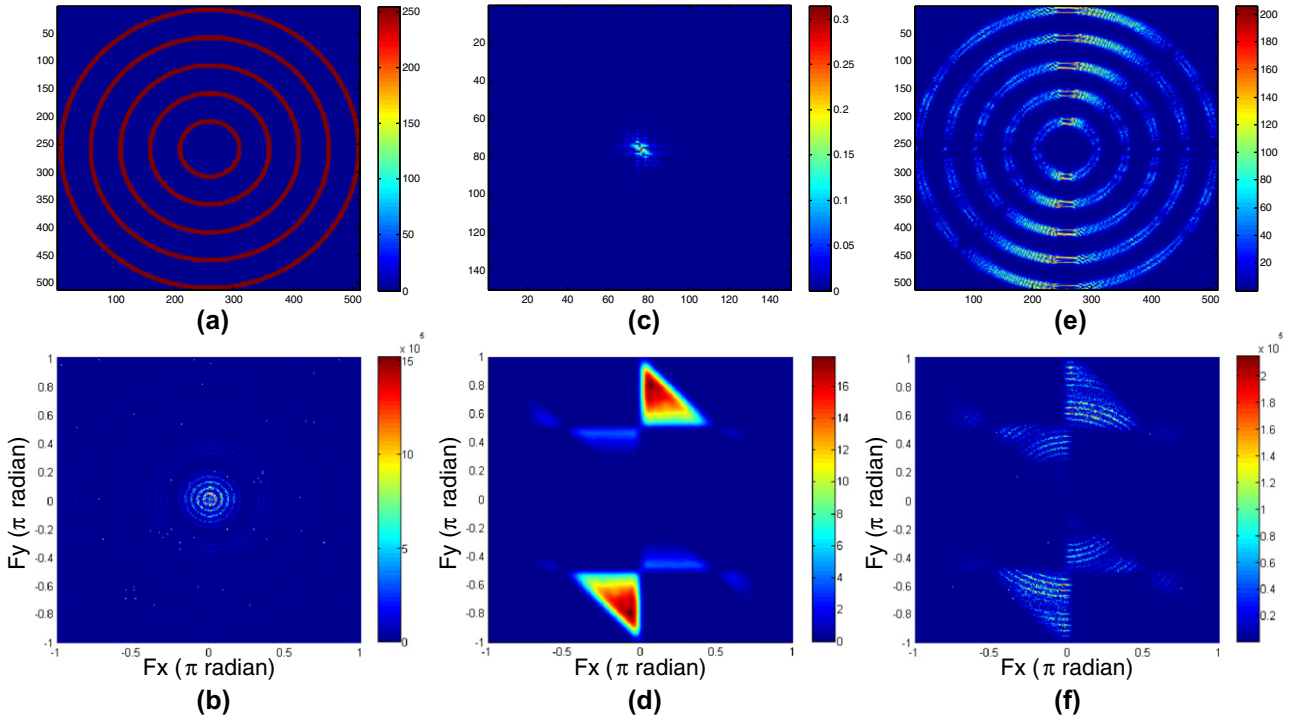

Fig. 14. (a) Input signal in spatial domain. (b) Input signal in frequency domain. (c) Filter response of DF_LH 4-0 in spatial domain. (d) Filter response of DF_LH 4-0 in frequency domain. (e) Output signal in spatial domain. (f) Output signal in frequency domain.

Table 6
ZC context table for directional subbands.

Directional subband	LH 4-0			LH 4-1			HL 4-2			HL 4-3			LH 4-3 HL 4-1 HH 4-0 HH4-2			LH 4-2 HL 4-0 HH 4-1 HH 4-3		
Context	D2+H	V	D1	D1+H	V	D2	D2+V	H	D1	D1+V	H	D2	D1	H+V	D2	D2	H+V	D1
8	$\geqslant 2$	X	X	$\geqslant 2$	X	X	$\geqslant 2$	X	X	$\geqslant 2$	X	X	2	X	X	2	X	X
7	1	$\geqslant 1$	X	1	$\geqslant 1$	X	1	$\geqslant 1$	X	1	$\geqslant 1$	X	1	$\geqslant 1$	X	1	$\geqslant 1$	X
6	1	0	$\geqslant 1$	1	0	$\geqslant 1$	1	0	$\geqslant 1$	1	0	$\geqslant 1$	1	0	$\geqslant 1$	1	0	$\geqslant 1$
5	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
4	0	2	X	0	2	X	0	2	X	0	2	X	0	$\geqslant 2$	X	0	$\geqslant 2$	X
3	0	1	X	0	1	X	0	1	X	0	1	X	0	1	X	0	1	X
2	0	0	2	0	0	2	0	0	2	0	0	2	0	0	2	0	0	2
1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

