Jointly Optimized Spatial Prediction and Block Transform for Video and Image Coding

Jingning Han, *Student Member, IEEE*, Ankur Saxena, *Member, IEEE*, Vinay Melkote, *Member, IEEE*, and Kenneth Rose, *Fellow, IEEE*

***Abstract*—This paper proposes a novel approach to jointly optimize spatial prediction and the choice of the subsequent transform in video and image compression. Under the assumption of a separable first-order Gauss-Markov model for the image signal, it is shown that the optimal Karhunen-Loeve Transform, given available partial boundary information, is well approximated by a close relative of the discrete sine transform (DST), with basis vectors that tend to vanish at the known boundary and maximize energy at the unknown boundary. The overall intraframe coding scheme thus switches between this variant of the DST named asymmetric DST (ADST), and traditional discrete cosine transform (DCT), depending on prediction direction and boundary information. The ADST is first compared with DCT in terms of coding gain under ideal model conditions and is demonstrated to provide significantly improved compression efficiency. The proposed adaptive prediction and transform scheme is then implemented within the H.264/AVC intra-mode framework and is experimentally shown to significantly outperform the standard intra coding mode. As an added benefit, it achieves substantial reduction in blocking artifacts due to the fact that the transform now adapts to the statistics of block edges. An integer version of this ADST is also proposed.**

***Index Terms*—Blocking artifact, discrete sine transform (DST), intra-mode, spatial prediction, spatial transform.**

I. Introduction

TRANSFORM coding is widely adopted in image and video compression to reduce the inherent spatial redundancy between adjacent pixels. The Karhunen–Loeve transform (KLT) possesses several optimality properties, e.g., in terms of high-resolution quantization (of Gaussians) and full decorrelation of the transformed samples. Practical considerations, however, limit the use of KLT. Its dependence on the signal results in high implementation complexity and added side information in the bitstream, as well as the absence of fast computation algorithm in general. The discrete cosine transform (DCT) has long been a popular substitute due to properties such as good energy compaction [1]. Standard video codecs such as H.264/AVC [2] implement transform coding within a block coder framework. Each video frame is partitioned into a grid of blocks, which may be spatially (intra-mode) or temporally (inter-mode) predicted, and then transformed via the DCT. The transform coefficients are quantized and entropy coded. Typical block sizes vary between 4×4 and 16×16. Such a block coder framework is motivated by the need to adapt to local signal characteristics, coding flexibility, and computational concerns. This paper focuses, in particular, on the intra-mode in video coding. Note that intra-mode coding does not exploit temporal redundancies, and thus, the concepts developed herein are generally applicable to still-image compression.

Although the motivation for employing a block coder is to separate the video frame into distinct regions, each of which with its own locally stationary signal statistics, invariably, the finite number of choices for block sizes and shapes results in residual correlation between adjacent blocks. In order to achieve maximum compression efficiency, intra-mode coding exploits the local anisotropy (for instance, the occurrence of spatial patterns within a frame) via the spatial prediction of each block from previously encoded neighboring pixels, available at block boundaries. The DCT has been demonstrated to be a good approximation for the KLT under certain Markovian assumptions [1], when there is no spatial prediction from pixels of adjacent blocks. However, its efficacy after boundary information has been accounted for is questionable. The statistics of the residual pixels close to known boundaries can significantly differ from the ones that are far off; the former might be better predicted from the boundary than the latter, and thus, one expects a corresponding energy variation across pixels in the residual block. The DCT is agnostic to this phenomenon. In particular, its basis functions achieve their maximum energy at both ends of the block. Hence, the DCT is mismatched with the statistics of the residual obtained after spatial prediction. This, of course, motivates the question of what practical transform is optimal or nearly optimal for the residual pixels after spatial prediction.

This paper addresses this issue by considering a joint optimization of the spatial prediction and the subsequent transformation. Under the assumption of a separable first-order

Manuscript received April 10, 2011; revised July 29, 2011; accepted September 04, 2011. Date of publication September 29, 2011; date of current version March 21, 2012. This work was supported in part by the University of California MICRO Program, by Applied Signal Technology Inc., by Qualcomm Inc., by Sony Ericsson Inc., and by the NSF under Grant CCF-0917230. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Hsueh-Ming Hang.

J. Han and K. Rose are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA (e-mail: jingning@ece.ucsb.edu; rose@ece.ucsb.edu).

A. Saxena was with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA. He is now with Samsung Telecommunications America, Richardson, TX 75082 USA (e-mail: ankur@ece.ucsb.edu).

V. Melkote was with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA. He is now with Dolby Laboratories, Inc., San Francisco, CA 94103 USA (e-mail: melkote@ece.ucsb.edu).

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/TIP.2011.2169976

Gauss–Markov model for the image pixels, prediction error statistics are computed based only on the available, i.e., already encoded (and reconstructed), boundaries. The mathematical analysis shows that the KLT of such intra-predicted residuals is efficiently approximated by a relative of the well-known discrete sine transform (DST) with appropriate frequencies and phase shifts. Unlike the DCT, this variant of the DST is composed of basis functions that diminish at the known block boundary, while retaining high energy at the unknown boundary. Due to such asymmetric structure of the basis functions, we refer to it as asymmetric DST (ADST). The proposed transform has significantly superior performance compared with the DCT in terms of coding gain as demonstrated by the simulations presented later in the context of ideal model scenarios. Motivated by this theory, a hybrid transform coding scheme is proposed, which allows choosing from the proposed ADST and the traditional DCT, depending on the quality and the availability of boundary information. Simulations demonstrate that the proposed hybrid transform coding scheme consistently achieves remarkable bit savings at the same peak signal-to-noise ratio (PSNR). Note that the intra-mode in H.264/AVC, which utilizes spatial prediction followed by the DCT, has been shown to provide better rate-distortion performance than wavelet-based Motion-JPEG2000 at low-to-medium frame/image resolutions [e.g., Common Intermediate Format (CIF) and Quarter CIF (QCIF)] [3]. Hence, the proposed block-based hybrid transform coding scheme is of significant benefit to still-image coding as well. A low-complexity integer version of the proposed ADST is also presented, which enables the direct deployment of the proposed hybrid transform coding scheme in conjunction with the integer DCT (Int-DCT) of the H.264/AVC standard.

A well-known shortcoming in image/video coding is the blocking effect; since the basis vectors of the DCT achieve their maximum energy at block edges, the incoherence in the quantization noise of adjacent blocks is magnified and exacerbates the notorious "blocking effect." Typically, this issue is addressed by post-filtering (deblocking) at the decoder to smooth the block boundaries, i.e., a process that can result in information loss, such as the undesirable blurring of sharp details. The proposed ADST provides the added benefit of alleviating this problem; its basis functions vanish at block edges with known boundaries, thus obviating the need for deblocking these edges. Simulation results exemplify this property of the proposed approach.

Highly relevant literature includes [4], where a first-order Gauss–Markov model was assumed for the images and it was shown that the image can be decomposed into a boundary response and a residual process given the closed boundary information. The boundary response is an interpolation of the block content from its boundary data, whereas the residual process is the interpolation error. Jain [4], [5] showed that the KLT of the residual process is exactly the DST when all boundaries are available, under the assumed Gauss–Markov model. However, in practice, blocks are sequentially coded, which implies that, when coding a particular block, available information is limited to only few (and not all) of its boundaries. Meiri and Yudilevich [6] attempted to solve this by first encoding the edges of the block, which border unknown boundaries. The remaining pixels of the block are now all enclosed within known boundaries and are encoded with a "pinned sine transform." However, the separate coding procedure required for block edges comes at a cost to coding efficiency. In the late 80s, it was experimentally observed that, under certain conditions, there exists some similarity in basis functions between the KLT of extrapolative prediction residual and a variant of the DST [7]. Alternatively, various transform combinations have been proposed in the literature. For instance, in [8], sine and cosine transforms are alternately used on image blocks to efficiently exploit inter block redundancy. Directional cosine transforms to capture the texture of block content have been proposed in [9]. More recently, mode-dependent directional transforms (MDDTs) have been proposed in [10], wherein different vertical and horizontal transforms are applied to each of the nine modes (of block size 4×4 and 8×8) in H.264/AVC intra-prediction. Unlike the proposed approach here, where a single ADST and the traditional DCT are used in combination based on the prediction context, the MDDTs are individually designed for each prediction mode based on training data of intra-prediction residuals in that mode and thus require the storage of 18 transform matrices at either dimension. Another class of transform coding schemes motivated by the need to reduce blockiness employs overlapped transforms, e.g., [11] and [12], to exploit inter block correlation. A recent related approach is [13] where intra coding is performed with the transform block enlarged to include available boundary pixels from previously encoded blocks.

