TOWARDS JOINTLY OPTIMAL SPATIAL PREDICTION AND ADAPTIVE TRANSFORM IN VIDEO/IMAGE CODING

Jingning Han, Ankur Saxena and Kenneth Rose

Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106
E-mail: {jingning,ankur,rose}@ece.ucsb.edu

ABSTRACT

This paper proposes a new approach to combined spatial (Intra) prediction and adaptive transform coding in block-based video and image compression. Context-adaptive spatial prediction from available, previously decoded boundaries of the block, is followed by optimal transform coding of the prediction residual. The derivation of both the prediction and the adaptive transform for the prediction error, assumes a separable first-order Gauss-Markov model for the image signal. The resulting optimal transform is shown to be a close relative of the *sine* transform with phase and frequencies such that basis vectors tend to vanish at known boundaries and maximize energy at unknown boundaries. The overall scheme switches between the above sine-like transform and discrete cosine transform (per direction, horizontal or vertical) depending on the prediction and boundary information. It is implemented within the H.264/AVC intra mode, is shown in experiments to significantly outperform the standard intra mode, and achieve significant reduction of the blocking effect.

Index Terms— Transform Coding, H.264 Intra mode, Image Compression, Blocking effect

1. 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 transformed samples. However, practical use of KLT is limited due to its high computational complexity. Amongst a variety of alternative transforms, the discrete cosine transform (DCT) has been shown to offer good energy compaction [1] for image compression and to attain performance close to that of KLT.

Practical considerations, such as underly the H.264/AVC intra mode, dictate transform coding implementation within a block coder with typical blocks of size 4×4 to 16×16. A DCT-based block coder suffers from blocking effect, i.e., a disturbing discontinuity at the block boundaries. Although post-filtering can smooth the boundaries, it incurs information loss, such as blurring sharp details.

Much research effort has been leveraged to reduce the blocking effect. In [2], 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 [2] [3] showed the KLT of the residual process to be the discrete sine transform (DST) when the boundary conditions are available in both directions. A related approach by Meiri and Yudilevich [4] first encodes the block content and then separately the boundaries, and finally applies a "pinned sine transform". However these approaches sacrifice some efficiency as they require separate coding procedures for block boundaries. To attain better compression in image coding, various transforms have been combined and proposed in literature. For example, [5] proposed alternate usage of sine and cosine transforms on image blocks to efficiently exploit inter block redundancy. Another approach of using directional cosine transforms to capture the texture of block content efficiently has been developed in [6].

In this paper, we address the related problem of jointly optimizing spatial prediction and the corresponding transform bases. We assume a separable Gauss-Markov model for the images and compute the prediction error statistics based only on the available decoded boundary. The mathematical analysis shows that the optimal transform is a relative of the known sine transform with appropriate phase shifts and frequencies. We propose a hybrid coding scheme that allows choosing from the proposed sine transform and the DCT. Simulations show that the proposed hybrid transform coding scheme significantly reduces the blocking effect and attain 7%-10% bit savings at same PSNR. Also note that the intra mode in H.264/AVC, which utilizes spatial prediction and DCT transform, has been shown to have better rate-distortion performance than wavelet-based Motion-JPEG2000 at low and medium resolution frame/image such as QCIF and CIF [7]. Thus our proposed block-based hybrid transform coding scheme is a strong contender in the context of general image coding as well. A related problem of jointly performing prediction and transform coding has also been studied in [8], where the emphasis was on designing a low-complexity intra-predictive transform.

The paper is organized as follows. Sec. 2 presents a mathematical analysis for spatial prediction in video coding. Sec. 3 describes the hybrid transform coding scheme and outlines the implementation details of the hybrid coding scheme in H.264/AVC intra mode and the revised entropy coder. Simulation results are presented in Sec. 4 followed by conclusions in Sec. 5.

