

R-D Performance Upper Bound of Transform Coding for 2-D Directional Sources

Shuyuan Zhu, *Student Member, IEEE*, Siu-Kei Au Yeung, and Bing Zeng, *Member, IEEE*

Abstract—Traditionally, any 2-D transform (such as 2-D DCT) is implemented through two separable 1-D transforms along the vertical and horizontal dimensions. Such a framework is however not most suitable for a 2-D directional source in which the dominant directional information is neither horizontal nor vertical. In this letter, we attempt to determine the R-D performance upper bound for block-based transform coding schemes applied on such 2-D directional sources. It is not a surprise that the Karhunen-Loève transform (KLT) plays a critical role here. Specifically, we show that a nonseparable KLT can be determined directly from the given 2-D directional source model to yield the R-D performance upper bound. We also show that there exists a significant gap between this upper bound and the R-D performance that can be achieved by using the traditional 2-D DCT.

Index Terms—Discrete cosine transform, Karhunen-Loève transform, R-D performance, transform coding.

I. INTRODUCTION

TRANSFORM coding is an efficient compression technique to remove redundancy in image and video signals. Some well-known transforms developed in the past several decades include the Walsh-Hadamard transform (WHT), the discrete cosine transform (DCT), and the discrete sine transform (DST). The performance of these transforms is generally compared against that of the so-called Karhunen-Loève transform (KLT) which is known to be optimal in the sense of energy compaction. Among all block-based transforms with a fixed transformation matrix, DCT turns out to be the most popular one as it is proven to be the best substitute to KLT under first-order Markov conditions; these conditions are approximately met in many image and video signals [1], [2]. Because of this, DCT has been adopted in almost all transform-based image and video coding standards, such as JPEG, H.261/263, MPEG-1/2/4, and H.264/AVC.

In practical image and video coding, 2-D DCT is always implemented separately through two 1-D transforms along the vertical and horizontal dimensions, respectively. This feature significantly decreases the computational complexity. However, it is clear that such a separable 2-D transform cannot yield the best energy compaction for any 2-D directional source in which the dominant directional information is along neither the

horizontal nor vertical direction—a common scenario in many image/video blocks. In our earlier works [3], [4], a directional DCT framework has been developed to offer an improved coding performance for these directional blocks. Nevertheless, this directional 2-D DCT is still separable: the first 1-D transform is applied along the most dominating direction in each image/video block (a search is needed to determine this direction); the second 1-D transform is based on how the coefficients obtained after the first transform are placed—see [4] for details. Both theoretical analysis and experimental results have demonstrated that this directional DCT framework truly outperforms the traditional 2-D DCT for image blocks with dominating directional details along neither horizontal nor vertical direction. In the meantime, several other methods, such as the lifting-based directional DCT-like transform [5], the direction-adaptive partitioned block transform (DA-PBT) [6], and the singular value decomposition (SVD) based directional transform [7], have also been proposed to improve the coding performance for image blocks with directional information.

A highly interesting question arises here: what is the R-D performance upper bound for a 2-D directional source?¹ More specifically, what is the truly optimal transform to reach such an upper bound? Clearly, the R-D performance can be slightly improved by using KLT in each dimension (either in the traditional 2-D DCT or the directional 2-D DCT). However, can we further improve the R-D performance by considering some *nonseparable* transforms? In this letter, we present some preliminary results towards this question: we will develop the nonseparable KLT so as to reach the upper bound for a general 2-D directional source model and demonstrate that there exists a significant gap between this upper bound and the R-D performance that can be achieved by using the traditional 2-D DCT (which is separable).

II. TRADITIONAL 2-D DCT AND NONSEPARABLE 2-D KLT WITHOUT DIRECTIONAL INFORMATION

Let's use $\mathbf{x}_{2D} = [x_{i,j}]_{N \times N}$ to denote an image block of size $N \times N$ and $\mathbf{C} = [c_{i,j}]_{N \times N}$ the transform matrix of the N -point 1-D DCT. We follow the commonly-used circular model to assume that the correlation between two samples is calculated as

$$E\{x_{k,l} \cdot x_{p,q}\} = \rho \sqrt{(k-p)^2 + (l-q)^2} \quad (1)$$

where $0 < \rho < 1$, (k, l) and (p, q) represent the coordinates of two pixels.