The approach we develop in this paper is derived within the general setting adopted for intra coding in current standards where prediction from boundaries is followed by the transform coding of the prediction residual. We derive the optimal transform for the prediction residuals under Markovian assumptions, show that it is signal independent and has fast implementation, and demonstrate its application to practical image/video compression. Some of our preliminary results were presented in [14].

The rest of this paper is organized as follows: Section II presents a mathematical analysis for spatial prediction and residual transform coding in video/image compression. Section III describes the proposed hybrid transform coding scheme and outlines the implementation details in the H.264/AVC intra coding framework. An integer version of the proposed transform coding scheme is then provided in Section IV.

II. Joint Spatial Prediction and Residual Transform Coding

This section describes the mathematical theory behind the proposed approach in the context of prediction of a 1-D vector from one of its boundary points. The optimal transform (KLT) for coding the prediction residual vector is considered, which, under suitable approximations, yields the ADST that we referred to in Section I. We then consider the effect of prediction from a "noisy" boundary since, in practice, only the reconstructed version of the boundary is available to the decoder (and not its original or exact value). Finally, the proposed ADST is compared with the traditional DCT in terms of coding gain

under assumed model conditions. The simplified exposition presented here then motivates the hybrid coding approach for 2-D video/image blocks presented in Section III.

A. One-Dimensional Derivation

Consider a zero-mean unit-variance first-order Gauss–Markov sequence, i.e.,

$$x_k = \rho x_{k-1} + e_k \tag{1}$$

where ρ is the correlation coefficient and e_k is a white Gaussian noise process with variance $1-\rho^2$. Let $\underline{x} = [x_1, x_2, \ldots, x_N]^T$ denote the random vector to be encoded given x_0 as the available (one-sided) boundary. Superscript T denotes the matrix transposition. Recursion (1) translates into the following set of equations:

$$\begin{aligned} x_1 &= \rho x_0 + e_1 \\ x_2 - \rho x_1 &= e_2 \\ &\vdots \\ x_N - \rho x_{(N-1)} &= e_N \end{aligned} \tag{2}$$

or in compact notation

$$Q\underline{x} = \underline{b} + \underline{e} \tag{3}$$

where

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & \ldots \\ -\rho & 1 & 0 & 0 & \ldots \\ 0 & -\rho & 1 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & 0 & -\rho & 1 \end{pmatrix} \tag{4}$$

and $\underline{b} = [\rho x_0, 0, \ldots, 0]^T$ and $\underline{e} = [e_1, e_2, \ldots, e_N]^T$ capture the boundary information and the innovation process, respectively. It can be shown that Q is invertible, and thus

$$\underline{x} = Q^{-1}\underline{b} + Q^{-1}\underline{e} \tag{5}$$

where superscript -1 indicates the matrix inversion. As expected, the "boundary response" or prediction $Q^{-1}\underline{b}$ in (5) satisfies

$$Q^{-1}\underline{b} = [\rho x_0, \rho^2 x_0, \ldots, \rho^N x_0]^T. \tag{6}$$

The prediction residual, i.e.,

$$\underline{y} = Q^{-1}\underline{e} \tag{7}$$

is to be compressed and transmitted, which motivates the derivation of its KLT. The autocorrelation matrix of $\underline{y}$ is given by

$$\begin{aligned} R_{\underline{yy}} &= E\{\underline{y}\underline{y}^T\} = Q^{-1}E\{\underline{e}\underline{e}^T\}(Q^T)^{-1} \\ &= (1-\rho^2)Q^{-1}(Q^T)^{-1}. \end{aligned} \tag{8}$$

Thus, the KLT for $\underline{y}$ is a unitary matrix that diagonalizes $Q^{-1}(Q^T)^{-1}$ and, hence, also the more convenient matrix:

$$P_1 = Q^T Q = \begin{pmatrix} 1+\rho^2 & -\rho & 0 & 0 & \ldots \\ -\rho & 1+\rho^2 & -\rho & 0 & \ldots \\ 0 & -\rho & 1+\rho^2 & -\rho & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & -\rho & 1+\rho^2 & -\rho \\ 0 & \ldots & 0 & -\rho & 1 \end{pmatrix}. \tag{9}$$

Although P_1 is Toeplitz, note that the element at the bottom right corner is different from all the other elements on the principal diagonal, i.e., it is not $1+\rho^2$. This irregularity complicates an analytic derivation of the eigenvalues and eigenvectors of P_1. As a subterfuge, we approximate P_1 with

$$\hat{P}_1 = \begin{pmatrix} 1+\rho^2 & -\rho & 0 & 0 & \ldots \\ -\rho & 1+\rho^2 & -\rho & 0 & \ldots \\ 0 & -\rho & 1+\rho^2 & -\rho & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & -\rho & 1+\rho^2 & -\rho \\ 0 & \ldots & 0 & -\rho & 1+\rho^2-\rho \end{pmatrix} \tag{10}$$

which is obtained by replacing the bottom-right corner element with $1+\rho^2-\rho$. The approximation clearly holds for $\rho \to 1$, which is indeed a common approximation for image signals. Now, the unitary matrix T_S that diagonalizes $\hat{P}_1$ and, hence, an approximation for the required KLT of $\underline{y}$, has been shown, in another context, to be the following relative of the common DST [15]:

$$[T_S]_{j,i} = \left(\frac{2}{\sqrt{2N+1}} \sin \frac{(2j-1)i\pi}{2N+1} \right) \tag{11}$$

where $j, i \in \{1, 2, \ldots, N\}$ are the frequency and time indexes of the transform kernel, respectively. Needless to say, the constant matrix T_S is independent of the statistics of innovation e_k and can be used as an approximation for the KLT when full information on boundary x_0 is available.

Note that the rows of T_S (i.e., the basis functions of the transform) take smaller values in the beginning (closer to the known boundary) and larger values toward the end. For instance, consider the row with $j = 1$ (i.e., the basis function with the lowest frequency). In the case where $N \gg 1$, the first sample $(i = 1)$ is $(2/\sqrt{2N+1})\sin(\pi/2N+1) \approx 0$, whereas the last sample $(i = N)$ takes the maximum value $(2/\sqrt{2N+1})\sin(N\pi/2N+1) \approx (2/\sqrt{2N+1})$. We thus refer to matrix/transform T_S as the ADST.

B. Effect of Quantization Noise on the Optimal Transform

The prior derivation of the ADST followed from the KLT of $\underline{y}$, whose definition assumes the exact knowledge of boundary x_0. In practice, however, only a distorted version of this boundary is available (due to quantization). For instance, in the context of block-based video coding, we have access only to *reconstructed* pixels of neighboring blocks. We thus consider

here the case when the boundary is available as the distorted value, i.e.,

$$\hat{x}_0 = x_0 + \delta \tag{12}$$

where δ is a zero-mean perturbation of the actual boundary x_0. Note that this model covers as special case the instance where no boundary information is available. We now rewrite the first equation in (2) as

$$x_1 = \rho\hat{x}_0 + f_1 \tag{13}$$

where $f_1 = -\rho\delta + e_1$, and (5) as

$$\underline{x} = Q^{-1}\underline{\hat{b}} + Q^{-1}\underline{\hat{e}} \tag{14}$$

with $\underline{\hat{b}} = [\rho\hat{x}_0, 0, \ldots, 0]^T$ and $\underline{\hat{e}} = [f_1, e_2, \ldots, e_N]^T$.

