Directional Discrete Cosine Transforms: A Theoretical Analysis

Jingjing Fu and Bing Zeng

Department of Electrical and Electronic Engineering
The Hong Kong University of Science and Technology
Clearwater Bay, Kowloon, Hong Kong, emails: {jjfu, eezeng@ust.hk}

Abstract - *Nearly all block-based transform techniques developed so far for image and video coding applications choose the 2-D discrete cosine transform (DCT) of a square block shape. With almost no exception, this conventional DCT is always implemented separately through two 1-D transforms, along the vertical and horizontal directions, respectively. In one of our recent works, we have developed a directional DCT framework in which the first transform may choose to follow a direction other than the vertical or horizontal one, while the second transform is arranged to be a horizontal one. Compared to the conventional DCT, our directional DCT framework has been demonstrated to provide a better coding performance for image blocks that contain directional edges – a popular scenario in many image and video signals. In this paper, we attempt to pursue an in-depth theoretical analysis to understand how the coding gain is produced in the directional DCT framework and how big it can be.*

Index Terms — Image and video coding, Directional transforms, Coding efficiency

1. INTRODUCTION

It is known that each digital image or video frame contains a lot of directional edges/details, and such edge/detail orientation varies greatly from one image block to another. When the conventional 2-D DCT is applied to an image block, the first DCT will be performed along the vertical or horizontal direction. If the edge/detail orientation within this image block does not follow the vertical or horizontal direction, many non-zero AC components are produced after the transform, thus making the coding rather inefficient. By recognizing this defect, we have recently developed a directional DCT framework in which the first transform can choose to follow the dominating direction (which can be different from the vertical or horizontal direction) within each image block, while the second transform can be implemented according to the arrangement of coefficients after the first transform. This directional framework has been demonstrated to provide a remarkable coding gain for all image blocks that contain directional edges other than the vertical/horizontal ones. However, there are a number of important issues that need to be solved before it can be applied to practical image and video coding applications. One of these issues is the theoretical analysis so as to understand why a coding gain is obtained and how big it can be. In this paper, we will pursue such a study.

The rest part of this paper is organized as follows. In Section 2, we present a brief review of the directional DCT framework. In Section 3, we model the 2-D co-variance matrix using an elliptical function rotated by an angle (representing the dominating direction). Then, an in-depth theoretical analysis is presented in Section 4 where both coding gain and energy packing efficiency (EPE) are examined. Some solid supports can be obtained from the analysis results to present a clear explanation on the coding efficiency improvement achieved in the directional DCT framework. Finally, Section 5 presents some concluding remarks and future works.

2. THE DIRECTIONAL DCT FRAMEWORK: A REVIEW

In this directional DCT framework [1,2], we have selected 8 directional modes, namely, vertical (Mode-0), horizontal (Mode-1), diagonal down-left (Mode-3), diagonal down-right (Mode-4), vertical-right (Mode-5), horizontal-down (Mode-6), vertical-left (Mode-7), and horizontal-up (Mode-8). In fact, these modes are selected in the same way as those used in H.264 [3] to define the intra-prediction modes. Clearly, Mode-0 and Mode-1 are the same as the conventional 2-D DCT. In the meantime, it is easy to find that only the diagonal down-left mode and the vertical-right mode are the essential modes, whereas other modes can be derived by a flipping/transposing operation.

A. Directional DCT for the diagonal down-left mode

As shown in Fig. 1, the DCT's in the first step will be performed along the diagonal down-left direction, i.e., for each diagonal line with $i+j=k$, $k=0,\cdots,2N-2$. There are totally $2N-1$ diagonal down-left (Mode-3) DCT's to be done, whose lengths are $[N_k]=[1,2,\cdots,N-1,N,N-1,\cdots,2,1]$. All coefficients after these DCT's are expressed into a group of column vectors, see Fig. 2 for the case $N=8$.

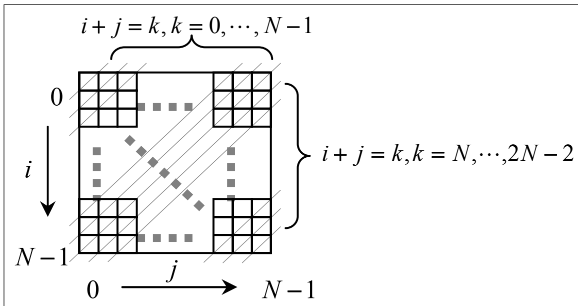

Fig. 1. An $N\times N$ image block in which the first 1-D DCT will be performed along the diagonal down-left direction.