2. SPATIAL PREDICTION AND RESIDUAL TRANSFORM CODING

Consider a zero-mean, unit variance, first-order Gauss-Markov sequence

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

This work was supported in part by the University of California MICRO Program, Applied Signal Technology Inc., Qualcomm Inc., and Sony Ericsson Inc.

where ρ is the correlation coefficient, and e_k is a white Gaussian noise process with variance $(1-\rho^2)$. The autocorrelation of sequence $\{x_k\}$ is: $R(m,n) = E(x_m x_n) = \rho^{|m-n|}$. Let $\underline{x} = [x_1, x_2, \cdots, x_N]^T$ denote the random vector to be encoded given x_0 as the available (one-sided) boundary. The recursion (1) translates into a 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 matrix-vector form:

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

where

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

and $\underline{b} = [\rho x_0, 0, \cdots, 0]^T$ captures the boundary data. Since Q is invertible, we may equivalently write

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

a decomposition of the signal into the sum of the "boundary response" or prediction $Q^{-1}\underline{b}$, and the prediction residual $\underline{y} = Q^{-1}\underline{e}$. The residual $\underline{y}$ must be compressed and transmitted. The autocorrelation matrix of $\underline{y}$ is given by:

$$R_{\underline{y}\underline{y}} = E(\underline{y}\underline{y}^T) = Q^{-1}E(\underline{e}\underline{e}^T)Q^{-T} = (1-\rho^2)Q^{-1}Q^{-T}. \tag{6}$$

The optimal transform (KLT) for $\underline{y}$ is a unitary matrix that diagonalizes $Q^{-1}Q^{-T}$, and hence also the more convenient $P_1 = Q^T Q$:

$$P_1 = \begin{pmatrix} 1+\rho^2 & -\rho & 0 & 0 & \dots \\ -\rho & 1+\rho^2 & -\rho & 0 & \dots \\ 0 & -\rho & 1+\rho^2 & -\rho & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & -\rho & 1+\rho^2 & -\rho \\ 0 & \dots & 0 & -\rho & 1 \end{pmatrix}. \tag{7}$$

While P_1 is Toeplitz, in general it is difficult to find its eigenvalues and eigenvectors because of the irregularity at the bottom-right corner (see e.g., [9]). As a subterfuge, we approximate the bottom-right corner element 1 by $1+\rho^2-\rho$. The approximation is clearly good for $\rho \to 1$ or $\rho \to 0$. We thus consider

$$\hat{P}_1 = \begin{pmatrix} 1+\rho^2 & -\rho & 0 & 0 & \dots \\ -\rho & 1+\rho^2 & -\rho & 0 & \dots \\ 0 & -\rho & 1+\rho^2 & -\rho & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & -\rho & 1+\rho^2 & -\rho \\ 0 & \dots & 0 & -\rho & 1+\rho^2-\rho \end{pmatrix} \tag{8}$$

whose KLT is explicitly calculated in [9] as a sinusoidal transform:

$$[T_S]_{j,i} = \left(\frac{2}{\sqrt{2N+1}} \sin \frac{(2j-1)i\pi}{2N+1} \right) \tag{9}$$

where $j, i = 1, 2, \cdots, N$. Note that T_S is independent of the statistics of the innovation e_k and hence the optimal transform conditioned on x_0 is given by a constant matrix.

We next consider the case where the exact boundary value x_0 is unavailable. For example, we may only have access to the reconstructed boundary, $\hat{x}_0 = x_0 + \delta$ (which also covers the special case where no boundary information is available.) The modification applies only to the first equation in (2), which becomes:

$$x_1 = \rho\hat{x}_0 + f_1. \tag{10}$$

If we assume that δ is independent of, and small in magnitude relative to x_0, then f_1 may be considered independent of $\{e_2, e_3, \dots, e_N\}$. Thus,

$$\begin{aligned} E(f_1^2) &= E(x_1 - \rho\hat{x}_0)^2 \\ &= 1 - \rho^2 + \rho^2 E(\delta^2) = 1 - \rho^2 + \rho^2\sigma^2, \end{aligned} \tag{11}$$

where $\sigma^2 = E(\delta^2)$, and we may reuse (5), i.e.,

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

where in this case $\underline{b} = [\rho\hat{x}_0, 0, \cdots, 0]^T$ and $\underline{e} = [f_1, e_2, \cdots, e_N]^T$. The autocorrelation matrix of the residual vector is:

$$\begin{aligned} R &= E(Q^{-1}\underline{e}\underline{e}^T Q^{-T}) \\ &= Q^{-1} diag(1-\rho^2+\rho^2\sigma^2, 1-\rho^2, \cdots, 1-\rho^2) Q^{-T} \\ &= (1-\rho^2)\left(Q^T diag(\frac{1}{1+\frac{\rho^2\sigma^2}{1-\rho^2}}, 1, \cdots, 1) Q \right)^{-1} \\ &= (1-\rho^2)P_2^{-1}. \end{aligned} \tag{12}$$