We denote by $\underline{\hat{y}} = Q^{-1}\underline{\hat{e}}$ the prediction residual when the boundary is distorted. As before, we require the KLT of $\underline{\hat{y}}$, which diagonalizes the corresponding autocorrelation matrix $R_{\underline{\hat{y}}\underline{\hat{y}}}$. Note that

$$\begin{aligned} E\left\{f_1^2\right\} &= E\{(e_1 - \rho\delta)^2\} \\ &= 1 - \rho^2 + \rho^2 E\{\delta^2\} = 1 - \rho^2 + \rho^2\sigma^2 \end{aligned} \tag{15}$$

where $\sigma^2 = E\{\delta^2\}$. The aforementioned follows from the fact that δ is independent of innovations $\underline{e}$, since x_0 itself is independent of $\underline{e}$. Thus

$$R_{\underline{\hat{y}}\underline{\hat{y}}} = E\left\{Q^{-1}\underline{\hat{e}}\underline{\hat{e}}^T(Q^T)^{-1}\right\} = (1-\rho^2)P_2^{-1} \tag{16}$$

where

$$P_2 = \begin{pmatrix} \rho^2 + \frac{1-\rho^2}{1-\rho^2+\rho^2\sigma^2} & -\rho & 0 & \ldots \\ -\rho & 1+\rho^2 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & 1+\rho^2 & -\rho \\ 0 & \ldots & -\rho & 1 \end{pmatrix}. \tag{17}$$

The KLT of $\underline{\hat{y}}$ is thus a unitary matrix that diagonalizes P_2. We now consider two separate cases.

Case 1—Small distortion: Suppose that $\sigma^2 \ll 1 - \rho^2$, which is usually the case when the quantizer resolution is medium or high. The top-left corner element in P_2 thus approaches $\rho^2 + 1$. Furthermore, when $\rho \to 1$, we can reapply the earlier subterfuge to the bottom-right corner element and replace it with $1 + \rho^2 - \rho$. Then, P_2 simply becomes $\hat{P}_1$ of (10), and the required diagonalizing matrix, i.e., the transform, is once again the ADST matrix T_S.

Case 2—Large distortion: The other extreme is when no boundary information is available or the energy of the quantization noise is high. In this case, we have $\sigma^2 \gg 1 - \rho^2$. The top-left corner element of P_2 is then

$$\rho^2 + \frac{1-\rho^2}{1-\rho^2+\rho^2\sigma^2} \approx \rho^2 \tag{18}$$

and can be approximated as $1 + \rho^2 - \rho$ when $\rho \to 1$. Thus, P_2 can be approximated as

$$\hat{P}_2 = \begin{pmatrix} 1+\rho^2-\rho & -\rho & 0 & \ldots \\ -\rho & 1+\rho^2 & -\rho & \ldots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & 1+\rho^2 & -\rho \\ 0 & \ldots & -\rho & 1+\rho^2-\rho \end{pmatrix} \tag{19}$$

whose KLT can be shown to be the conventional DCT [1] (see Appendix A), which we henceforth denote by T_C. This also implies that the DCT is the optimal transform in the case where no boundary information is available, and the transform is directly applied to the pixels instead of the residuals.

C. Quantitative Analysis Comparing the ADST and the DCT to the KLT

The previous discussion argued that the ADST or the DCT closely approximate the KLT under limiting conditions on the correlation coefficient ρ or the quality of boundary information indicated by σ. We now quantitatively compare the performance of the DCT and the proposed ADST against that of the KLT (of $\underline{y}$ or $\underline{\hat{y}}$) in terms of coding gains [16] under the assumed signal model, at different values of ρ and σ.

First, consider the case when there is no boundary distortion. Let the prediction residual $\underline{y}$ be transformed to

$$\underline{z} = A\underline{y} = [z_1, z_2, \ldots, z_N]^T \tag{20}$$

with an $N \times N$ unitary matrix A. The objective of the encoder is to distribute a fixed number of bits to the different elements of $\underline{z}$ such that the average distortion is minimized. This *bit-allocation problem* is addressed by the well-known water filling algorithm (see, e.g., [16]). Under assumptions such as a Gaussian source, a high-quantizer resolution, a negligible quantizer overload, and with non-integer bit-allocation allowed, it can be shown that the minimum distortion (mean squared error) obtainable is proportional to the geometric mean of the transform domain sample variances $\sigma_{z_i}^2$, i.e.,

$$D_A \propto \left(\prod_{i=1}^{N}\sigma_{z_i}^2\right)^{1/N} \tag{21}$$

where, for Gaussian source, the proportionality coefficient is independent of transform A. These variances can be obtained as the diagonal elements of the autocorrelation matrix of $\underline{z}$, i.e.,

$$R_{\underline{z}\underline{z}} = E[\underline{z}\underline{z}^T] = AE[\underline{y}\underline{y}^T]A^T = (1-\rho^2)AP_1^{-1}A^T \tag{22}$$

where we have used (8) and (9). The coding gain in decibels of any transform A is now defined as

$$\mathcal{G}_A = 10\log_{10}(D_{\mathbf{I}}/D_A). \tag{23}$$

Here, $\mathbf{I}$ is the $N \times N$ identity matrix, and hence, $D_{\mathbf{I}}$ is the distortion resulting from the direct quantization of the untransformed vector $\underline{y}$. The coding gain $\mathcal{G}_A$ thus provides a compar-

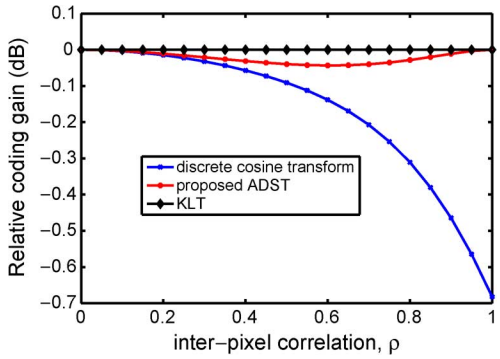


Fig. 1. Theoretical coding gains of the ADST and the DCT relative to the KLT, all applied to the prediction residuals (given the exact knowledge of the boundary), plotted versus the inter-pixel correlation coefficient. The block dimension is 4 × 4.

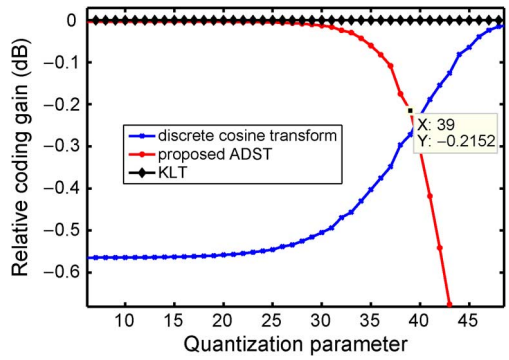


Fig. 2. Theoretical coding gains of ADST and DCT relative to the KLT plotted versus QP. The inter-pixel correlation coefficient is 0.95.

ison of the average distortion incurred with and without transformation A. Note that, for any given A (including the ADST, the DCT, and the KLT of $\underline{y}$), computing $R_{\underline{zz}}$ and, hence, $\sigma^2_{z_i}$ does not require making any approximations for P_1. However, when A is the KLT of $\underline{y}$, $R_{\underline{zz}}$ is a diagonal matrix (with diagonal elements equal to the eigenvalues of P_1^{-1}), the transform coefficients z_i are uncorrelated, and the coding gain reaches its maximum.