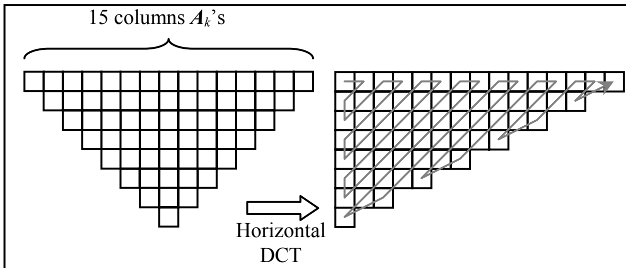

Fig. 2. Example of $N = 8$: arrangement of coefficients after the first DCT (left) and arrangement of coefficients after the second DCT as well as the modified zig-zag scanning (right).

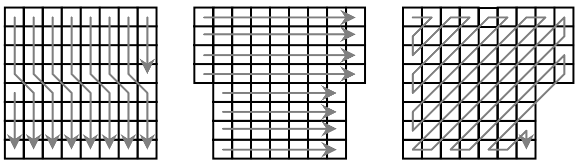

Fig. 3. Directional DCT for the vertical-right mode (Mode 5).

Notice that each column is of a different length, with the DC component placed at top, followed by the first AC component and so on. To complement the arrangement of coefficients after the diagonal DCT in the first step, the second DCT is arranged to be a horizontal one and thus applied to each row, see Fig. 2. The right part of Fig. 2 shows a modified zig-zag scanning that will be used to convert the 2-D coefficient block into a 1-D sequence so as to facilitate the runlength-based variable length coding (VLC).

B. Extension to other directional modes

Extension to other modes is straightforward. For instance, for the diagonal down-right mode (Mode-4), we can simply flip it (either vertically or horizontally), and then the flipped image block will fall into the case as discussed above.

As the second example, let us consider the vertical-right mode (Mode-5) where the block size is selected at 8×8. Referring to Fig. 3, the directional DCT in the first step follows the vertical-right direction and will generate a coefficient block as shown in the middle part of Fig. 3. Next, the second DCT will be a horizontal one again (of length 9 for the first 4 rows and 7 for the last 4 rows, respectively) and all coefficients after this DCT will be pushed to left. Finally, a modified zig-zag scanning is defined as in the right part of Fig. 3.

For Modes-6 to 8, we can simply flip or transpose the image block first and then the manipulated block will fall into the case of Mode-5.

C. DC separation and ΔDC correction

As both directional DCT's formulated above are of a different length in different directional line, they cannot be applied directly on image blocks, because they would suffer from the so-called *mean weighting defect* [4].

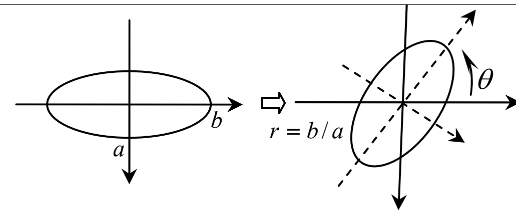

Fig. 4. An elliptical function and its rotated version used to model the co-variance matrix.

One way to solve this problem is to modify the weighting factor used in the DCT matrix. However, the DCT matrix after such modification will become a non-unitary one, whereas the transform coding theory suggests that the use of a non-unitary transform is highly disadvantageous for coding efficiency, because it would suffer from the so-called *noise weighting defect* [4].

To solve this dilemma, Kauff and Schuur proposed a novel method that consists of two steps: (1) DC separation (the mean value is subtracted from each original image block and the DC component produced after the transform, although maybe non-zero, is discarded) and (2) ΔDC correction (a correction term is calculated at the decoding stage after the first IDCT), in their work on SA-DCT [4]. This method can be readily applied in our case. It has been demonstrated in [4] that the ΔDC method is consistently better than modifying the weighting factor by 1-2 dB, and such result has been confirmed in our experiments [1,2]. Therefore, the ΔDC method is always adopted in the following theoretical analysis of this directional DCT framework.

