Transforms for intra prediction residuals based on prediction inaccuracy modeling

Citation: Cai, Xun, and Jae S. Lim. "Transforms for Intra Prediction Residuals Based on Prediction Inaccuracy Modeling." 2015 IEEE International Conference on Image Processing (ICIP), 27-30 September, 2015, Quebec City, Canada, IEEE, 2015. pp. 4401–4405.

As Published: http://dx.doi.org/10.1109/ICIP.2015.7351638

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Persistent URL: http://hdl.handle.net/1721.1/108261

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Transforms for Intra Prediction Residuals Based on Prediction Inaccuracy Modeling

Xun Cai *Student Member, IEEE,* and Jae S. Lim, *Fellow, IEEE*

***Abstract*—In intra video coding and image coding, the directional intra prediction is used to reduce spatial redundancy. Intra prediction residuals are encoded with transforms. In this paper, we develop transforms for directional intra prediction residuals. Specifically, we observe that the directional intra prediction is most effective in smooth regions and edges with a particular direction. In the ideal case, edges can be predicted fairly accurately with an accurate prediction direction. In practice, an accurate prediction direction is hard to obtain. Based on the inaccuracy of prediction direction that arises in the design of many practical video coding systems, we can estimate the residual covariance and propose a class of transforms based on the estimated covariance function. The proposed method is evaluated by the energy compaction property. Experimental results show that with the proposed method, the same amount of energy in directional intra prediction residuals can be preserved with a significantly smaller number of transform coefficients.**

***Index Terms*—Intra Coding, Image Coding, Intra Prediction Residuals, Transform, Karhunen Loève Transform**

I. Introduction

In transform-based image and video coding, transforms are applied to images and prediction residuals, and the transform coefficients are encoded. With a proper choice of the transform, a large amount of energy can be preserved with a small number of large transform coefficients. This is known as the energy compaction property of transforms [1], [2]. A better energy compaction allows the image and video signal to be encoded with fewer coefficients, while preserving a certain level of image quality.

It is well known that for a random signal with a known covariance function, the linear transform with the best energy compaction property is the Karhunen Loève transform (KLT) [3]. The KLT of typical images has been investigated both theoretically and empirically. It has been noted that the KLT basis functions of typical images are close to the two-dimensional discrete cosine transform (2D-DCT) [4]. The 2D-DCT is also the KLT of a random process characterized by the first-order Markov model for images. As a reasonable approximation to the KLT for images, the 2D-DCT is extensively used in many image and video coding systems [5]–[10].

In various image and video coding systems, prediction is used to reduce the correlation. The prediction residuals, rather than image intensities, are encoded by transforms. To compute the optimal transform for residual signals, it is necessary to obtain the covariance function. A substantial amount of effort has been spent on modeling and estimating covariance functions for prediction residuals. Successful systems with transforms that consider the characteristics of residual signals have been developed. They will be reviewed in Section II.

In this paper, we develop a new class of transforms for directional intra prediction residuals. Specifically, we observe that the directional intra prediction is most effective for edges with clear directionality in typical images. In the ideal case, edges can be predicted fairly accurately if an accurate prediction direction is used. In practice, an accurate prediction direction is hard to obtain. Based on the inaccuracy of prediction direction, we estimate the residual covariance as a function of the coded boundary gradient. We propose to use the KLT of the estimated residual covariance.

This paper is organized as follows. In Section II, we review previous research on transform design for image and video coding systems. In Section III, we discuss our proposed method. We observe that directional intra prediction residuals display non-stationary characteristics. These non-stationary characteristics are modeled by prediction inaccuracy in the proposed model. We derive the model for the horizontal prediction direction and extend it to arbitrary prediction directions. In Section IV, we show experimental results of the proposed method based on the energy compaction property. The proposed transforms are used in addition to the DCT or the asymmetrical discrete sine transform (ADST). A considerable amount of saving in the number of transform coefficients is observed with the hybrid transform. In Section V, we conclude the paper.

II. Previous Research

In image and video compression, transforms are used to reduce the spatial correlation in images and prediction residuals. Transforms are designed primarily by covariance modeling and covariance estimation. In the first approach, the transforms are based on covariance modeling. In this covariance modeling approach, the signals of interest are represented with a model. The model results in a covariance function that is used to obtain the KLT. In the work reported in [11], typical images are represented with a first-order auto-regressive Markov model. It is shown that the KLT basis functions of this model are close to the DCT when the pixels are highly correlated. This model is a reasonable approximation for typical image signals, particularly in a local region. The 2D-DCT is extensively used in many image and video coding systems.