Fig. 1 compares the ADST and the DCT in terms of their coding gains, relative to the KLT; specifically, it depicts $\mathcal{G}_{T_s} - \mathcal{G}_{\mathrm{KLT}}$ and $\mathcal{G}_{T_c} - \mathcal{G}_{\mathrm{KLT}}$ versus the correlation coefficient ρ. Note that, although the derivation of the ADST, i.e., T_s, assumed that $\rho \to 1$, it, in fact, approximates the KLT closely even at other values of the correlation coefficient ρ, when the boundary is exactly known, i.e., without any distortion. The maximum gap between the ADST and the KLT or the maximum loss of optimality is less than 0.05 dB (and occurs at $\rho \approx 0.65$). In comparison, the DCT poorly performs (by about 0.56 dB) for the practically relevant case of high correlation ($\rho \approx 0.95$). At low correlation ($\rho \to 0$), the autocorrelation matrix of the prediction residual, i.e., $R_{\underline{yy}} \approx \mathbf{I}$, and, hence, any unitary matrix, including the ADST and the DCT, will function as a KLT. The block length used in obtaining the results of Fig. 1 was $N = 4$, but similar behavior can be observed at higher values of N. We emphasize that these theoretical coding gains are obtained with respect to the prediction residuals, which are to be transmitted instead of the original samples.

We now consider the case where the boundary is distorted, i.e., when the vector to be transformed is $\underline{\hat{y}}$. The KLT in this case is defined via matrix P_2 of (17), and similar to (22), the coding gain of transform A is obtained from the diagonal elements of $AP_2^{-1}A^T$. Note that coding gains will now be a function of the boundary distortion σ^2, which, in practice, is a result of the quantization of the transformed coefficients. In order to enhance the relevance of the discussion that follows, we first describe a mapping from σ^2 to the quantization parameter (QP) commonly used in the context of video compression to control the bitrate/reconstruction quality, so that the performance of the ADST and the DCT can be directly compared via coding gains at different *QP values*. Since the transforms considered are unitary, assuming a uniform high rate quantizer, the variance of boundary distortion can be shown to be $\Delta_Q^2/12$, where Δ_Q is the quantizer step size associated with the QP value Q. Let the image pixels (luminance components) be modeled as Gaussian with a mean of 128 and a variance of $127/(\sqrt{2}\mathrm{erf}^{-1}(0.99))$, where $\mathrm{erf}^{-1}(\cdot)$ is the inverse error function.[1] However, note that, so far, the discussion assumed the unit variance for the source samples x_i (see description of (1) in Section II-A). We therefore normalize the image pixel model to the unit variance, and hence, the "normalized" variance σ^2 of the boundary distortion δ maps to the true distortion $\Delta_Q^2/12$ as follows:

$$\sigma^2 = \frac{\Delta_Q^2}{12}\frac{\sqrt{2}\mathrm{erf}^{-1}(0.99)}{127}. \tag{24}$$

The aforementioned mapping is used to compare the performance of the ADST and the DCT relative to the KLT at different values of the boundary distortion indicated in terms of the QP. Fig. 2 provides such a comparison at the inter-pixel correlation of $\rho = 0.95$. The following observations can be made:

1) At low values of QP (i.e., reliable block boundary), as discussed in Section II-B, the ADST outperforms the DCT and performs close to the KLT.
2) At high values of QP (e.g., $QP > 40$), the pixel boundary is very distorted and unsuitable for prediction. In this case, the DCT performs better than the ADST.
3) Typically, in image and video coding, QP values of practical interest range between 20 and 40. As evident from Fig. 2, the ADST should be the transform of choice in these cases when intra prediction is employed.

III. Hybrid Transform Coding Scheme

We extended the theory proposed so far in the framework of 1-D sources to the case of 2-D sources, such as images. H.264/AVC intra coding predicts the block adaptively using its (upper and/or left) boundary pixels and performs a DCT separately on the vertical and horizontal directions, i.e., the block of pixels is assumed to conform to a 2-D *separable model*. The previous simulations with a 1-D Gauss–Markov model indicate that, for typical quantization levels, the ADST can provide better coding performance than the DCT when the pixel boundary is available along a particular direction. Therefore, we herein jointly optimize the choice of the transform in conjunction with the adaptive spatial prediction of the standard and refer to this paradigm as the hybrid transform coding scheme.

[1]This choice of parameters effectively requires that the CDF increases by 0.99, when pixel value goes from 0 to 255.

A. Hybrid Transform Coding With a 2-D Separable Model

Let $x_{i,j}$ denote the pixel in the ith row and the jth column of a video frame or image. The first-order Gauss–Markov model of (1) is extended to two dimensions via the following separable model for pixels[2]:

$$x_{i,j} = \rho x_{i-1,j} + \rho x_{i,j-1} - \rho^2 x_{i-1,j-1} + e_{i,j} \tag{25}$$

where, as in Section II-A, the source samples $x_{i,j}$ are assumed to have zero mean and unit variance. The innovations $e_{i,j}$ are independent identically distributed (i.i.d.) Gaussian random variables. Note that the aforementioned model results in the following inter pixel correlation:

$$E\{x_{i,j}x_{k,m}\} = \rho^{|i-k|+|j-m|}. \tag{26}$$

Now, consider an $N \times N$ block of pixels, i.e., X, containing pixels $x_{i,j}$, with $i,j \in 1,2,\ldots,N$. We can rewrite (25) for block X via the compact notation, i.e.,

$$QXQ^T = B + E \tag{27}$$

where Q is defined by (4) and E is the innovation matrix with elements $e_{i,j}$, $i,j \in 1,2,\ldots,N$. By expanding QXQ^T, it can be shown that matrix B contains non-zero elements only in the first row and the first column, with

$$\begin{aligned} B(1,1) &= \rho x_{0,1} + \rho x_{1,0} - \rho^2 x_{0,0} \\ B(1,j) &= \rho x_{0,j} - \rho^2 x_{0,j-1} \qquad \forall j = \{2\ldots N\} \\ B(i,1) &= \rho x_{i,0} - \rho^2 x_{i-1,0} \qquad \forall i = \{2\ldots N\}. \end{aligned} \tag{28}$$

In other words, B contains the boundary information from two sides of X (i.e., for the top and left boundaries of the block). With this mathematical framework in place, we now describe the proposed hybrid transform coding scheme to encode block X.

1) Prediction and Transform in the Vertical Direction: Let us consider the jth column in the image, denoted as $\{c_i = x_{i,j}\}$. By (25), $c_i = \rho c_{i-1} + \beta_i$, where

$$\beta_i = x_{i,j} - \rho x_{i-1,j}. \tag{29}$$

The pixels are Gaussian random variables, and hence, so are variables $\{\beta_i\}$. Furthermore

$$E\{\beta_i\beta_{i+h}\} = \rho^{2+|h|} + \rho^{4+|h|} - \rho^{3+|h-1|} - \rho^{3+|h+1|} = 0$$

for any integer $h \neq 0$. Therefore, $\{\beta_i\}$ are indeed i.i.d. Gaussian random variables. Hence, the $\{c_i\}$ sequence effectively follows a 1-D Gauss–Markov model akin to (1), with innovations given by $\{\beta_i\}$. Thus, the arguments for optimal prediction and transform for 1-D vectors developed in Section II hold for individual image columns.

Case 1—Top boundary is available: When the top boundary of X is available, (6) is employed to predict the jth column of X as $Q^{-1}[c_0, 0, \ldots, 0]^T$ (recall that $c_0 = x_{0,j}$). The ADST can now be applied on the resulting column $\underline{\beta}_j = [\beta_1, \ldots, \beta_N]^T$ of residual pixels. This process (individually) applied to the N columns results in the following matrix Y_S of transform coefficients:

$$Y_S = T_S Q^{-1}[\underline{\beta}_1, \underline{\beta}_2, \ldots, \underline{\beta}_N] = [\underline{t}_1, \ldots, \underline{t}_N]^T[\underline{\beta}_1, \ldots, \underline{\beta}_N]$$

where $\underline{t}_i^T$ is the ith row in matrix $T_S Q^{-1}$. Now, consider the perpendicular direction, i.e., the rows in Y_S. The ith row is denoted as

$$[\gamma_1, \gamma_2, \ldots, \gamma_N] = \underline{t}_i^T[\underline{\beta}_1, \underline{\beta}_2, \ldots, \underline{\beta}_N]. \tag{30}$$

If the left boundary of block X is also available, the left boundary of the row is $\gamma_0 = \underline{t}_i^T \underline{\beta}_0$. Let

$$\epsilon_k = \gamma_k - \rho\gamma_{k-1} = \underline{t}_i^T(\underline{\beta}_k - \rho\underline{\beta}_{k-1}) = \underline{t}_i^T \begin{pmatrix} e_{1,k} \\ e_{2,k} \\ \vdots \\ e_{N,k} \end{pmatrix} \tag{31}$$

where we have used (25) and (29). Since innovations $e_{i,j}$ are Gaussian random variables, so is $\{\epsilon_k\}$. For any integer $h \neq 0$

$$E\{\epsilon_k\epsilon_{k+h}\} = \underline{t}_i^T E\left\{ \begin{pmatrix} e_{1,k} \\ e_{2,k} \\ \vdots \\ e_{N,k} \end{pmatrix} \begin{pmatrix} e_{1,k+h} \\ e_{2,k+h} \\ \vdots \\ e_{N,k+h} \end{pmatrix}^T \right\} \underline{t}_i = 0. \tag{32}$$

Therefore, $\{\epsilon_k\}$ are i.i.d. Hence, sequence $[\gamma_0, \gamma_1, \gamma_2, \ldots, \gamma_N]$ conforms to (1).