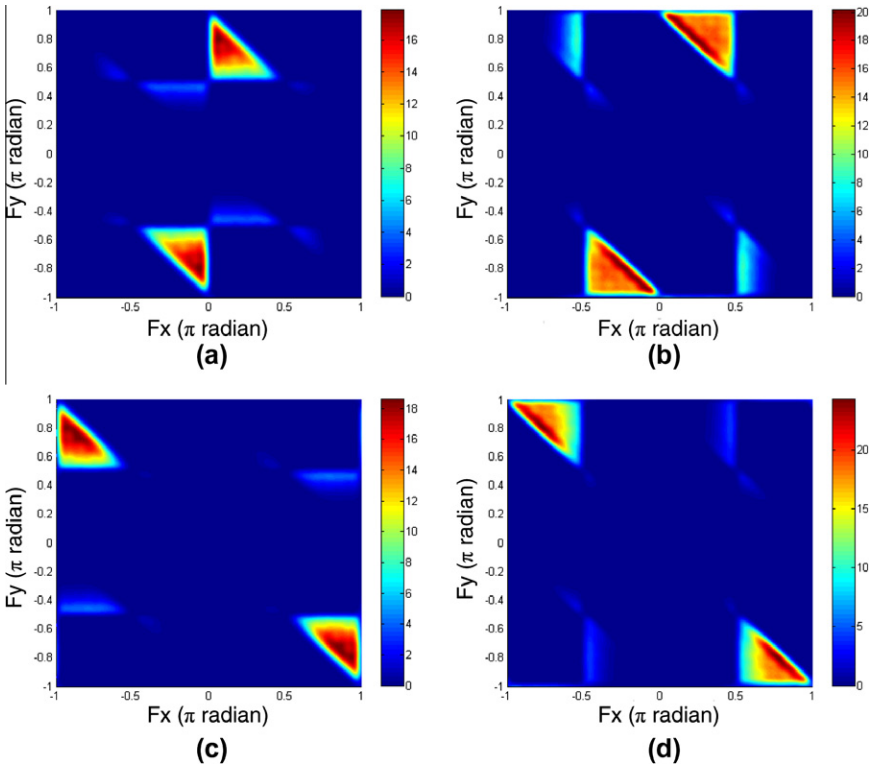

Fig. 15. Frequency magnitude responses of (a) LH 4-0 (b) LH 4-2 (c) HH 4-0 (d) HH 4-2.

$$L = -\log_2 \prod_{i=1}^{n} P(x_i | x^{i-1}). \tag{37}$$

The original ESCOT considers only the 2-D DWT coefficients in the horizontal and the vertical directions. Yet, the coefficients in a certain directional subband may cluster along one specific direction (different from the vertical or horizontal directions). The original context table fails to handle this case well. Therefore, we redesign the context models of ESCOT.

In Fig. 13(a), the 13 subbands produced by WBCT are labeled as "LL", "HH 4-0", "LH 4-0", "HL 4-0", and likewise. In Fig. 13(b), the edges passing through A can be H-A-H (0^O), V-A-V (90^O), D1-A-D1 (45^O), and D2-A-D2 (-45^O). We denote the 0^O, 90^O, 45^O, -45^O directions as "H", "V", "D1", and "D2", respectively.

In Fig. 14, we examine the effect of the directional filter LH 4-0 (DF_LH 4-0). A concentric-circle pattern, which has edges of all directions, is used as the input pattern. Fig. 14 (a) and (b) show this input signal and its frequency spectrum. Fig. 14 (c) shows the spatial filter impulse response of DF_LH 4-0, which is roughly along

Table 7
Abbreviations for the adopted tools in the image coding scheme.

Directional Transform	
SLF	Short length directional Filter.
LLF	Long length directional Filter.
NDF	No directional Filter.
Decision	
NDS1	No decision, applying directional transform on all subbands (LH^1, HL^1, and HH^1).
NDS2	No decision, directional transform not applied.
WDS	With decision, applying directional transform on the chosen wavelet subbands.
Entropy coder	
O	ESCOT with the original ZC context tables.
P	ESCOT with the proposed ZC context tables.

the H direction (slightly tilted to the D2 direction). Fig. 14 (d) shows the filter frequency magnitude response of DF_LH 4-0, whose energy clusters mainly along the vertical axis. In Fig. 14 (e), the filtered output image contains mainly the spatial edges aligned with the H direction (slightly tilted to the D2 direction). Fig. 14 (f) shows the frequency spectrum of filtered signals. Evidently, the dominated directions of the LH 4-0 outputs are H and D2. Hence, "H and D2" are the filtered directions of LH 4-0.

Similarly, we identify the filtered directions of the other directional subbands. The filtered directions of LH 4-1 are "H and D1", those of HL 4-2 are "V and D2", and those of HL 4-3 are "V and D1". The filtered directions of the four corner subbands (LH 4-2, HH 4-3, HH 4-1, and HL 4-0) are D2. And those of the other four corner subbands (LH 4-3, HH 4-2, HH 4-0, HL 4-1) are D1.

ESCOT uses three types of context models or *context tables* – the zero coding tables (ZC), the sign coding tables (SC) and the magnitude refinement tables (MR). ESCOT codes bit-planes from the most significant bit-plane to the least significant bit-plane. ESCOT starts with ZC to code the beginning zeros until it hits the first non-zero bit. ESCOT uses ZC to code the magnitude of first non-zero bit