The KLT of R also diagonalizes the matrix P_2 defined above, which is explicitly:

$$P_2 = \begin{pmatrix} \rho^2 + \frac{1-\rho^2}{1-\rho^2+\rho^2\sigma^2} & -\rho & 0 & \dots \\ -\rho & 1+\rho^2 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 1+\rho^2 & -\rho \\ 0 & \dots & -\rho & 1 \end{pmatrix} \tag{13}$$

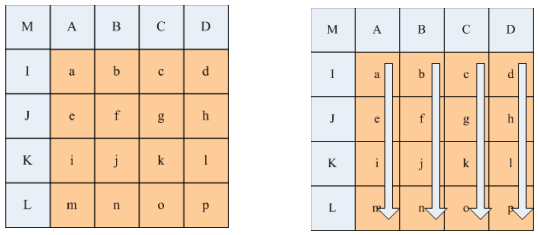
Assuming that $\sigma^2 \ll 1-\rho^2$ (small δ), and that ρ approaches either 1 or 0, we reapply the earlier subterfuge at the bottom-right corner element to replace 1 with $1+\rho^2-\rho$, while at the top-left corner element we note that $\rho^2 + \frac{1-\rho^2}{1-\rho^2+\rho^2\sigma^2}$ approaches ρ^2+1. The resulting optimal transform is again the constant matrix T_S, despite the fact that the boundary is distorted by reconstruction error.

The other case is when no boundary information is available. Here we have $\sigma^2 \gg 1-\rho^2$ (since the δ must now be large). The top-left corner element of P_2 is then:

$$\rho^2 + \frac{1-\rho^2}{1-\rho^2+\rho^2\sigma^2} = \rho^2 + 1 - \frac{\rho^2\sigma^2}{1-\rho^2+\rho^2\sigma^2}, \tag{14}$$

and can be approximated as $1+\rho^2-\rho$ when ρ goes to 1 or 0. P_2 can thus be approximated by:

$$\hat{P}_2 = \begin{pmatrix} 1+\rho^2-\rho & -\rho & 0 & \dots \\ -\rho & 1+\rho^2 & -\rho & \dots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 1+\rho^2 & -\rho \\ 0 & \dots & -\rho & 1+\rho^2-\rho \end{pmatrix} \tag{15}$$



(a) DC mode (b) Vertical mode

Fig. 1. Examples of intra prediction mode.

whose KLT is exactly the conventional DCT, that we denote T_C.

Thus, it will be advantageous to switch between the sine (T_S) and cosine (T_C) transforms depending on the availability of the horizontal/vertical image block boundaries. We next present the hybrid transform coding and its implementation in H.264/AVC standard.

3. HYBRID TRANSFORM CODING SCHEME

3.1. Transform Coding

The analysis in the previous section motivates the incorporation of switching between sine (9) and cosine transforms into H.264/AVC intra mode. We refer to the proposed scheme as hybrid transform coding. For clarity of presentation, we temporarily disregard the Integer Discrete Cosine Transform adopted in the H.264/AVC standard, and assume a separable 2-D image model.

In H.264/AVC intra mode, there are nine candidate prediction modes for blocks of size 4x4 . Among them, Vertical (Mode 0), Horizontal (Mode 1) and DC (Mode 2) are the most frequently used modes [6]. We focus on these three modes in this work to illustrate our ideas. Analysis for other directional modes is similar.

The DC mode is depicted in Figure 1(a). The prediction at position a-p is the mean of pixels A-D and I-L. Choosing this mode implies both the upper (A-D) and the left (I-L) boundaries to be good reference candidates. Hence, we propose to use a 2-D (boundary response) prediction followed by the sine transform T_S on both vertical and horizontal direction in this case.

Note that our derivation assumed zero mean. Hence when operating on a block, it is necessary to *remove* its local mean $\overline{x}_b$ defined as the average of all reference pixels from the boundary before prediction (For example, for the 4×4 block size shown in Fig. 1(a) , $\overline{x}_b$ is the mean of the pixels $M, A-D$ and $I-L$). Later, we add $\overline{x}_b$ back to the predicted pixel. The operations for 2-D prediction can be summarized as follows: Let $x(0,0) = M - \overline{x}_b$, $x(0,1) = A - \overline{x}_b$, $x(1,0) = I - \overline{x}_b$, etc. are the modified zero-mean boundaries. Let X denote the $N \times N$ matrix of the pixels in the block to be encoded. We have $QXQ^T = B+E$, where B contains the two-side boundary information and E is the residual matrix with all its elements form a white noise process.