Case 2—Top boundary is unavailable: When the top-boundary information is unavailable, no prediction can be performed in the vertical direction, and the transformation is to be directly applied on the pixels. As previously discussed in Case 1, every column in X follows the 1-D AR model, and thus, the optimal transform, as suggested in Section II-B, is the DCT. The transform coefficients of the column vectors are now

$$Y_C = T_C X = [\underline{p}_1, \ldots, \underline{p}_N]^T[\underline{x}_1, \ldots, \underline{x}_N]$$

where $\underline{x}_k = [x_{1,k}, \ldots, x_{N,k}]^T$ is the kth column in X and $\underline{p}_i^T$ is the ith row in T_C. Now, consider the ith row in Y_C, denoted as

$$[\gamma_1, \gamma_2, \ldots, \gamma_N] = \underline{p}_i^T[\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_N] \tag{33}$$

with boundary sample $\gamma_0 = \underline{p}_i^T \underline{x}_0$, when the left-boundary $\underline{x}_0$ is available. Let

$$\begin{aligned} \epsilon_k &= \gamma_k - \rho\gamma_{k-1} = \underline{p}_i^T[\underline{x}_k - \rho\underline{x}_{k-1}] \\ &= \underline{p}_i^T \begin{pmatrix} x_{1,k} - \rho x_{1,k-1} \\ x_{2,k} - \rho x_{2,k-1} \\ \vdots \\ x_{N,k} - \rho x_{N,k-1} \end{pmatrix}. \end{aligned} \tag{34}$$

Again, as in (32), it can be shown that $\{\epsilon_k\}$ are i.i.d. Gaussian random variables and, thus, sequence $[\gamma_0, \gamma_1, \gamma_2, \ldots, \gamma_N]$ follow the AR model in (1).

2) Prediction and Transform in the Horizontal Direction: Note that, irrespective of the availability of the top boundary,

[2]For simplicity, a constant inter pixel correlation coefficient ρ is assumed in our model. We note that a more complicated model with spatially adaptive ρ is expected to further improve the overall coding performance.

the results in Cases 1 and 2 of Section III-A1 show that the elements of each row of the block obtained by transforming the columns of X follow the 1-D AR model. Thus, the conclusions in Section II can be again applied to each individual row of this block of transformed columns.

Case 1—Left boundary is available: The prediction for $[\gamma_1, \gamma_2, \ldots, \gamma_N]$ is $Q^{-1}[\gamma_0, 0, \ldots, 0]$, where γ_0 is computed as indicated in either Case 1 or 2 of Section III-A1. The residuals can now be transformed by the application of the ADST on each row and can be then encoded.

Case 2—Left boundary is unavailable: The DCT is directly applied on to the row vectors. No prediction is employed.

In summary, the hybrid transform coding scheme accomplishes the 2-D transformation of a block of pixels as two sequential 1-D transforms separately performed on rows and columns. The choice of 1-D transform for each direction is dependent on the corresponding prediction boundary condition.

1 Vertical transform: Employ the ADST if the top boundary is used for prediction; use the DCT if the top boundary is unavailable.
2 Horizontal transform: Employ the ADST if the left boundary is used for prediction; use the DCT if the left boundary is unavailable.

B. Implementation in H.264/AVC

We now discuss the implementation of the aforementioned hybrid transform coding scheme in the framework of the H.264/AVC. The standard intra coder defines nine candidate prediction modes, each of which corresponding to a particular spatial prediction direction. Among these, Vertical (Mode 0), Horizontal (Mode 1), and direct-current (DC) (Mode 2) modes are the most frequently used. We focus on these three modes to illustrate the basic principles and demonstrate the efficacy of the approach. Implementation for remaining directional modes can be derived along similar lines.

The standard DC mode is illustrated in Fig. 3(a) for the 4×4 block of pixels denoted $a, b, \ldots, p$. All the 16 pixels share the same prediction, which is the mean of the boundary pixels M, $A-D$, and $I-L$. The standard encoder follows up this prediction with a DCT in both vertical and horizontal directions. Note that the DC mode implies that both the upper and left boundaries of the block are available for prediction. The proposed hybrid transform coding scheme, when incorporated into H.264/AVC, modifies the DC mode as follows: the columns are first predicted as described in Section III-A and the residues are then transformed via the ADST. The same process is repeated on rows of the block of transformed columns.

The standard Vertical mode shown in Fig. 3(b) only uses the top boundary for prediction, while the left boundary is assumed unavailable. The standard encoder then sequentially applies the DCT in vertical and horizontal directions. In contrast, when the proposed hybrid transform coding scheme is incorporated into H.264/AVC, the Vertical mode is modified as follows: the columns of prediction residuals are first transformed with the ADST, and subsequently, the DCT is applied in the horizontal direction. In a similarly modified Horizontal mode, the ADST is applied to the rows, and the DCT is applied to the columns.

Note that our derivations through Section III-A of the hybrid transform coding scheme assumed zero-mean source samples. In practice, however, the image signal has a local mean value that varies across regions. Hence, when operating on a block, it is necessary to *remove* its local mean, from both the boundary and the original blocks of pixels to be coded. The hybrid transform coding scheme operates on these mean-removed samples. The mean value is then added back to the pixels during reconstruction at the decoder. In our implementation, the mean is simply calculated as the average of reconstructed pixels in the available boundaries of the block. This method obviates the need to transmit the mean in the bitstream.

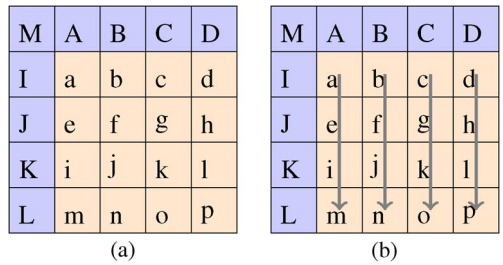


Fig. 3. Examples of intra-prediction mode. (a) DC mode, both upper and left boundaries (M, $A-D$, and $I-L$) are used for prediction. (b) Vertical mode, only the upper pixels ($A-D$) are considered as effective boundary for prediction.

C. Entropy Coding

Here, we discuss some practical issues that arise in the implementation of the proposed hybrid transform within the H.246/AVC intra mode. The entropy coders in H.264/AVC, i.e., typically context-adaptive binary arithmetic coding or context-adaptive variable-length coding, are based on run-length coding. Specifically, the coder first orders the quantized transform coefficients (indexes) of a 2-D block in a 1-D sequence, at the decreasing order of expected magnitude. A number of model-based lookup tables are then applied to efficiently encode the non-zero indexes and the length of the tailing zeros. The efficacy of such entropy coder schemes relies on the fact that nonzero indexes are concentrated in the front end of the sequence. A zigzag scanning fashion [2] is employed in the standard since the lower frequency coefficients in both dimensions tend to have higher energy. Our experiments with the hybrid transform coder show that the same zigzag sequence is still applicable to the modified DC mode but does not hold for the modified Vertical and Horizontal modes. A similar observation has been reported in [10], where MDDTs are applied to the prediction error. We note that this phenomenon tends to be accentuated in our proposed hybrid transform scheme, which optimally exploits the true statistical characteristics of the residual.