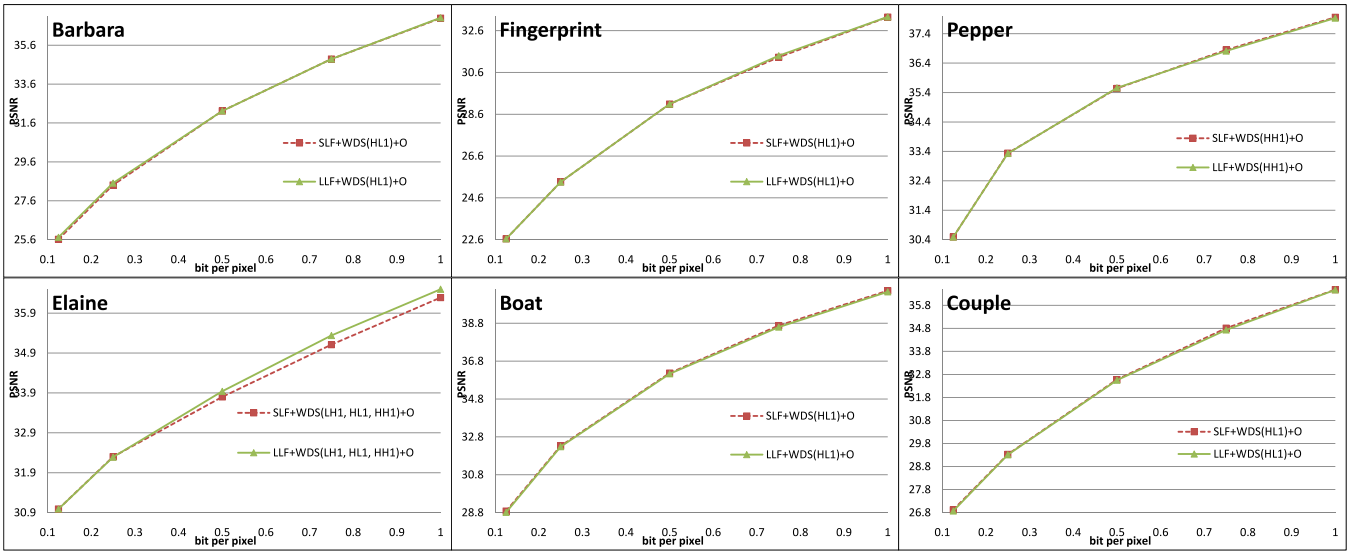

Fig. 16. PSNR of the image coding schemes with SLF and LLF ("SLF+WDS+O" and "LLF+WDS+O").

Table 8
Run time of the image coding schemes with SLF and LLF.

Scheme	SLF+WDS(HL^1)+O (Barbara, Fingerprint, Boat, Couple, average)	SLF+WDS(HH^1)+O (Pepper)	SLF+WDS(LH1, HL^1, HH^1)+O (Elaine)
Run time	4.547 s	4.550 s	8.023 s
Scheme	LLF+WDS(HL^1)+O (Barbara, Fingerprint, Boat, Couple, average)	LLF+WDS(HH^1)+O (Pepper)	LLF+WDS(LH1, HL^1, HH^1)+O (Elaine)
Run time	23.031 s	23.026 s	62.484 s

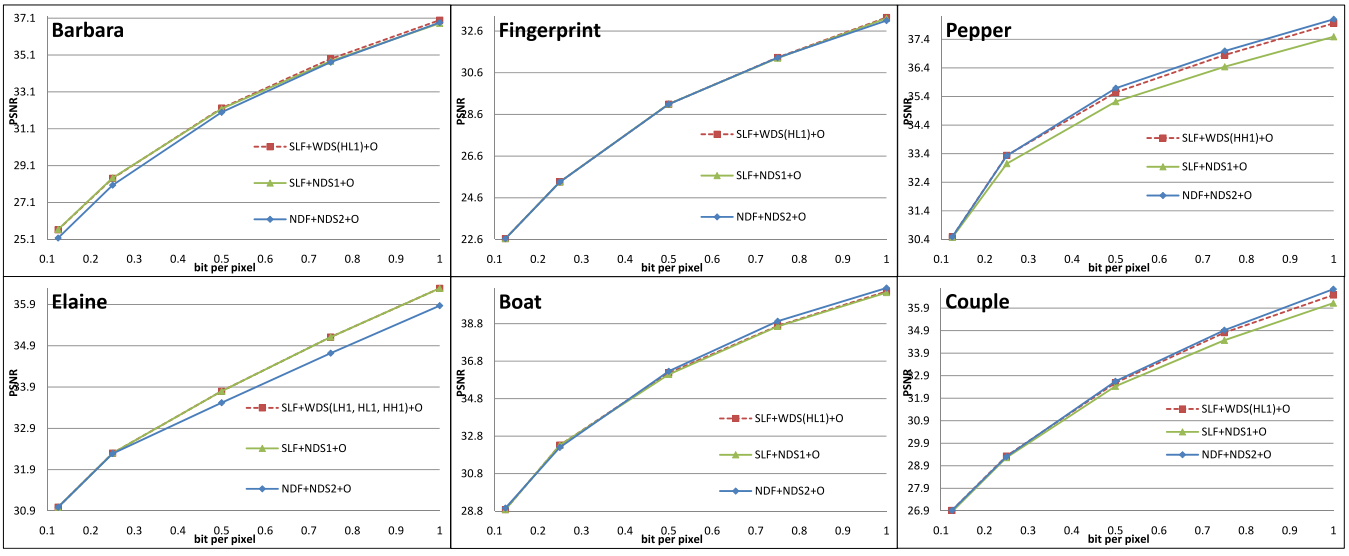

Fig. 17. PSNR of the image coding schemes with and without decision ("SLF+WDS+O", "SLF+NDS1+O", and "NDF+NDS2+O").