Here we consider 2-D prediction and pixel $x(i,j)$, the $\{i,j\}$ element of X can be written as:

$$x(i,j) = \rho x(i-1,j) + \rho x(i,j-1) - \rho^2 x(i-1,j-1) + e_{i,j} \quad (16)$$

where $e(i,j)$ denote the $\{i,j\}$ element of matrix E.

By expanding QXQ^T, it can be shown that only the elements in the top-most row and left-most column of matrix B are non-zero and given as:

$$\begin{aligned} B(1,1) &= \rho x(0,1) + \rho x(1,0) - \rho^2 x(0,0), \\ B(1,j) &= \rho x(0,j) - \rho^2 x(0,j-1) \qquad \forall j = \{2...N\}, \\ B(i,1) &= \rho x(i,0) - \rho^2 x(i-1,0) \qquad \forall i = \{2...N\}. \end{aligned} \quad (17)$$

Thus, the prediction block is:

$$X_b = Q^{-1} B Q^{-T} + \overline{x}_b \cdot 1(N,N). \quad (18)$$

where $1(N,N)$ denote the $N \times N$ matrix with all elements taking the value 1. The residues $X - X_b$ are then transformed into frequency domain by taking the transform as $T_S^T (X - X_b) T_S$.

Next we consider the Vertical mode shown in Figure 1(b). When this mode is chosen, the image block tends to have vertical edges in its content and hence only the upper boundary is reliable. Here the local mean $\overline{x}_b$ is calculated based on the top boundary (using pixels A-D for the figure shown), and prediction is performed in the vertical direction, i.e., $X_b(i,j) = \rho^i x(0,j) + \overline{x}_b$. The sine transform is applied in the vertical direction, while the cosine transform is applied in the horizontal direction of the residual matrix, i.e. $T_S^T (X - X_b) T_C$. The encoding of the Horizontal mode is performed in an analogous manner.

3.2. Entropy Coding

In this section, we discusses some practical issues that arise during the implementation of proposed hybrid transform in H.246/AVC intra mode.

Prior to the entropy coding in H.264/AVC, the quantized transform coefficients of a block are scanned in a zig-zag fashion [7], since the lower frequency coefficients in both dimensions tend to have higher energy. Our experiments using hybrid transform coding show it to be indeed true for DC mode, but this does not hold for Vertical and Horizontal modes.

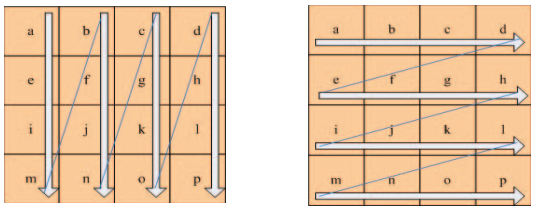
For instance, when we encode the luma component of sequence $carphone.yuv$ in Intra mode, and compute an average of the absolute values of prediction errors in Vertical mode across 4500 blocks of size 4x4, we obtain the following matrix:

$$\begin{pmatrix} 4.252473 & 3.966052 & 4.123201 & 4.240333 \\ 5.640063 & 5.319919 & 5.542041 & 5.879496 \\ 7.136241 & 6.663669 & 7.003822 & 7.358138 \\ 8.645683 & 8.521358 & 9.207284 & 9.456385 \end{pmatrix}. \quad (19)$$

Clearly the coefficients increase along vertical direction, which was expected as the prediction is performed using the Vertical boundary. We thus need to take the sine transform in the vertical direction and the cosine transform in the horizontal direction. The resulting transform coefficients will have more energy in the left-most column in the transformed coefficient matrix. The energy across the subsequent columns increase as we go forward in the transformed coefficient matrix. Hence we propose to use the scanning order as shown in Fig. 2(a) for the Vertical mode. (In our experiments when we tried the standard zig-zag scanning in H.264/AVC, while using the hybrid transform in Vertical mode, we did not notice any performance gains because of the mismatch of the transform coefficients and scanning order.) In a similar vein, we use the scanning order shown in Fig. 2(b) for the Horizontal mode.