A. Variances of Transform Coefficients

Let's use $\mathbf{X}_{2D} = [X_{m,n}]_{N \times N}$ to denote the DCT coefficients. After performing the traditional 2-D DCT, the variance of each DCT coefficient can be computed as

¹Notice that we use the PSNR value to measure the involved distortion (D); hence one refers to the PSNR-R upper bound, which actually corresponds to the Shannon R-D lower bound.

Manuscript received April 26, 2009; revised June 10, 2009. First published June 23, 2009; current version published July 29, 2009. This work was supported in part by an RGC research grant of the HKSAR government. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Adrian Munteanu.

The authors are with the ECE Department, the Hong Kong University of Science and Technology, Clearwater Bay, Kowloon, Hong Kong SAR, China (e-mail: eezshy@ust.hk; jeffay@ust.hk; eezeng@ust.hk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LSP.2009.2026115

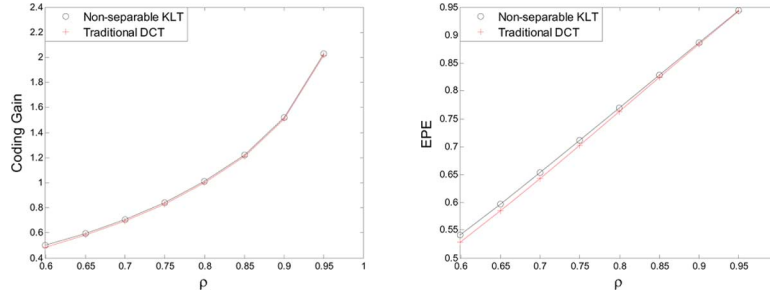


Fig. 1. Comparison of the traditional 2-D DCT and nonseparable 2-D KLT (without considering the directional information) in terms of coding gain and EPE.

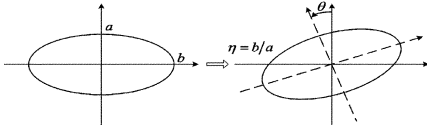


Fig. 2. Elliptical function and its rotated version for modeling the covariance matrix of a 2-D directional source.

$$\begin{aligned}\sigma_{m,n}^2 &= E\{X_{m,n} \cdot X_{m,n}\} \\ &= \sum_{k,l,p,q=0}^{N-1} E\{x_{k,l}x_{p,q}\} \cdot (c_{m,k}c_{n,l}c_{m,p}c_{n,q})\end{aligned}\quad (2)$$

where $m, n = 0, 1, \dots, N-1$.

Furthermore, an $N \times N$ image block $\mathbf{x}_{2D} = [x_{i,j}]_{N \times N}$ can be converted into a single column-vector through concatenating all columns, one after another. Then, we use (1) to compute the covariance matrix \mathbf{R}_{2D} (of size $N^2 \times N^2$) for this column-vector. By the eigen-vector decomposition, we can finally determine the KLT matrix, denoted as \mathbf{K} of size $N^2 \times N^2$. Clearly, the resulted KLT is *nonseparable*!

Notice that the computation involved in determining the KLT matrix may be quite large, as \mathbf{K} is of size $N^2 \times N^2$. Fortunately, such computations can be performed off-line and need to be done only once.

Applying this nonseparable KLT to the column-vector formed above, we know that the covariance matrix in the transform domain can be computed as follows:

$$\mathbf{R}'_{2D} = \mathbf{K} \cdot \mathbf{R}_{2D} \cdot \mathbf{K}^T. \quad (3)$$

According to KLT, all nonzero elements of \mathbf{R}'_{2D} are positioned along the diagonal line and these elements are the variances of the corresponding KLT coefficients.

B. Coding Gain

We can calculate the coding gain for any selected transform using the variances of all transform coefficients as follows:

$$G = -\frac{1}{2(N^2 - 1)} \left(\sum_{m,n=0, m+n \neq 0}^{N-1} \log_2 \sigma_{m,n}^2 \right). \quad (4)$$

Notice that we do not count the first coefficient, i.e., the DC component in (4), because of the DC prediction or other prediction methods used to reduce the bit consumption in most image and video coding schemes.

The comparison between the traditional 2-D DCT and the nonseparable 2-D KLT is shown in Fig. 1 (left). It can be seen

that the nonseparable KLT consistently outperforms the traditional 2-D DCT, but the gap between these two transforms is very insignificant.