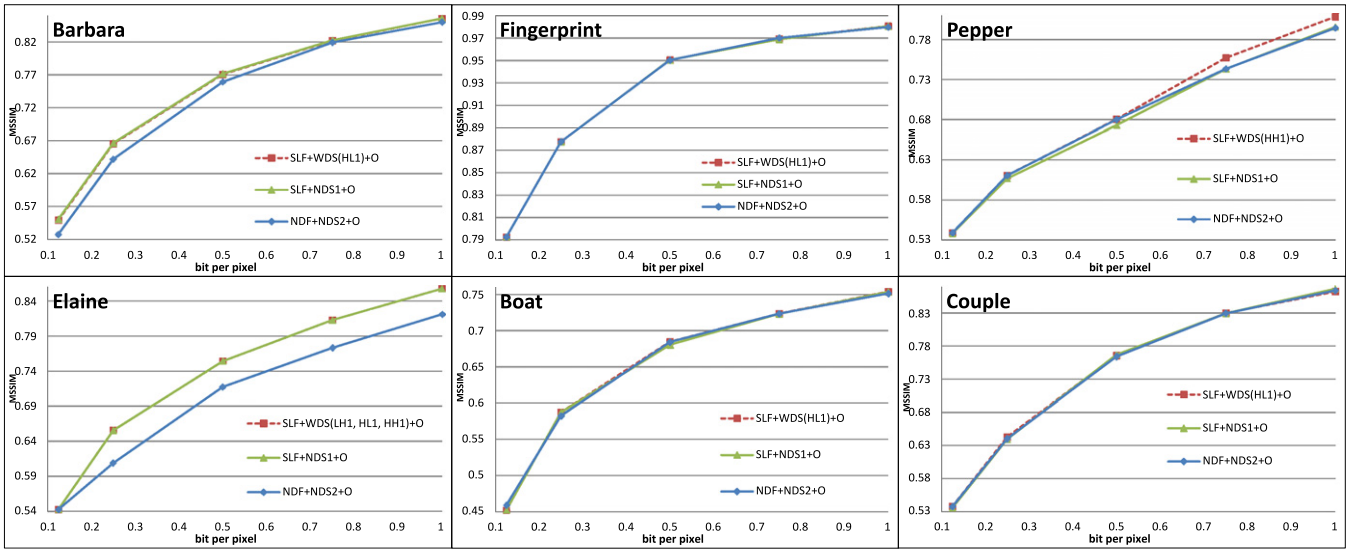

Fig. 18. MSSIM of the image coding schemes with and without decision ("SLF+WDS+O", "SLF+NDS1+O", and "NDF+NDS2+O").

Table 9
Average Run Time of the Image Coding Schemes with and without Decision.

Scheme	SLF+WDS+O	SLF+NDS1+O	NDF+NDS2+O
Run time	4.804 s	8.206 s	2.688 s

and SC to code its sign. For the remaining bits, ESCOT uses MR to code their magnitudes. To match the characteristics of the WBCT coefficients, we alter the ZC context table in ESCOT. For the coefficients in the ordinary 2-D wavelet subbands, we adopt the ZC context table (Table 5) in EBCOT [9]. But for the coefficients in the directional subbands, the proposed Table 6 is the ZC context table.

In Table 5 and Table 6, each "context" denotes a model, and the numbers of non-zero coefficients are listed under the directions, H, V, and D1+D2, and X denotes "Do not care". Fig. 13(b) shows the neighbors and their notations we use in the entropy coding. The neighbors include vertical neighbors (V), horizontal neighbors (H), left-lower and right-upper neighbors (D1), and left-upper and right-lower neighbors (D2). To code coefficient A in a wavelet subband of a bit-plane, we first calculate the number of non-zero coefficients in all directions. For 2-D wavelets, based on the subband location and the non-zero coefficient patterns, we decide which context in Table 5 is to be used to code this bit of coefficient A. Similarly, we code the coefficients in the other directional subbands using Table 6.

Fig. 15 shows the frequency responses of the WBCT directional filters. We notice the aliasing phenomenon in WBCT. Because the directional filters are not ideal filters, their outputs contain aliasing components. Thus, the outputs of a certain filter populate not only along one direction but also along another direction (with less energy). Consequently, the context model in arithmetic coding becomes less accurate or its coding efficiency is reduced. We may reduce aliasing by adopting a sharper (and thus longer) filter but the computation time would then increase.