To experimentally illustrate this point, we encoded the luminance component of 20 frames of the *carphone* sequence in Intra mode at QP $= 20$ and computed the average of the absolute values of quantized transform coefficients across blocks of size 4×4 for which the encoder selected the modified Horizontal mode. The following matrix of average values was obtained:

$$\begin{pmatrix} 2.49 & 0.66 & 0.19 & 0.04 \\ 2.00 & 0.51 & 0.14 & 0.02 \\ 1.61 & 0.38 & 0.11 & 0.01 \\ 1.12 & 0.25 & 0.06 & 0.00 \end{pmatrix}. \quad (35)$$

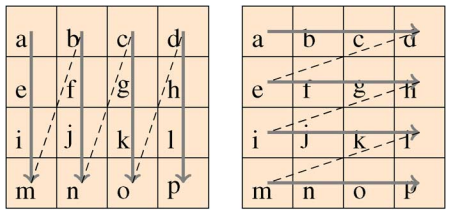


Fig. 4. Scanning order in hybrid transform coding scheme. (a) Horizontal mode; scan columns sequentially from left to right. (b) Vertical mode; scan rows sequentially from top to bottom.

This observation is consistent with our proposal of a scanning order, as shown in Fig. 4(a), for the Horizontal mode, which results in a coefficient sequence with decreasing order of the expected coefficient magnitude, thus enhancing the efficiency of the entropy coder. In a similar vein, we use the scanning order shown in Fig. 4(b) for the Vertical mode.

IV. Integer Hybrid Transform

A low-complexity version of the DCT, called the Int-DCT, has been proposed in [17] and adopted by the H.264/AVC standard [2]. The scheme employs an integer transform with orthogonal (instead of orthonormal) bases and embeds the normalization in the quantization of each coefficient, thus requiring calculations with simple integer arithmetic and significantly reducing the complexity. The overall effective integer transform T_{IC} is a close element-wise approximation to the DCT T_C. Specifically

$$T_C \approx T_{\mathrm{IC}} = \Lambda_C H_C = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{10}} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{10}} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{pmatrix} \tag{36}$$

where Λ_C is a diagonal matrix and H_C is the transform matrix used in H.264/AVC. For a given quantization step size Δ, the encoder can perform the transform and the quantization of vector $\underline{y}$, and the output after quantization is Y_q given as

$$\begin{aligned} Y_q &= \left\lceil (T_{\mathrm{IC}}\underline{y}) \otimes \left(\frac{1}{\Delta} \cdot 1(N,1) \right) \right\rceil \\ &= \left\lceil (\Lambda_C H_C \underline{y}) \otimes \left(\frac{1}{\Delta} \cdot 1(N,1) \right) \right\rceil \\ &= \left\lceil (H_C \underline{y}) \otimes \left(\Lambda_C \frac{1}{\Delta} \cdot 1(N,1) \right) \right\rceil \end{aligned} \tag{37}$$

where $\otimes$ denotes element-wise multiplication, $\lceil \cdot \rceil$ denotes the rounding function, and $1(N,1)$ is the "all 1's" column vector of N entries. Since all elements in H_C are integers, the transformation only requires additions and shifts. The resulting coefficients are quantized with weighted step sizes that incorporate the normalization factor Λ_C. The efficacy of the Int-DCT has been reported in [17]. The Int-DCT is directly used instead of the floating-point DCT in H.264/AVC to avoid drift issues due to the floating-point arithmetic and for the ease of implementation of Int-DCT via adders and shifts.

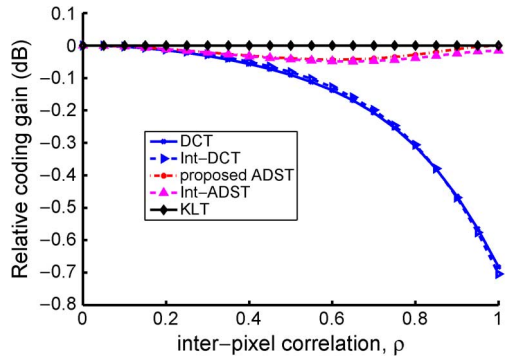


Fig. 5. Theoretical coding gains of the ADST/Int-ADST and the DCT/Int-DCT, all applied to the prediction residuals (given exact knowledge of the boundary), plotted versus the inter pixel correlation coefficient. The block dimension is 4×4.

Analogous to the Int-DCT, we now propose an integer version of the proposed ADST, namely, the Int-ADST. The floating-point sine transform is repeated herein as

$$[T_S]_{j,i} = \left(\frac{2}{\sqrt{2N+1}} \sin \frac{(2j-1)i\pi}{2N+1} \right) \tag{38}$$

whose elements are, in general, irrational numbers. The Int-DST T_{IS} approximates T_S as follows:

$$\begin{aligned} T_S \approx T_{\mathrm{IS}} &= \Lambda_S H_S \\ &= \begin{pmatrix} \frac{1}{\sqrt{147}} & 0 & 0 & 0 \\ 0 & \frac{7}{\sqrt{147}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{147}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{147}} \end{pmatrix} \begin{pmatrix} 3 & 5 & 7 & 8 \\ 1 & 1 & 0 & -1 \\ 8 & -3 & -7 & 5 \\ 5 & -8 & 7 & -3 \end{pmatrix}. \end{aligned} \tag{39}$$

When applied to vector $\underline{y}$, the Int-DST can be implemented as follows:

$$\begin{aligned} Y_q &= \left\lceil (T_{\mathrm{IS}}\underline{y}) \otimes \left(\frac{1}{\Delta} \cdot 1(N,1) \right) \right\rceil \\ &= \left\lceil (\Lambda_S H_S \underline{y}) \otimes \left(\frac{1}{\Delta} \cdot 1(N,1) \right) \right\rceil \\ &= \left\lceil (H_S \underline{y}) \otimes \left(\Lambda_S \frac{1}{\Delta} \cdot 1(N,1) \right) \right\rceil. \end{aligned} \tag{40}$$

Again, all the elements in H_S are integers; thus, computing $(H_S\,\underline{y})$ only requires addition and shift operations. We evaluate the coding performance of T_{IS} relative to the KLT and compare it with the original ADST T_S and DCT/Int-DCT at different values of inter pixel correlation ρ in Fig. 5. The performance of the Int-ADST is very close to that of the ADST and the KLT, and it substantially outperforms the DCT/Int-DCT at high values of ρ. The maximum gap between coding gains of the Int-ADST and the ADST is about 0.02 and 0.05 dB between the Int-ADST and the KLT.

V. Simulation Results

The proposed hybrid transform coding scheme is implemented within the H.264/AVC Intra framework. All the nine prediction modes are enabled, among which the DC, Vertical,

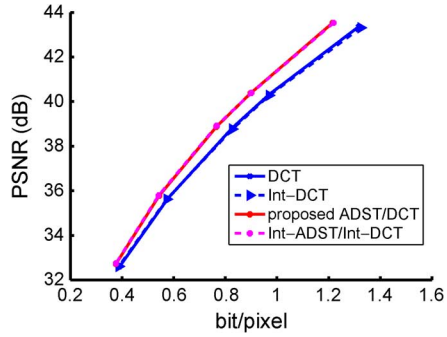


Fig. 6. Coding performance on *carphone* at the QCIF resolution. The proposed hybrid transform coding scheme (ADST/DCT) and its integer version (Int-ADST/Int-DCT) are compared with H.264 Intra coding using the DCT and the Int-DCT.

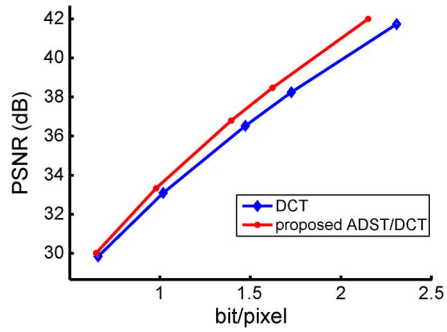


Fig. 7. Coding performance on *harbor* at the CIF resolution. The proposed hybrid transform coding scheme (ADST/DCT) is compared with H.264/AVC Intra using the DCT.