Xun Cai and Jae S. Lim are with the Research Laboratory of Electronics, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 02139 USA. E-mail: cx2001@mit.edu, jslim@mit.edu.

This paper is an extension of the conference paper "Transforms for Intra Prediction Residuals Based on Prediction Inaccuracy Modeling", submitted to and recently accepted by IEEE ICIP 2015

Manuscript received May 13, 2015;
Manuscript revised Aug 7, 2015;
Manuscript accepted Sep 1, 2015;

In the recent work reported in [12], the covariance function for intra prediction residuals is investigated. Based on the observation that pixels in a block can be predicted more accurately when they are closer to the boundary, a first-order Markov model with the deterministic boundary is proposed. This model results in the ADST. The ADST shows a significant performance improvement over the DCT for directional intra prediction residuals. As a result, it is used as an alternative to the DCT to encode intra prediction residuals in the HEVC system [8], [13].

Models for motion-compensated residuals have been investigated. In the work reported in [14]–[16], it is observed that many one-dimensional anisotropic structures arise in motion-compensated residuals. By modeling the motion-compensated residuals with one-dimensional first-order Markov models along a certain direction, a set of one-dimensional transforms of many directions is derived for encoding motion-compensated residuals. A significant coding gain was reported with directional 1D transforms. Similar ideas have been applied to lifting wavelet transforms and disparity-compensated residuals [17], [18].

In the work reported in [19]–[21], motion-compensated residuals are modeled as stationary random processes. As opposed to the first-order Markov model, different covariance functions are proposed to account for the notion that the motion-compensated residuals are less correlated than still images.

The second approach is based on covariance estimation from video data. The covariance estimation process can be performed either through an offline process or on-the-fly during the encoding and decoding processes. In the methods based on offline covariance estimation, the covariance function is computed by analyzing a set of typical video sequences in an offline process. A set of signals that shares similar statistics is used to compute the empirical covariance function, and this covariance function is used to compute an empirical KLT. The KLT is used in the video coding system. Since the transform is computed offline, it does not change throughout the encoding and decoding processes. A variety of transforms based on this approach have been proposed. In the work reported in [22], intra prediction residuals from the same prediction mode are grouped to estimate the covariance function for that mode. Based on the covariance function for each mode, a set of mode-dependent transforms is proposed. In [23], multiple transforms are proposed for each intra prediction mode. To group the residual signals that share similar characteristics, a method based on the K-Singular Value Decomposition (K-SVD) [24] is used. In addition, multiple transforms for motion-compensated residuals have also been investigated in [23].

In the methods based on online covariance estimation, the covariance function is estimated during the encoding and decoding processes from encoded video data. The KLT from the estimated covariance is then obtained. The estimation process may choose to use different portions of encoded information for better adaptivity. As a result, transforms based on online covariance estimation are usually adaptive. We note that the coded information is known to both the encoder and the decoder. In addition, the covariance estimation and the KLT computation rules are synchronized at the encoder and the decoder. As a result, the transmission of the transform basis functions is usually not necessary. Many transforms in this approach have been proposed. In the work reported in [25], it is observed that the statistics of intra residual signals depend on the template in the encoded region for the current block. The covariance function is estimated from similar patches with matching templates in certain regions of encoded video data. In the work reported in [26], it is observed that the covariance function of the motion-compensated residuals can be obtained from the reference block used in motion-compensation. In many cases, the motion-compensated residuals arise due to a slight amount of translation and rotation of the displaced reference block. Based on these assumptions, a set of simulated residual blocks are generated from the reference block. The covariance function is estimated from the simulated residual blocks for encoding the current block. In the work reported in [27], the second-order statistics of the residual signals are investigated. It is observed that a strong correlation exists between the residual frame and the gradient information in the reference frame. A non-linear relationship between the residual variance and the gradient magnitude is obtained and transmitted. To encode the current residual block, the KLT is obtained from a covariance function

$$\gamma(n,m) = \rho^{|n-m|}\sigma_n\sigma_m$$

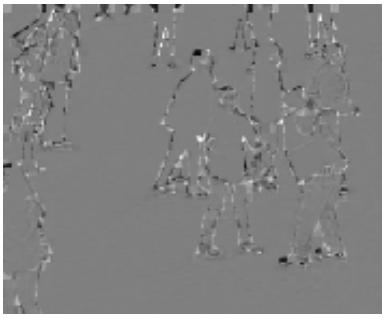
The optimal ρ is estimated and transmitted along with the non-linear function. The variance parameters σ_n^2 and σ_m^2 are estimated from the gradient of the encoded reference frame on a pixel-by-pixel basis.