3. MODELING OF 2-D CO-VARIANCE MATRIX

In the 1-D case, the co-variance among a number of consecutive pixels is usually modeled as a Toplitz matrix. In the 2-D case, the modeling will be much more complicated. For example, if no directional edges are assumed and the vertical and horizontal directions are treated equally, a circular function can be used to model the 2-D co-variance matrix. However, when a particular directional orientation is assumed within an image block, an elliptical function rotated by an angle θ turns to be a good model, as shown in Fig. 4. Analytically, the element of the co-variance matrix can be calculated as follows:

$$E\{x(m,n)x(p,q)\} = \rho(m-p, n-q) = \rho_0^{\sqrt{\eta^2 d_1^2(\theta) + d_2^2(\theta)}} \quad (1)$$

where $0 < \rho_0 < 1$, (m,n) and (p,q) represent the coordinates of two image pixels,

$$d_1(\theta) = (m-p)\cos\theta + (n-q)\sin\theta$$

$$d_2(\theta) = (n-q)\cos\theta - (m-p)\sin\theta \,,$$

and $\eta = b/a \geq 1$ represents the eccentricity of the elliptical function (shown in Fig. 4).

4. THEORETICAL ANALYSIS RESULTS

Let us denote an image block (with the mean value subtracted from it) of size $N \times N$ as $\boldsymbol{x} = [x(i,j)]_{N \times N}$ and use

$\boldsymbol{C} = [c_N(i,j)]_{N\times N}$ to represent the transform matrix of the N-point DCT. After performing the conventional 2-D DCT, the transform coefficients can be expressed as

$$\boldsymbol{X} = \left[X(k,l)\right]_{N\times N} = \boldsymbol{C}\cdot\boldsymbol{x}\cdot\boldsymbol{C}^T \tag{2}$$

Then, the variance (i.e., energy) of each DCT coefficient can be calculated as

$$\sigma_{k,l}^2 = E\{X(k,l)\cdot X(k,l)\} = \sum_{m,n,p,q=0}^{N-1} E\{x(m,n)x(p,q)\}\times \left(c_N(k,m)c_N(n,l)c_N(k,p)c_N(q,l)\right) \tag{3}$$

Next, let us apply the Mode-3 DCT on each diagonal down-left line and arrange the transform coefficients of each line into a column vector (as explained earlier in Fig. 2):

$$\boldsymbol{A}_k = \left[A_{0,k}, A_{1,k}, \cdots, A_{N_k-1,k}\right]^T, \quad k = 0,\cdots,2N-2 \tag{4}$$

where

$$A_{l,k} = \sum_{t=0}^{N_k-1} c_{N_k}(l,t)\cdot x_D(t) \tag{5}$$

and $\left[x_D(t)\middle| t=0,\cdots,N_k-1\right]^T$ is obtained by arranging the k-th diagonal down-left line into a column vector.

Next, a horizontal DCT is applied along each horizontal line: $[A_{l,k}]_{k=l,\cdots,2N-2-l}$ for $l=0,\cdots,N-1$. The coefficients after the second-step DCT's are pushed horizontally to left and denoted as $[\hat{A}_{l,k}]_{k=0,\cdots,2N-2-2l}$ for $l=0,\cdots,N-1$. Finally, we can calculate the variance of each $\hat{A}_{l,k}$ as

$$\hat{\sigma}_{l,k}^2 = E\left\{\hat{A}(l,k)\cdot\hat{A}(l,k)\right\} \tag{6}$$

with $k=0,\cdots,2N-2-2l$ for $l=0,\cdots,N-1$.

For the diagonal down-right mode and other directional modes, the corresponding variances of all transform coefficients can be calculated in a similar way.

A. *Coding gain*

Once the variances of all individual transform coefficients are calculated, the coding gain can be calculated as follows:

$$G_{0/1} = -\frac{1}{2(N^2-1)}\left(\sum_{k,l=0,l+k\neq 0}^{N-1}\log_2\left(\sigma_{k,l}^2\right)\right) \tag{7}$$

for the conventional 2-D DCT (Mode-0/1), and

$$G_3 = -\frac{1}{2(N^2-1)}\sum_{l=0}^{N-1}\left(\sum_{k=0,k+l\neq 0}^{2N-2-2l}\log_2\left(\hat{\sigma}_{l,k}^2\right)\right) \tag{8}$$

for the Mode-3 2-D DCT. The coding gain for other directional 2-D DCT's can be calculated similarly using their variances.

Since the mean value has been subtracted from each original image block, there are only N^2-1 terms involved in the computation of Eq. (7). On the other hand, although the DC component in the diagonal down-left 2-D DCT may be non-zero, it was indicated earlier that this DC value is always discarded. Therefore, there are also only N^2-1 terms involved in the computation of Eq. (8). It will also be the same case for other directional DCT modes.