TABLE I
RELATIVE BIT SAVINGS OF THE ADST/DCT IN COMPARISON WITH DCT AT CERTAIN OPERATING POINT SEQUENCES AT THE QCIF RESOLUTION

bit savings (%)	PSNR (dB)		
Test Sequence	42	38	34
foreman	6.1	4.7	3.6
carphone	10.1	9.1	6.4
coastguard	4.7	3.4	2.6
highway	4.0	3.0	1.4
container	4.4	4.7	5.1
salesman	6.2	5.4	4.3

TABLE II
RELATIVE BIT SAVINGS OF THE ADST/DCT IN COMPARISON WITH DCT AT CERTAIN OPERATING POINT SEQUENCES AT THE CIF RESOLUTION

bit savings (%)	PSNR (dB)		
Test Sequence	42	38	34
waterfall	8.3	6.0	4.3
container	4.3	4.4	3.9
harbour	11.0	8.5	7.5
foreman	4.6	3.4	1.5
mobile	3.9	3.7	4.1
flower	2.7	2.5	2.6
bus	8.3	5.3	4.9
city	5.9	5.0	4.7

and Horizontal modes are modified, as proposed in Section III. To encode a block, every prediction mode (with the associated transform) is tested, and the decision is made by rate-distortion optimization.

For quantitative comparison, the first ten frames of *carphone* at QCIF resolution are encoded in Intra mode using the proposed hybrid transform coding scheme, the conventional DCT, and their integer versions, respectively. The coding performance is shown in Fig. 6. Clearly, the proposed approach provides better compression efficiency than DCT, particularly at medium-to-high quantizer resolution. When quantization is coarse (i.e., the bitrate is low), the boundary distortion is significant, and the DCT approaches the performance of the ADST, as discussed in Section II-C. When the integer version of either transform is employed in the encoder, the same performance as its floating-point counterpart is observed. We have, in fact, established that the performance of the integer versions is generally hardly distinguishable from the floating-point version of the same transform. Hence, from now on, to avoid unnecessary clutter in the presentation of the remainder of the experiments, we will only show results for the original floating-point transforms. More simulation results of test sequences at the QCIF resolution are shown in Table I, where the operating points are chosen as 42, 38, and 34 dB of PSNR, and the coding performance is evaluated in terms of relative bit savings. The coding performance in the context of sequences at the CIF resolution is demonstrated in Fig. 7 and Table II.

Competing transforms in image coding are compared in terms of compression performance, complexity, and perceptual quality [18]. We next focus on the perceptual comparison. The decoded frames of *carphone* at the QCIF resolution using the proposed codec and H.264/AVC Intra coder at 0.3 bits/pixel are shown in Fig. 8. The basis vectors of the DCT maximize their energy distribution at both ends; hence, the discontinuity at block boundaries due to quantization effects (commonly referred to as blockiness) are magnified [see Fig. 8(c)]. Although the deblocking filter, which is, in general, a low-pass filter applied across the block boundary, can mitigate such blockiness, it also tends to compromise the sharp curves, e.g., the face area in Fig. 8(d). In contrast with the DCT, the basis vectors of the proposed ADST minimize their energy distribution as they approach the prediction boundary and maximize it at the other end. Hence, they provide smooth transition to neighboring blocks, without recourse to an artificial deblocking filter. Therefore, the proposed hybrid transform coding scheme provides consistent reconstruction and preserves more details, as shown in Fig. 8(b).

VI. CONCLUSION

In this paper, we have described a new compression scheme for image and intra coding in video, which is based on a hybrid transform coding scheme in conjunction with the intra prediction from available block boundaries. A new sine transform, i.e., the ADST, has been analytically derived for the prediction residuals in the context of intra coding. The proposed scheme switches between the sine transform and the standard DCT depending on the available boundary information, and the resulting hybrid transform coding has been implemented within the H.264/AVC intra mode. An integer low-complexity version of the proposed sine transform has been also derived,

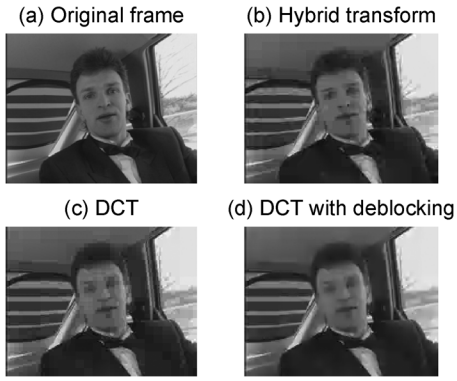


Fig. 8. Perceptual comparison of reconstructions of *carphone* Frame 2 at the QCIF resolution. (b) The proposed hybrid transform coding scheme is compared with (a) the original frame and (c) H.264/AVC Intra (d) without deblocking filter.

which can be directly implemented within the H.264/AVC system and avoids any drift due to the floating-point operation. The theoretical analysis of the coding gain shows that the proposed ADST has a performance that is very close to the KLT and substantially outperforms the conventional DCT for intra coding. The proposed transform scheme also efficiently exploits inter block correlations, thereby reducing the blocking effect. Simulation results demonstrate that the hybrid transform coding scheme outperforms the H.264/AVC intra-mode both perceptually and quantitatively.

APPENDIX A
PROOF OF CASE 2 IN SECTION II-B

Claim: The KLT for $\hat{P}_2$ in (19) is DCT given by

$$[T_C]_{j,i} = \left(\alpha \cos\left(\frac{\pi(j-1)(2i-1)}{2N}\right)\right) \tag{41}$$

where $j, i \in \{1, 2, \ldots, N\}$ are the frequency and time indexes of the transform kernel, respectively, and

$$\alpha = \begin{cases} \sqrt{\frac{1}{N}}, & j = 1 \\ \sqrt{\frac{2}{N}}, & j = 2, 3, \ldots, N. \end{cases}$$

The eigenvalue associated with the jth eigenvector $\underline{v}_j$ is

$$\lambda_j = 1 + \rho^2 - 2\rho \cos\left(\frac{\pi(j-1)}{N}\right).$$

Proof: To verify this statement, let us consider a vector quantity that measures by how much we deviate from the eigenvalue/eigenvector condition, i.e.,

$$\underline{w}_j = \hat{P}_2 \underline{v}_j - \lambda_j \underline{v}_j = (\hat{P}_2 - \lambda_j I)\underline{v}_j \tag{42}$$

where I denotes the identity matrix. The first entry of $\underline{w}_j$ is

$$\begin{aligned} &\alpha\rho\left\{\left[2cos\left(\frac{\pi(j-1)}{N}\right) - 1\right]\cos\left(\frac{\pi(j-1)}{2N}\right)\right. \\ &\quad \left. - \cos\left(\frac{3\pi(j-1)}{2N}\right)\right\} \\ &= 2\alpha\rho\cos\left(\frac{\pi(j-1)}{N}\right) cos\left(\frac{\pi(j-1)}{2N}\right) \\ &\quad - \alpha\rho\left(cos\left(\frac{\pi(j-1)}{2N}\right) + \cos\left(\frac{3\pi(j-1)}{2N}\right)\right) = 0. \end{aligned}$$

The last entry of $\underline{w}_j$ can be computed as

$$\begin{aligned} &\alpha\rho\left[2\cos\left(\frac{\pi(j-1)}{N}\right) - 1\right]\cos\left(\frac{\pi(j-1)(2N-1)}{2N}\right) \\ &\quad - \alpha\rho\cos\left(\frac{\pi(j-1)(2N-3)}{2N}\right) \\ &= 2\alpha\rho\cos\left(\frac{\pi(j-1)}{N}\right)\cos\left(\frac{\pi(j-1)(2N-1)}{2N}\right) \\ &\quad - \alpha\rho\left(\cos\left(\frac{\pi(j-1)(2N-1)}{2N}\right)\right. \\ &\qquad \left. + \cos\left(\frac{\pi(j-1)(2N-3)}{2N}\right)\right) \\ &= 2\alpha\rho\cos\left(\frac{\pi(j-1)}{N}\right)\cos\left(\frac{\pi(j-1)(2N-1)}{2N}\right) \\ &\quad - 2\alpha\rho\cos\left(\frac{\pi(j-1)(2N-2)}{2N}\right)\cos\left(\frac{\pi(j-1)}{2N}\right). \end{aligned}$$