In the methods discussed above, it is important to adapt the transforms to the characteristics of the signals to be encoded. This is achieved primarily with two approaches. In the first approach, an adaptive transform is chosen from a predefined set of transforms. In this case, designing a reasonable set of transforms may become a difficult task. In the second approach, the statistics is obtained directly from coded data. In this case, it may be hard to ensure robust estimation based on a limited number of available samples. These observations motivate the modeling in the method that we propose in this paper. In the proposed method, the process that generates residual signals is first studied. A model that summarizes the residual generation process is proposed. The proposed model allows a more robust estimation of the covariance function only from a small number of coded pixels. An estimated covariance function is proposed based on the proposed model. It is adaptive to the content to be encoded. This approach is discussed in detail in the following sections.

III. Proposed Method

In this section, we describe the proposed transforms for directional intra prediction residuals. First, we discuss the characteristics of directional intra prediction residuals based on empirical observations in Section III-A. In Section III-B, we discuss the model that characterizes these empirical observations. Specifically, we model the directional intra prediction as the result of prediction inaccuracy. From this model, we

An intra frame



Intra prediction residual

Fig. 1: Intra frame and intra prediction residual

can estimate the residual covariance based on the gradient of the coded boundary. The proposed model is first discussed in the horizontal prediction and extended to arbitrary directions. The statistics of the proposed method are analyzed in Section III-C. The KLT of the covariance function is proposed in Section III-D. Finally, we discuss the gradient computation on a discrete sampling grid in Section III-E.

A. Characteristics of directional intra prediction residuals

The characteristics of intra prediction residuals are significantly different from those of still images. Figure 1 shows an example of a still frame and its intra prediction residual.[1] For a typical still image, we observe that image intensities tend to be stationary in most smooth regions of the image. For the intra prediction residuals, we observe that most regions are close to zero, as a consequence of the effective intra prediction in smooth regions. In the regions where sharp edges and busy textures arise, the intra prediction becomes less effective, and the residuals become much larger in these regions.

To carefully investigate the characteristics of directional intra prediction residuals on a block-by-block basis, we show a 4x4 block of intra prediction residual in Figure 2. In this 4x4 block, the vertical prediction is used. We note that intensities of the directional prediction residuals tend to increase along the prediction direction, as the distance from the boundary increases. This observation is typical in many video sequences. It has been investigated in previous research, such as the work reported in [12]. In addition, we note that the residual signal along the direction orthogonal to the prediction direction displays significantly different characteristics. Specifically, the residual intensities change abruptly along the direction orthogonal to the prediction direction, as shown in Figure 2. This observation indicates that the residual signal may be highly non-stationary in the direction orthogonal to the prediction direction. The characteristics of the prediction residuals are very sensitive, not only to the prediction direction, but to the local change of the image data as well. In other words, the characteristics of the prediction residuals should not only be mode-dependent, but also data-dependent.

[1]The residual frame is shown with an offset of 128, to show the negative values.

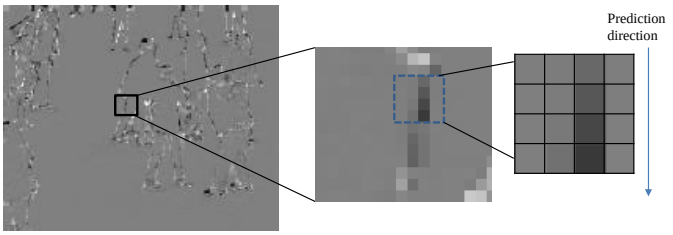


Fig. 2: Illustration of the derivation

The non-stationarity of residual signals can be interpreted by the prediction accuracy. In those regions where there are sharp discontinuities in the original frame, the prediction tends to be less accurate. Therefore, the residual intensities tend to be large relative to smooth regions. We wish to use this observation to predict the statistics of the residuals. This observation will be useful only when we can relate the local change of image data to coded data. We note that it is in general not possible to estimate the statistics of the residual signal from the same region that is yet to be encoded.