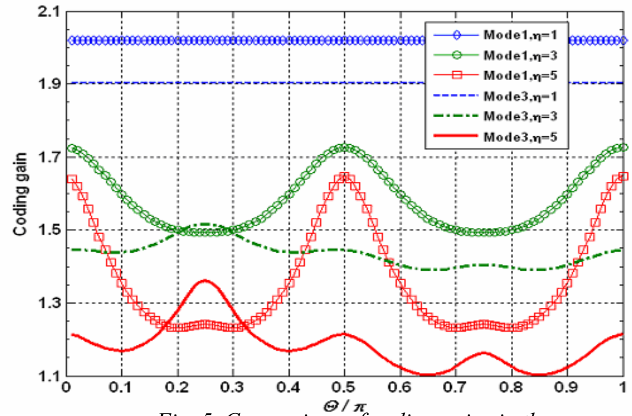

Fig. 5. Comparison of coding gains in the conventional DCT (Mode-0/1) and Mode-3 DCT.

For the first set of analysis results, we present in Fig. 5 a comparison between the conventional 2-D DCT and Mode-3 2-D DCT where the horizontal axis represents the angle θ ($N=8$). When the eccentricity $\eta = 1$ - no directional orientation, the coding gain is constant for both DCT's and the conventional DCT turns out to be always better than the Mode-3 DCT. When $\eta > 1$ - which is the case we are much more interested in, it is easy to understand that the best angle for Mode-3 is $\theta = \pi/4$. We can observe this fact exactly in Fig. 5. As a matter of fact, the coding gain of Mode-3 DCT at this angle may surpass that of the conventional DCT if η becomes large. More interestingly, Mode-3 DCT may turn out to be better than the conventional DCT over a small interval of angle around $\pi/4$.

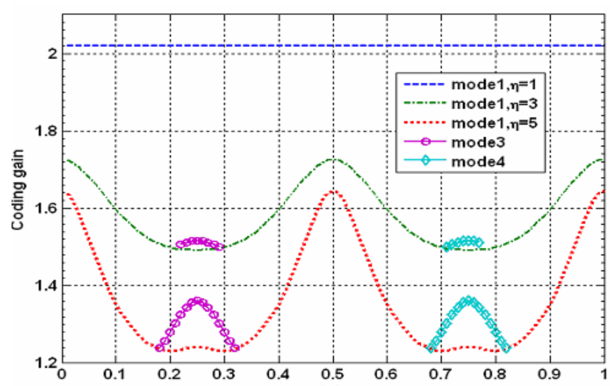

Fig.6 Comparison of all directional modes in terms of coding gain

On the other hand, it is also observed from Fig. 5 that Mode-3 DCT becomes worse than the conventional DCT when θ falls out of this small interval. This is natural because Mode-3 is used to handle the diagonal down-left direction only. When θ takes other values, other directional modes should be used. For instance, Mode-4 should be chosen for a small interval around $\theta = 3\pi/4$. Similarly, Mode-1 (horizontal) should be chosen around $\theta = 0$. Mode-0 (vertical) around $\theta = \pi/2$, etc.

Following this idea, we can choose a directional mode for each small range of θ such that the selected mode is better than all other modes. Some corresponding results are shown in Fig. 6 ($N = 8$). However, it is seen from this figure that only two diagonal modes have been selected eventually, whereas Modes-5 to 8 have been excluded. This also seems natural because Modes-5 to 8 are very close to Mode-0/1 (differing by $\pi/16$ only).

A general observation from Figs. 5 and 6 is that the coding gain improvement achieved Mode-3/4 (compared with the conventional DCT) is usually small, but increases as the eccentricity η goes larger. Moreover, with a larger η, the interval of θ where Mode-3/4 will be chosen (so as to obtain the optimal coding gain) becomes wider.

B. Energy packing efficiency (EPE)

In a practical coding application, a quantizer will always be used and it typically quantizes many transform coefficients (high frequency ones) to zero. To reflect this reality, the energy packing efficiency (EPE) proposed in [5] can be used instead, where EPE is defined as the energy portion contained in the first M_0 transform coefficients in respect to the total energy (contained in all M coefficients), i.e.,

$$EPE(M_0) = \sum_{p=0}^{M_0-1} E\{X_p^2\} \Big/ \sum_{p=0}^{M-1} E\{X_p^2\}. \qquad (9)$$

For the 8×8 block size, $M = 64$ and we choose $M_0 = 6$ to perform the EPE-based analysis in the following. Six coefficients with the significant variance are selected. For the conventional DCT, these six coefficients are the first 6 coefficients along the zig-zag scanning order, while it is distinct for directional DCT.