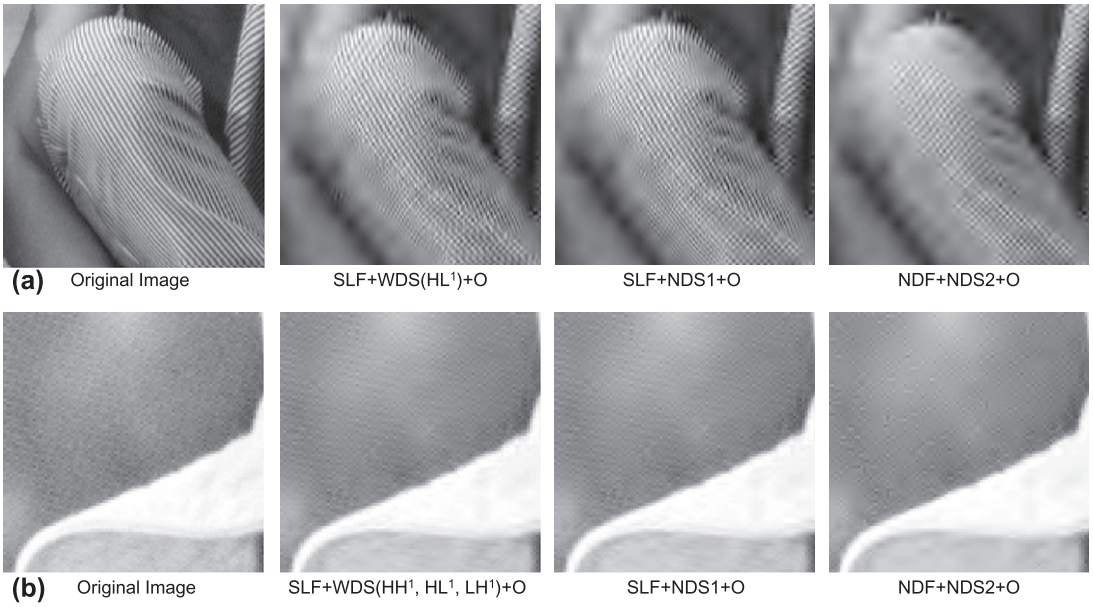

Fig. 19. (a) Portions of the original and the reconstructed images of *Barbara* at 0.125bpp. (b) Portions of the original and the reconstructed images of *Elaine* at 0.5bpp.

Table 10
PSNR of the image coding schemes with the original and the new ZC context tables (directional filters = SLF).

Test image	Coding Shceme	0.125 bpp	0.25 bpp	0.5 bpp	0.75 bpp	1.0 bpp
Barbara	SLF+WDS(HL^1)+O	25.62	28.41	32.22	34.89	36.99
	SLF+WDS(HL^1)+P	25.79	28.53	32.33	34.96	37.11
Fingerprint	SLF+WDS(HL^1)+O	22.64	25.36	29.09	31.33	33.25
	SLF+WDS(HL^1)+P	22.64	25.52	29.09	31.33	33.25
Pepper	SLF+WDS(HH^1)+O	30.49	33.34	35.54	36.85	37.96
	SLF+WDS(HH^1)+P	30.6	33.37	35.61	36.82	37.95
Elaine	SLF+WDS(LH^1, HL^1, HH^1)+O	30.99	32.3	33.8	35.11	36.29
	SLF+WDS(LH^1, HL^1, HH^1)+P	31.09	32.31	33.84	35.12	36.37
Boat	SLF+WDS(HL^1)+O	28.88	32.32	36.17	38.68	40.52
	SLF+WDS(HL^1)+P	28.9	32.42	36.26	38.78	40.58
Couple	SLF+WDS(HL^1)+O	26.92	29.33	32.58	34.81	36.48
	SLF+WDS(HL^1)+P	26.92	29.39	32.6	34.85	36.63

Table 11
PSNR of the image coding schemes with the original and the new ZC context tables (directional filters = LLF).

Test image	Coding Shceme	0.125 bpp	0.25 bpp	0.5 bpp	0.75 bpp	1.0 bpp
Barbara	LLF+WDS+O	25.72	28.51	32.22	34.89	37.01
	LLF+WDS+P	25.86	28.71	32.41	34.96	37.11
Fingerprint	LLF+WDS+O	22.64	25.36	29.09	31.33	33.25
	LLF+WDS+P	22.64	25.52	29.09	31.41	33.26
Pepper	LLF+WDS+O	30.49	33.33	35.56	36.81	37.93
	LLF+WDS+P	30.6	33.37	35.62	36.9	38.07
Elaine	LLF+WDS+O	30.99	32.29	33.94	35.34	36.5
	LLF+WDS+P	31.09	32.33	34	35.38	36.53
Boat	LLF+WDS+O	28.81	32.28	36.13	38.6	40.46
	LLF+WDS+P	28.8	32.39	36.22	38.67	40.58
Couple	LLF+WDS+O	26.87	29.31	32.55	34.73	36.47
	LLF+WDS+P	26.93	29.37	32.56	34.79	36.53

Table 12
Run time of the image coding schemes with different ZC context tables.

Scheme	SLF+WDS(HL^1)+O (Barbara, Fingerprint, Boat, Couple, average)	SLF+WDS(HH^1)+O (Pepper)	SLF+WDS(LH^1, HL^1, HH^1)+O (Elaine)
Run time	4.547 s	4.550 s	8.023 s
Scheme	SLF+WDS(HL^1)+P (Barbara, Fingerprint, Boat, Couple, average)	SLF+WDS(HH^1)+P (Pepper)	SLF+WDS(LH^1, HL^1, HH^1)+P (Elaine)
Run time	4.203 s	4.177 s	7.813 s
Scheme	LLF+WDS(HL^1)+O (Barbara, Fingerprint, Boat, Couple, average)	LLF+WDS(HH^1)+O (Pepper)	LLF+WDS(LH^1, HL^1, HH^1)+O (Elaine)
Run time	23.031 s	23.026 s	62.484 s
Scheme	LLF+WDS(HL^1)+P (Barbara, Fingerprint, Boat, Couple, average)	LLF+WDS(HH^1)+P (Pepper)	LLF+WDS(LH^1, HL^1, HH^1)+P (Elaine)
Run time	22.391 s	22.386 s	62.256 s

6. Experimental results

We have discussed the three proposed tools that enhance a WBCT image coding scheme in computation and/or complexity reduction. They are short length 2-D filters, a mean-shift-based decision, and new ZC context tables for ESCOT. In this section, we examine the impact of each tool towards the system performance. And, putting them together, we compare the overall performance between the 2-D DWT image coding scheme, the original WBCT image coding scheme, and the proposed WBCT image coding scheme with three new tools.