C. Energy Packing Efficiency

To evaluate the coding efficiency in the case when some quantization is carried out on transform coefficients, we can measure the energy packing efficiency (EPE) [8]. EPE is defined as the ratio of the energy contained in the first m transform coefficients to the total energy (contained in all M coefficients):

$$EPE = \frac{\sum_{k=0}^{m-1} E\{X_k^2\}}{\sum_{k=0}^{M-1} E\{X_k^2\}}. \quad (5)$$

For the 8×8 block size, $M = 64$ and we choose the first six transform coefficients with the *largest* variances, i.e., $m = 6$, in (5). The reason of choosing $m = 6$ is because that there are usually only a few nonzero quantized coefficients in each coded 8×8 block at a moderate bit-rate — $m = 6$ represents this scenario approximately. This choice of m will also be used in the experimental results presented in the next section.

The comparison results in terms of EPE between the traditional 2-D DCT and the nonseparable 2-D KLT are shown in Fig. 1 (right). It can be seen that the nonseparable KLT also outperforms the traditional DCT method consistently, but this margin still remains to be very insignificant.

The above comparison results of coding gain and EPE clearly demonstrate that DCT is truly a good approximation of the KLT when the directional information is not considered.

III. NONSEPARABLE 2-D KLT WITH CONSIDERATION OF DIRECTIONAL INFORMATION

When a particular directional orientation is assumed within an image block, we choose a rotated elliptical function to model the correlation between two pixels within the image block:

$$E\{x_{k,l} \cdot x_{p,q}\} = \rho \sqrt{\eta^2 d_1^2(\theta) + d_2^2(\theta)} \quad (6)$$

where $\eta = b/a \geq 1$ represents the ratio of long and short radius of the elliptical function and

$$\begin{cases} d_1(\theta) = (k-p) \cos \theta + (l-q) \sin \theta \\ d_2(\theta) = (l-q) \cos \theta - (k-p) \sin \theta. \end{cases} \quad (7)$$

Compared with (1) that assumes a circular function, (6) selects an elliptical function rotated by an angle θ , as shown in Fig. 2. It is clear that two models become identical if $\eta = 1$.

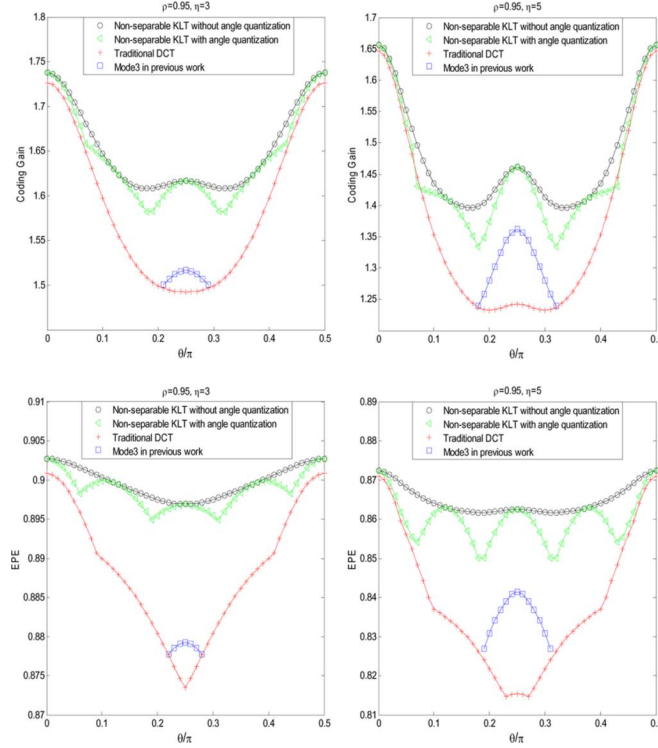


Fig. 3. Comparison of the traditional 2-D DCT, directional 2-D DCT, and nonseparable 2-D KLT in terms of coding gain and EPE.

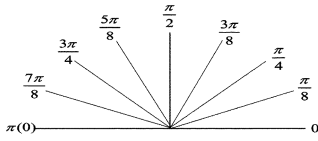


Fig. 4. Eight angles used to approximate the original orientations.