4. SIMULATION RESULTS

We demonstrate the performance of our proposed coding scheme by comparing it with the H.264/AVC intra mode. The hybrid coding scheme described above is implemented in JM11.0 with Main



(a) Vertical mode (b) Horizontal mode

Fig. 2. Scanning order of Vertical and Horizontal modes.

Profile, Level 4.0. For visual comparison we show frame 2 of the $carphone\_qcif.yuv$ sequence encoded at 0.3 bits/pixel (Fig. 3). The proposed hybrid transform coding scheme provides smoother reconstruction and conserves more details, especially around the face area. For quantitative comparison, we show the rate-distortion

Fig. 3. Visual comparison of reconstructions ($carphone$).

curves for the first 10 frames of QCIF sequences $foreman$ and $carphone$ in Fig. 4. The proposed hybrid transform coding scheme gives higher compression efficiency especially at medium to high bit-rate due to its superior energy compaction property and upto 10% bit-savings are observed at PSNR higher that 40 dB. (At low bit rates, most of the transform coefficients are 0 and both the proposed hybrid coding scheme and the DCT in H.264/AVC have similar rate-distortion performance, since no residual information is transmitted).

It should be mentioned that during our simulations, we set the correlation coefficient ρ equal to 0.95 for simplicity. The overall performance could be improved by adapting the local correlation coefficient in an image and calculating it explicitly from the neighboring blocks. Further, the proposed hybrid transform method can be directly combined with scalar quantization to make it an integer transform and reduce the computational complexity. We leave both these directions as part of future work.

5. CONCLUSIONS

This paper describes an image (and video) compression technique that is based on a hybrid transform coding scheme in conjunction

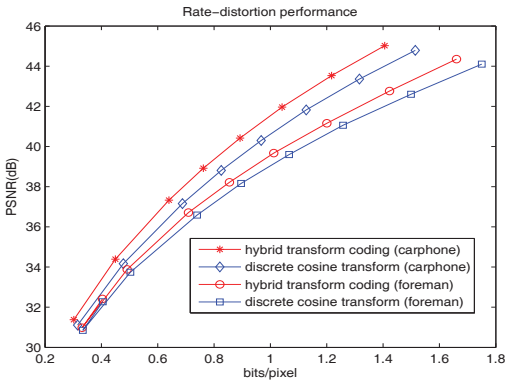


Fig. 4. Rate-distortion curves of $foreman$ and $carphone$.

with intra prediction from available block boundaries. It alternates between sinusoidal transforms with appropriate phase and frequency parameters (specifically, a variant of the known sine transform, and the standard cosine transform) depending on the boundary condition. It efficiently exploits inter-block correlations. Simulation results demonstrate that the hybrid transform coding scheme outperforms H.264/AVC intra mode both subjectively and quantitatively.

6. ACKNOWLEDGMENT

The authors are grateful to Emrah Akyol for seeing the potential of revisiting this problem, and helpful early discussions.

7. REFERENCES

[1] K. R. Rao and P. Yip, "Discrete cosine transform-algorithms, advantages and applications," *Academic Press*, 1990.

[2] A. K. Jain, "A fast karhunen-loeve transform for a class of random processes," *IEEE Trans. on Commn.*, vol. 43, pp. 1023–1029, Sept 1976.

[3] A. K. Jain, "Image coding via a nearest neighbors image model," *IEEE Trans. on Commn.*, vol. 23, pp. 318–331, Mar 1975.

[4] A. Z. Meiri and E. Yudilevich, "A pinned sine transform image coder," *IEEE Trans. on Commn.*, vol. 29, pp. 1728–1735, Dec 1981.

[5] K. Rose, A. Heiman, and I. Dinstein, "Dct/dst alternate-transform image coding," *IEEE Trans. on Commn.*, vol. 38, pp. 94–101, Jan 1990.

[6] B. Zeng and J. Fu, "Directional discrete cosine transforms, a new framework for image coding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 18, pp. 305–313, Mar 2008.

[7] 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," *SPIE*, vol. 5266, pp. 129–137, Oct 2003.

[8] J. Xu, F. Wu, and W. Zhang, "Intra-predictive transforms for block-based image coding," *IEEE Trans. on Signal Processing*, vol. 57, pp. 3030–3040, Aug 2009.

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