To estimate the residual statistics only from the coded data, we consider the process of directional intra prediction. Specifically, we consider the sensitivity of prediction to the accuracy of the prediction direction. In a smooth region where pixels share similar intensities, the prediction accuracy is less sensitive to the prediction direction. On the other hand, in the regions where sharp discontinuities exist in the original frame, the prediction is very sensitive to the accuracy of the prediction direction. A small disturbance of the prediction direction away from the actual direction may lead to a large prediction error. This observation leads to a model that estimates the residual covariance only from the coded boundary. In the following subsections, we discuss the model in detail.

B. Prediction inaccuracy modeling

In this section, we discuss the proposed model for directional intra prediction residuals. Specifically, we relate the residual intensities to the prediction inaccuracy and boundary

gradient. We first derive a simplified model for the horizontal prediction to illustrate the idea. We then extend the simplified model to arbitrary prediction directions.

1) Model for horizontal prediction: We first establish the notations for the proposed model. We consider a rectangular block to be encoded, and we use the following notations:

- $f(m,n)$: current block to be encoded
- $\hat{f}(m,n)$: predicted block, obtained by copying the coded left boundary $f(0,n)$ along the horizontal direction.
- $r(m,n)$: residual block, obtained by subtracting $\hat{f}(m,n)$, the predicted block, from $f(m,n)$, the current block.

In the above notations, m is the horizontal coordinate, $m=0$ corresponds to the coded left boundary that is used for prediction, n corresponds to the vertical coordinate and $m,n \geq 1$ is the area to be encoded. This is illustrated in Figure 3.

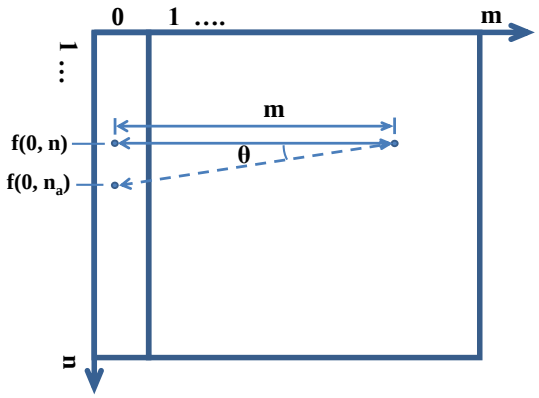


Fig. 3: Illustration of the derivation: horizontal

We first note that the residual is obtained by subtracting the prediction from the current block:

$$r(m,n) = f(m,n) - \hat{f}(m,n) \tag{1}$$

The prediction is obtained by horizontal prediction:

$$\hat{f}(m,n) = f(0,n) \tag{2}$$

In addition, we assume that the accurate prediction direction is characterized by a random variable $\theta(m,n)$ taking small values. This θ can be assumed, for example, uniformly distributed in all directions near the horizontal direction. Suppose we denote n_{a} as the location of the accurate prediction in the coded boundary. Ignoring the difference between the intensities of the current pixel and the perfect prediction, we obtain:

$$f(m,n) \approx f(0,n_{\mathrm{a}}) \tag{3}$$

where

$$n_{\mathrm{a}} = n + m\tan(\theta(m,n)) \approx n + m\theta(m,n) \tag{4}$$

for small θ. This can be seen from the geometry shown in Figure 3.

From equations (1), (2), (3) and (4), we obtain:

$$\begin{aligned} r(m,n) &= f(m,n) - \hat{f}(m,n) \approx f(0,n_{\mathrm{a}}) - f(0,n) \\ &\approx (n_{\mathrm{a}} - n)\frac{\partial f(0,n)}{\partial n} \approx m\theta(m,n)\frac{\partial f(0,n)}{\partial n} \end{aligned} \tag{5}$$

for small θ and therefore small $n_a - n$.

Equation (5) indicates that the residual intensity is proportional to the distance m and to the boundary gradient. In addition, the residual intensity depends on how inaccurate the prediction direction is away from the actual direction, characterized by a random variable θ.

2) Model for arbitrary prediction directions: For an arbitrary prediction direction, the same idea from the horizontal prediction applies. We can model the residual signal as a function of boundary gradient and the prediction inaccuracy. The geometry is slightly more involved. In Figure 4, we illustrate the geometry for the derivation of the model. We derive the model when the left boundary is used in prediction. The upper boundary case can be derived by symmetry.