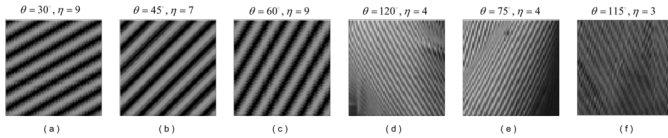


Fig. 5. Test images used in our experimental simulations: (a)-(c) three 64×64 images of sine waves with fixed rotation angles, and (d)-(f) three 64×64 image portions chosen from “Barbara”.

Given θ and η , the variances of transform coefficients can be obtained from (2) and (6) for the traditional 2-D DCT. Based on the covariance matrix R_{2D} that is now computed by using (6), we can derive the corresponding KLT matrix and thereafter compute the variances of KLT coefficients from (3). Finally, with all variances computed, we can measure the coding gain and EPE for the two transform methods. Clearly, both coding gain and EPE are symmetric around $\theta = \pi/2$. Therefore, we only need to calculate both measurements with θ from 0 to $\pi/2$. Some comparison results between these two methods are presented in Fig. 3. It is clear that, when $\eta > 1$, the KLT-based method yields a significant gap (in terms of both coding gain and EPE) over the DCT-based method, especially when the rotation angle θ is away from 0 and 90° (corresponding to the horizontal and vertical directions). Naturally, the largest gap of both coding-gain and EPE measurements between these two methods is found to occur at $\theta = \pi/4$.

In practical coding applications, one needs some overhead bits to signal the angle parameter θ for each block. To keep these overhead bits at minimum, we can only accommodate a small number of directions. Here, we select 8 angles according to Fig. 4 and others need to be quantized to the nearest one. The coding gain and EPE for the nonseparable 2-D KLT after such an angle quantization are also shown in Fig. 3. It is seen that the nonseparable 2-D KLT with angle quantization still performs significantly better than the traditional 2-D DCT.

Furthermore, the corresponding results obtained with the directional 2-D DCT [3], [4] are included in Fig. 3 for comparison purpose. It is clear that there still exists a quite noticeable gap between the directional 2-D DCT and the nonseparable 2-D KLT. Notice that only the diagonal mode (Mode-3/4) in the directional DCT yields a better performance than the traditional DCT.

Another important observation worthy of being mentioned is as follows: the first m coefficients with the largest variances may not always be the first m ones along the (zig-zag) scanning order in both traditional 2-D DCT and directional 2-D DCT, whereas this is always true in the nonseparable 2-D KLT. The implication is that extra bits are thus needed to indicate those coefficients (with the largest variances), thus further decreasing the coding efficiency.

Finally, notice that, strictly speaking, the truly optimal KLT depends on the involved signal, which thus allows the maximal energy compaction but also requires many overhead-bits for coding the KLT matrix. Alternately, people usually assume some correlation models, such as the ones given in (1) and (6), to obtain nonadaptive KLT basis functions.

IV. EXPERIMENTAL RESULTS

In this section, we provide some simulation results with estimated ρ , θ , and η to verify our theoretical analysis. We apply

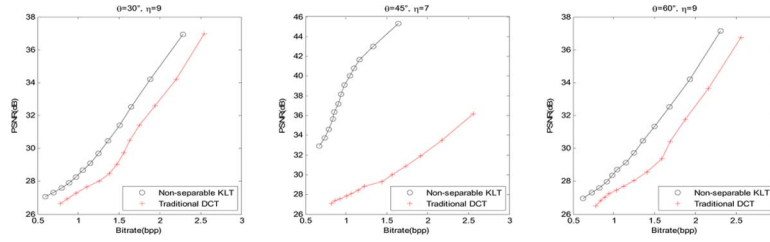


Fig. 6. R-D performances for test images of sine waves with fixed rotation angles.

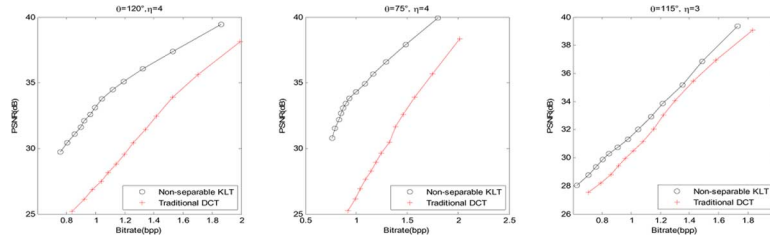


Fig. 7. R-D performances for the image portions chosen from "Barbara."

the traditional 2-D DCT and the nonseparable 2-D KLT to two groups of test images. The first one includes synthetic images of sine waves, with size 64×64 and rotated with some fixed angles, as shown in Fig. 5(a)–(c). The second group represents some 64×64 portions of the image "Barbara," as shown in Fig. 5(d)–(f).