A few abbreviations are explained below. The original WBCT image coding scheme can apply directional filtering to either all subbands (NDS1) or no subband (NDS2). With our decision mechanism (WDS), we adaptively choose the subbands for directional filtering. Moreover, the original WBCT scheme uses long length directional filters (LLF), and our proposed image coding scheme uses short length directional filters (SLF) instead. The no directional filtering (NDF) situation appears when either the WDS declares that no subband needs directional filtering or the NDS2 strategy is adopted. There are two options for ESCOT: the original context tables (O) designed for 2-D DWT coefficients or the proposed context tables (P) fine-tuned for the WBCT coefficients. Table 7 summarizes all the aforementioned abbreviations.

The notation of an image coding scheme consists of three parts: the directional transform type, the decision, and the coder tables. For example, the 2-D DWT image coding scheme is "NDF+NDS2+O", the original WBCT image coding scheme is "LLF+NDS1+O", and our proposed coding scheme with three tools is "SLF+WDS(HL^1, HH^1)+P". Note that the subbands selected by WDS are listed in the parenthesis after WDS, and thus "WDS(LH^1, HL^1, and HH^1)" is the same as "NDS1".

Our test images are listed in Table 2. The experimental platform is Matlab r2008b on a PC with Intel Core 2 Quad Q9400 CPU. First,

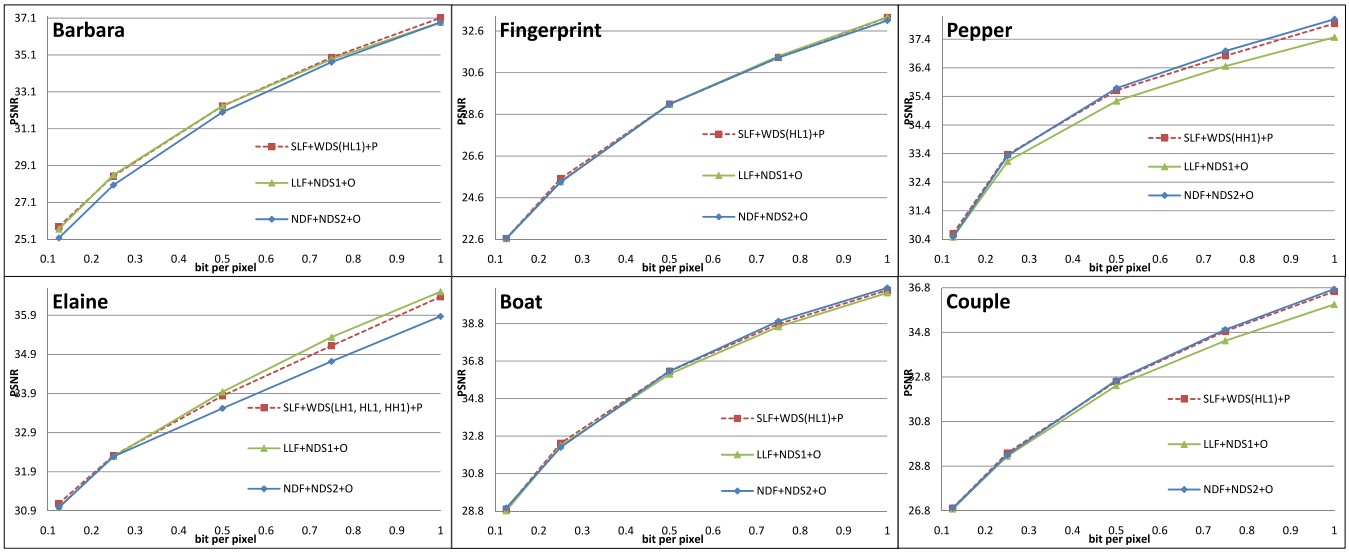

Fig. 20. PSNR of the 2-D DWT scheme (NDF+NDS2+O), the original WBCT scheme (LLF+NDS1+O), and the proposed scheme with three new tools (SLF+WDS+P).

we show the impacts of filter length in terms of PSNR and run time by comparing "SLF+WDS+O" and "LLF+WDS+O". Fig. 16 shows their PSNR at various bitrates (bit per pixel, bpp). Obviously, the image coding scheme with SLF has similar PSNR performances as that with LLF. Table 8 shows the run time of these two schemes and the image coding scheme with SLF consumes only 10–20% computational time of that with LLF.