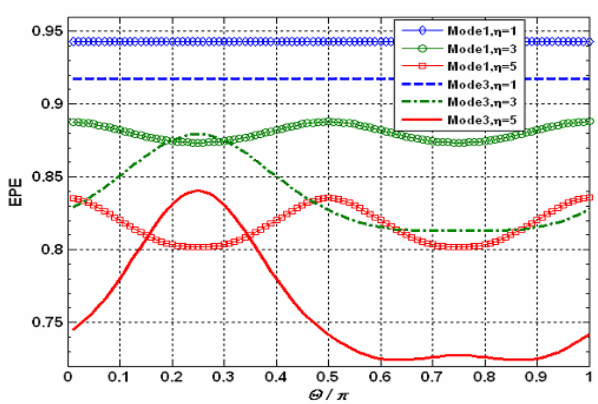

Fig. 7. Comparison of EPE's of the conventional DCT (Mode-0/1) and the Mode-3 DCT.

Similar to the coding gain based analysis, we first present in Fig. 7 a comparison between the conventional DCT and Mode-3 DCT. It is found that almost over the same interval around $\theta = \pi/4$, Mode-3 DCT yields a larger EPE than the conventional DCT.

A detailed comparison among all directional modes is shown in Fig. 8. It is seen that when η is relatively small, a similar observation (as discussed earlier) holds, i.e., only Modes-3/4 have been chosen. However, when η increases, e g., $\eta = 5$, all Modes-3 to 8 have been chosen and together they construct an EPE curve that is much better than that of the conventional DCT.

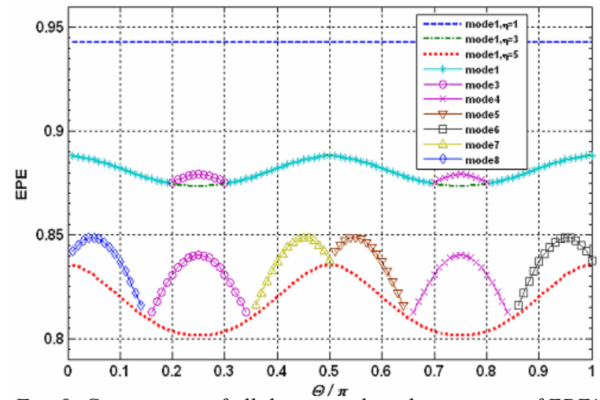

Fig. 8. Comparison of all directional modes in terms of EPE's.

5. CONCLUDING REMARKS AND FUTURE WORKS

A comprehensive theoretical analysis of the directional DCT framework has been presented in this paper where we considered both the coding gain based analysis and the EPE-based analysis. The analysis results explicitly explained where a coding efficiency improvement can be achieved in each of the directional modes defined in this new framework (compared with the conventional DCT). They also gave some numerical ranges on how big the improvements can be.

In the meantime, the analysis results also showed that the improvement decreases as the eccentricity η becomes smaller. More importantly, it was observed that the coding gain or EPE curve becomes constant when $\eta = 1$ and there always exists a (small) gap between the conventional DCT and any non-conventional mode. The reason is because any of Modes-3 to 8 has to perform transforms of different lengths to different directional lines. We anticipate that this drawback can be avoided if we combine one image block with some of its neighboring blocks – which is indeed one of our future works.

References

[1] B. Zeng and J.-J. Fu, "Directional discrete cosine transforms for image coding," in *Proc. of IEEE ICME-2006*, pp.721-724 July 2006, Toronto, Canada.

[2] B. Zeng and J.-J. Fu, "Directional discrete cosine transforms – a new framework for image coding," submitted to *IEEE Trans. on Circuits and Systems for Video Technology*, 2006.

[3] ITU-T Rec. H.264 | ISO/IEC 14496-10 (AVC), "Advanced video coding for generic audiovisual services", Mar. 2005.

[4] P. Kauff and K. Schuur, "Shape-adaptive DCT with block-based DC separation and ΔDC correction," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 8, pp. 237-242, June 1998.

[5] H. Kitajima, "Energy packing efficiency of the Hadamard transform", *IEEE Trans. on Commun*, vol. COM-24, pp. 1256-1258, Nov. 1976