Since

$$\begin{aligned} &\cos\left(\frac{\pi(j-1)(2N-2)}{2N}\right)\cos\left(\frac{\pi(j-1)}{2N}\right) \\ &\quad = \cos\left(\pi(j-1) - \frac{\pi(j-1)(2N-2)}{2N}\right) \\ &\qquad \times \cos\left(\pi(j-1) - \frac{\pi(j-1)}{2N}\right) \\ &\quad = \cos\left(\frac{\pi(j-1)}{N}\right)\cos\left(\frac{\pi(j-1)(2N-1)}{2N}\right) \end{aligned}$$

the last entry of $\underline{w}_j$ is zero. The remaining kth element, where $k \in \{2, 3, \ldots, N-1\}$, is computed as

$$\begin{aligned} &-\alpha\rho\cos\left(\frac{\pi(j-1)(2k-3)}{2N}\right) - \alpha\rho\cos\left(\frac{\pi(j-1)(2k+1)}{2N}\right) \\ &\quad +2\alpha\rho\cos\left(\frac{\pi(j-1)}{N}\right)\cos\left(\frac{\pi(j-1)(2k-1)}{2N}\right) = 0. \end{aligned}$$

Hence, all the entries of $\underline{w}_j$ are zero, i.e., $\underline{w}_j = \underline{0}$, which implies that $\hat{P}_2\underline{v}_j = \lambda_j\underline{v}_j$. Indeed, $\underline{v}_j$ is the eigenvector of $\hat{P}_2$, and λ_j is the eigenvalue associated with $\underline{v}_j$, for all $j \in \{1, 2, \ldots, N\}$. Therefore, DCT T_C is the KLT for $\hat{P}_2$ in (19).

REFERENCES

[1] K. R. Rao and P. Yip, *Discrete Cosine Transform-Algorithms, Advantages and Applications*. New York: Academic, 1990.

[2] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[3] D. Marpe, V. George, H. L. Cycon, and K. U. Barthel, "Performance evaluation of Motion-JPEG2000 in comparison with H.264/AVC operated in pure intra coding mode," in *Proc. SPIE*, Oct. 2003, vol. 5266, pp. 129–137.

[4] A. K. Jain, "A fast Karhunen–Loeve transform for a class of random processes," *IEEE Trans. Commun.*, vol. COM-24, no. 9, pp. 1023–1029, Sep. 1976.

[5] A. K. Jain, "Image coding via a nearest neighbors image model," *IEEE Trans. Commun.*, vol. COM-23, no. 3, pp. 318–331, Mar. 1975.

[6] A. Z. Meiri and E. Yudilevich, "A pinned sine transform image coder," *IEEE Trans. Commun.*, vol. COM-29, no. 12, pp. 1728–1735, Dec. 1981.

[7] N. Yamane, Y. Morikawa, and H. Hamada, "A new image data compression method-extrapolative prediction discrete sine transform coding," *Electron. Commun. Jpn.*, vol. 70, no. 12, pp. 61–74, 1987.

[8] K. Rose, A. Heiman, and I. Dinstein, "DCT/DST alternate-transform image coding," *IEEE Trans. Commun.*, vol. 38, no. 1, pp. 94–101, Jan. 1990.

[9] B. Zeng and J. Fu, "Directional discrete cosine transforms—A new framework for image coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 3, pp. 305–313, Mar. 2008.

[10] Y. Ye and M. Karczewicz, "Improved H.264 intra coding based on bi-directional intra prediction, directional transform, and adaptive coefficient scanning," in *Proc. IEEE ICIP*, Oct. 2008, pp. 2116–2119.

[11] H. S. Malvar and D. H. Staelin, "The LOT: Transform coding without blocking effects," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 4, pp. 553–559, Apr. 1989.

[12] T. Sikora and H. Li, "Optimal block-overlapping synthesis transforms for coding images and video at very low bitrates," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 2, pp. 157–167, Apr. 1996.

[13] J. Xu, F. Wu, and W. Zhang, "Intra-predictive transforms for block-based image coding," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3030–3040, Aug. 2009.

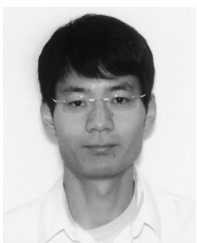
[14] J. Han, A. Saxena, and K. Rose, "Towards jointly optimal spatial prediction and adaptive transform in video/image coding," in *Proc. IEEE ICASSP*, Mar. 2010, pp. 726–729.

[15] W. C. Yueh, "Eigenvalues of several tridiagonal matrices," *Appl. Math. E-Notes*, vol. 5, pp. 66–74, Apr. 2005.

[16] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[17] H. S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 598–603, Jul. 2003.

[18] G. Strang, "The discrete cosine transform," *SIAM Rev.*, vol. 41, no. 1, pp. 135–147, Mar. 1999.



Jingning Han (S'10) obtained the B.S. degree in electrical engineering in 2007 from Tsinghua University, Beijing, China, and the M.S. degree in electrical and computer engineering in 2008 from the University of California, Santa Barbara, where he is currently working toward the Ph.D. degree.

He interned with Ericsson, Inc., during the summer of 2008 and in Technicolor, Inc., in 2010. His research interests include video compression and networking.

Mr. Han was the recipient of the outstanding teaching assistant awards from the Department of Electrical and Computer Engineering, University of California, Santa Barbara, in 2010 and 2011, respectively.

Ankur Saxena (S'06–M'09) was born in Kanpur, India, in 1981. He received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Delhi, India, in 2003 and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of California, Santa Barbara, in 2004 and 2008, respectively.

He has interned with Fraunhofer Institute of X-Ray Technology, Erlangen, Germany, and NTT DoCoMo Research Laboratories, Palo Alto, in the summers of 2002 and 2007, respectively. He is currently a Senior Research Engineer with Samsung Telecommunications America, Richardson, TX. His research interests span source coding, image and video compression, and signal processing.

Dr. Saxena was a recipient of the President Work study award during his Ph.D. and the Best Student Paper Finalist at the IEEE International Conference on Acoustics, Speech, and Signal Processing 2009.

Vinay Melkote (S'08–M'10) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology Madras, Chennai, India, in 2005 and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of California, Santa Barbara, in 2006 and 2010, respectively.

He interned with the Multimedia Codecs Division, Texas Instruments, India, during the summer of 2004 and with the Audio Systems Group, Qualcomm, Inc., San Diego, in 2006. He is currently with the Sound Technology Research Group, Dolby Laboratories, Inc., San Francisco, CA, where he focuses on audio compression and related technologies. His other research interests include video compression and estimation theory.

Dr. Melkote is a student member of the Audio Engineering Society. He was a recipient of the Best Student Paper Award at the IEEE International Conference on Acoustics, Speech, and Signal Processing 2009.

Kenneth Rose (S'85–M'91–SM'01–F'03) received the Ph.D. degree from the California Institute of Technology, Pasadena, in 1991.

He then joined the Department of Electrical and Computer Engineering, University of California at Santa Barbara, where he is currently a Professor. His main research activities are in the areas of information theory and signal processing, and include rate-distortion theory, source and source-channel coding, audio and video coding and networking, pattern recognition, and nonconvex optimization. He is interested in the relations among information theory, estimation theory, and statistical physics, and their potential impact on fundamental and practical problems in diverse disciplines.

Dr. Rose was a corecipient of the 1990 William R. Bennett Prize Paper Award of the IEEE Communications Society, as well as the 2004 and 2007 IEEE Signal Processing Society Best Paper Awards.