Next, we present the impacts of decision algorithm in terms of PSNR, MSSIM [35] and run time among "SLF+WDS+O", "SLF+NDS1+O", and "NDF+NDS2+O" (the 2-D DWT coding scheme). MSSIM represents *mean of structural similarity*. A higher MSSIM implies a better image subjective quality. Fig. 17 shows the PSNR of the image coding schemes with and without decision. The image coding scheme with decision ("SLF+WDS+O") has similar PSNR performance as those without decisions ("SLF+NDS1+O" and "NDF+NDS2+O"). Fig. 18 shows the MSSIM of the image coding schemes with and without decision. Our proposed image coding scheme with decision ("SLF+WDS+O") has similar MSSIM performance as "SLF+NDS1+O" and has better MSSIM than "NDF+NDS2+O". The visual quality improvement is most obvious on some pictures such as *Elaine*. Fig. 19 shows portions of the original and the reconstructed images of *Barbara* and *Elaine* generated by these three schemes. Noticeably, "SLF+WDS+O" and "SLF+NDS1+O" show more texture details than "NDF+NDS2+O". Table 9 shows the run time of these schemes. "SLF+WDS+O" saves about 50% computational time comparing to "SLF+NDS1+O" but it needs roughly 70% extra computational time comparing to "NDF+NDS2+O". In brief, the image coding scheme with decision, "SLF+WDS+O," achieves a good balance between quality and speed.

Next, we examine the effect of the new ESCOT context tables in terms of PSNR and run time. Table 10 shows the PSNR of the image coding schemes with the original and the new ZC context tables when the directional filters are SLF. And Table 11 shows the PSNR when the directional filters are LLF. The image coding schemes with the new ZC context tables ("SLF/LLF+WDS+P") have a slightly better PSNR performance than those with the original ZC context table ("SLF/LLF+WDS+O") in all cases. Moreover, Table 12 indicates that "SLF/LLF+WDS+P" consumes less computation time than its "SLF/LLF+WDS+O" counterpart in all cases. The context table of "O" considers 26 neighbors in a $3 \times 3 \times 3$ cubic but that of "P" considers only 8 neighbors in a 3×3 square. Clearly, "P" uses fewer neighbors and consumes less computation. Thus, our proposed context tables can also speed up slightly the coding process.

At last, we compare the performance of the entire image coding scheme for three candidates: "LLF+NDS1+O" (the original WBCT image coding scheme), "NDF+NDS2+O" (the 2-D DWT image coding scheme) and "SLF+WDS+P" (our proposed WBCT image coding scheme). Fig. 20 shows the PSNR of these three coding schemes. Generally, our proposed "SLF+WDS+P" has better average PSNR than the "NDF+NDS2+O" and its average PSNR is comparable with that of "LLF+NDS1+O". Table 13 shows their run time. Our proposed scheme "SLF+WDS+P" saves more than 92% computing time than "LLF+NDS1+O" (the original WBCT image coding scheme). On the other hand, it costs 67% extra computing time than "NDF+NDS2+O" (the 2-D DWT image coding scheme). Clearly, our proposed scheme offers a good balance between computational complexity and image visual quality.

7. Conclusions

The WBCT-based image coding approach is explored in this paper. We propose three components to enhance its performance. First, we design a short-length filter set (SLF) to speed up the filtering process. It provides similar coding performance but requires only 10% of computational complexity of the original long-length filters (LLF). Second, we construct a mean-shift-based decision process to decide if a higher subband (HH^1, HL^1, or LH^1) is appropriate for directional decomposition. Threshold values are carefully selected to identify the energy peaks in each candidate subband. Finally, we design new zero-coding (ZC) context tables for ESCOT because the coefficients produced by directional decomposition have different statistical characteristics among near-by coefficients. Compared with the conventional 2-D DWT coding scheme, our scheme provides better visual quality with a moderate additional computational cost. Compared with the original WBCT coding scheme, the proposed coding scheme provides comparable image quality (PSNR and MSSIM) but with significantly less computing time.

Table 13
Average run time of the 2-D DWT scheme (NDF+NDS2+O), the original WBCT scheme (LLF+NDS1+O), and the proposed scheme with three new tools (SLF+WDS+P).

Scheme	SLF+WDS+P	LLF+NDS1+O	NDF+NDS2+O
Run time	4.499 s	62.469 s	2.688 s

Acknowledgements

The first author would like to thank Dr. Jang-Jer Tsai for his valuable suggestions on improving the quality of this paper. This work was supported in parts by the NSC, Taiwan under Grant 98-2221-E-009-076, by the MOEA, Taiwan under Grant 99-EC-17-A-01-I1-0016, and by the Intelligent Information Communications Research Center, NCTU.