For the traditional DCT-based method, the simulation can be directly implemented based on the JPEG standard. However, for the nonseparable 2-D KLT method, since the transform is carried out simultaneously on two dimensions, i.e., on the concatenated column-vector, the energy distribution of the transform coefficients does not exactly follow the same ordering as used in the traditional 2-D DCT. Subsequently, we cannot follow the zig-zag scanning order in this case to do the run-length based variable length coding (VLC). In fact, there is a single column of transform coefficients in the nonseparable KLT case. Thus, we scan them along this column, from the beginning to the end. In the meantime, the quantization table of size 8×8 used in JPEG also needs to be modified in order to accommodate this single-column case. To this end, we need to zig-zag scan the 8×8 quantization matrix so as to convert it into a column-vector. Then, this column-vector will be used to perform the quantization on all KLT coefficients. It is worthy to be pointed out that the quantization table used in JPEG has been somehow optimized; whereas the corresponding table used in the KLT-based method has not been optimized in any sense. Even though, the results presented below demonstrate that the KLT-based method still offers a much better R-D performance in comparison to the DCT-based method.

In our simulations, we assume $\rho = 0.95$ for all images, while the parameters θ and η are selected by visually observing the given image data. In fact, one of our future works is how to select those parameters in a more proper way. We generate the corresponding nonseparable 2-D KLT matrix for the directional 2-D source model with given parameters and then apply it to the image. We did not perform the angle quantization on the first group of images, but did on the second group (according to the eight angles shown in Fig. 4). The R-D performances obtained from the traditional 2-D DCT-based and nonseparable 2-D KLT-based methods are shown in Figs. 6 and 7. It can be seen that the nonseparable 2-D KLT-based method provides

a remarkable improvement over the traditional DCT-based method for all test images.

V. CONCLUDING REMARKS

In this letter, we presented a theoretical study of the R-D performance of transform coding for a general 2-D directional source model. To this goal, we developed the nonseparable KLT so as to achieve the upper-bound for the R-D performance. We showed and then confirmed via experiments that there exists a significant gap between this upper-bound and the R-D performance that can be achieved by using the traditional 2-D DCT. Although the nonseparable 2-D KLT depends on the statistics of the image data and thus is not effective to be used in practical coding applications, the remarkable improvement demonstrated for directional 2-D sources indeed brings us with a broad imagination space as well as a strong attention: it could be highly possible to find some nonseparable transforms with *fixed* matrices that can approximate well the corresponding nonseparable KLTs for a given range of parameters ρ , θ , and η . This will be the central issue in our future works, and our approach is basically to try some functional fitting on the determined nonseparable KLT matrix for the given ρ , θ , and η .

REFERENCES

- [1] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Trans. Comput.*, vol. C-23, pp. 90–93, Jan. 1974.
- [2] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
- [3] J. Fu and B. Zeng, "Directional discrete cosine transforms: A theoretical analysis," in *ICASSP-2007*, Honolulu, HI.
- [4] B. Zeng and J. Fu, "Directional discrete cosine transforms—a new framework for image coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, pp. 305–313, Mar. 2008.
- [5] H. Xu, J. Xu, and F. Wu, "Lifting-based directional DCT-like transform for image coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, pp. 1325–1335, Oct. 2007.
- [6] C. Chang and B. Girod, "Direction-adaptive partitioned block transform for image coding," in *ICIP-2008*, San Diego, CA.
- [7] Y. Ye and M. Karczewicz, "Improved H.264 intra coding based on bi-directional intra prediction, directional transform, and adaptive coefficient scanning," in *ICIP-2008*, San Diego, CA.
- [8] H. Kitajima, "Energy packing efficiency of the Hadamard transform," *IEEE Trans. Commun.*, vol. COM-24, pp. 1256–1258, Nov. 1976.