References

[1] D. Taubman, M.W. Marcellin, JPEG2000: Image Compression Fundamentals, Standards, and Practice, Kluwer, Norwell, MA, 2002.
[2] M. Vetterli, "Wavelets, approximation and compression", IEEE Signal Proc. Mag., pp. 59–73, Sep. 2001.
[3] M. N. Do, "Directional Multiresolution Image Representations", Ph.D. Dissertation, Swiss Fed. Inst. Technol., Lausanne, Switzerland, Nov. 2001.
[4] D. Taubman, A. Zakhor, Orientation adaptive subband coding of images, IEEE Trans. Image Process. 3 (4) (1994) 421–437.
[5] D. Wang, L. Zhang, A. Vincent, F. Speranza, Curved wavelet transform for image coding, IEEE Trans. Image Process. 15 (8) (2006) 2413–2421.
[6] V. Chappelier, C. Guillemot, Oriented wavelet transform for image compression and denoising, IEEE Trans. Image Process. 15 (10) (2006) 2892–2903.
[7] W. Ding, F. Wu, X. Wu, S. Li, H. Li, Adaptive directional lifting-based wavelet transform for image coding, IEEE Trans. Image Process. 16 (2) (2007) 416–427.
[8] C.-L. Chang, B. Girod, Direction-adaptive discrete wavelet transform for image compression, IEEE Trans. Image Process. 16 (5) (2007) 1289–1302.
[9] D. Taubman, High performance scalable image compression with EBCOT, IEEE Trans. Image Process. 9 (7) (2000) 1158–1170.
[10] C. Tian , S.S. Hemami, "An embedded image coding system based on tarp filter with classification", in: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Process, Montreal, QC, Canada, May 2004, vol. 3, pp. 49–52.
[11] M.N. Do, M. Vetterli, The contourlet transform: an efficient directional decomposition multiresolution image representation, IEEE Trans. Image Process. 14 (12) (2005) 2091–2106.
[12] Y. Lu, M.N. Do, "The finer directional wavelet transform," in: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Philadelphia, PA, USA, March 2005.
[13] T.T. Nguyen, S. Oraintara, A class of multiresolution directional filter bank, IEEE Trans. Signal Process. 55 (3) (2007) 949–961.
[14] T.T. Nguyen, S. Oraintara, Multiresolution directional filterbanks: Theory, design, and applications, IEEE Trans. Signal Process 53 (10) (2005) 3895–3905.
[15] I.W. Selesnick, R.G. Baraniuk, N.G. Kingsbury, The dual-tree complex wavelet transform, IEEE Signal Process. Mag. 22 (2005) 123–151.
[16] R. Eslami, H. Radha, A new family of nonredundant transforms using hybrid wavelets and directional filter banks, IEEE Trans. Image Process. 16 (4) (2007) 1152–1167.
[17] R. Eslami, H. Radha, "Wavelet-based contourlet coding using an SPIHT-like algorithm," in: Proceedings of Conference Information Sciences, Systems, Princeton, NJ, USA, March 2004, pp. 784–788.
[18] A. Said, W. Pearlman, A new, fast and efficient image codec based on set partitioning in hierarchical trees, IEEE Trans. Circuits Syst. Video Technol. 6 (1996) 243–250.
[19] J.M. Shapiro, "An embedded hierarchical image coder using zerotrees of wavelet coefficients", in: IEEE Data Compression Conference, Snowbird, UT, 1993, pp. 214–223.
[20] J. Yang, Y. Wang, W. Xu, Q. Dai, Image coding using dual-tree discrete wavelet transform, IEEE Trans. Image Process. 17 (9) (2008) 1555–1569.
[21] P.J. Burt, E.H. Adelson, The Laplacian pyramid as a compact image code, IEEE Trans. Commun. 31 (4) (1983) 532–540.
[22] R.H. Bamberger, M.J.T. Smith, A filter bank for the directional decomposition of images: Theory and design, IEEE Trans. Signal Process. 40 (4) (1992) 882–893.
[23] S.-M. Phoong, C.W. Kim, P.P. Vaidyanathan, R. Ansari, A new class of two-channel biorthogonal filter banks and wavelet bases, IEEE Trans. Signal Process. 43 (3) (1995) 649–665.
[24] Y.P. Lin, P.P. Vaidyanathan, Theory and design of two-dimensional filter banks: a review, Multidimensional Syst. Signal Proc. 7 (1996) 263–330.
[25] A.V. Oppenheim, R.W. Schaffer, Discrete-Time Signal Processing, Englewood Cliffs, NJ, Prentice-Hall, 1989.
[26] J. Xu, Z. Xiong, S. Li, Y. Zhang, Three-dimensional embedded subband coding with optimized truncation (3D ESCOT), Appl. Comput. Harmonic Anal. 10 (2001) 290–315. Special issue on wavelet applications in engineering.
[27] X. Wu, "High-order context modeling and embedded conditional entropy coding of wavelet coefficients for image compression", in: Proceedings of 31st Asilomar Conference Signals, Systems, Computers, November 1997, pp. 1378–1382.
[28] S.-T. Hsiang, J.W. Woods, "Embedded image coding using zeroblocks of subband/wavelet coefficients and context modeling", IEEE Int. Conf. Circuits Systems (ISCAS) 3 (2000) 662–665.
[29] Z. Liu, L. Karam, "Mutual information-based analysis of JPEG2000 contexts", IEEE Trans. Image Process. 14 (4) (2005) 411–422.
[30] C.-H. Hung, H.-M. Hang, "Image coding using short wavelet-based contourlet transform", in: IEEE International Conference on Image Proceedings (ICIP), San Diego, CA, USA, October 2008.
[31] C.-H. Hung, H.-M. Hang, "Decision-directed adaptive wavelet image coding with directional decomposition," in: IEEE International Symposium on Circuits and Systems (ISCAS), Taipei, Taiwan, May 2009.
[32] Y. Cheng, Mean shift, mode seeking, and clustering, IEEE Trans. Pattern Anal. Mach. Intell. 17 (8) (1995) 790–799.
[33] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, IEEE Trans. Pattern Anal. Mach. Intell. 24 (5) (2002) 603–619.
[34] D. Comaniciu, P. Meer, "Mean shift analysis and applications", in: Proceedings of IEEE International Conference Computer Vision, 1999, pp. 1197–1203.
[35] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error measurement to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.