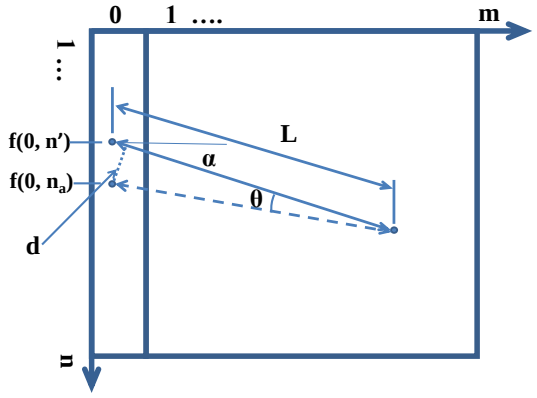


Fig. 4: Illustration of the derivation: arbitrary

When an arbitrary prediction is used, the current pixel is predicted from a pixel with a different boundary coordinate. Therefore, n is replaced by n'. In addition, the displacement from the accurate predictor to the predictor used is related to θ in a different way. Consider the geometry shown in Figure 4. The arc length resulting from the inaccurate prediction direction is $d \approx L\theta$. In this relation, L is the distance from the residual pixel to its boundary predictor. From the geometry shown in Figure 4, the displacement becomes $n_a - n \approx \frac{d}{\cos\alpha}$, where α is the angle from the prediction direction to the norm of the boundary.

Combining these results, we obtain the following estimation by analogy to the horizontal case:

$$r(m,n) \approx \frac{L}{\cos\alpha}\left.\frac{\partial f(0,n)}{\partial n}\right|_{n'}\theta(m,n) \tag{6}$$

Equation (6) indicates that the residual is proportional to the boundary gradient, evaluated at the position of the predictor. In addition, the residual is proportional to the distance from the current pixel to its boundary predictor scaled by a factor related with the prediction direction. We note that the general case is consistent with the horizontal case. When the horizontal prediction is used, $\alpha = 0$, $L = m$ and Equation (6) reduces

to Equation (5). As another example, when the diagonal prediction is used, $\alpha = \frac{\pi}{4}$. Equation (6) will be used to derive the covariance function for the residual signal in Section III-C.

C. Statistics based on prediction inaccuracy

From Equation (6), the randomness of residual signal in the proposed model originates from the randomness of the prediction inaccuracy θ. This observation implies that we can study the statistics of the residual signal by studying that of prediction inaccuracy. In this section, we study the mean, variance and covariance of the process, characterized by the proposed model.

1) Mean: We first note that $E[\theta(m,n)] = 0$. This is reasonable since the prediction direction inaccuracy would not generally be biased towards any side. This leads to:

$$E[r(m,n)] = \frac{L}{\cos\alpha} \left.\frac{\partial f(0,n)}{\partial n}\right|_{n'} E[\theta(m,n)] = 0 \quad (7)$$

2) Variance: Denote the variance function as $\sigma^2(m,n)$. We take the expectation of r^2, with respect to the random variable θ.

$$\begin{aligned}\sigma^2(m,n) &= E\left[r^2(m,n)\right] \\ &\approx \left[\frac{L}{\cos\alpha}\right]^2 \left[\frac{\partial f(0,n)}{\partial n}\right]^2 \Bigg|_{n'} E[\theta(m,n)^2] \end{aligned} \quad (8)$$

This relationship indicates that the residual variance is proportional to the squared distance and squared boundary gradient. In other words, residual intensity tends to be large where the boundary gradient at the predictor is large. In our model, the boundary gradient is an estimation of the amount of local change along the prediction direction. Therefore, this relationship also indicates that the residual is large when the estimated local change at the same location is large. This is consistent with the intuition discussed in Section III-A.

3) Covariance: Since the random process is zero-mean, Equation (6) and Equation (8) directly lead to the following covariance function:

$$Cov\left[r(m_1,n_1)r(m_2,n_2)\right] = \sigma(m_1,n_1)\sigma(m_2,n_2)R \quad (9)$$

where R is the factor that characterizes the correlation of the prediction inaccuracy, defined as

$$R = \frac{E[\theta(m_1,n_1)\theta(m_2,n_2)]}{\sqrt{E[\theta^2(m_1,n_1)]E[\theta^2(m_2,n_2)]}} \quad (10)$$

The relationship in Equation (9) indicates that the covariance function of the residual signal depends on the estimated residual standard deviation σ and the statistics of the prediction inaccuracy R. Specifically, this equation indicates that the non-stationarity of the residuals is reflected mostly by a drastic change of the residual variance function. By choosing a reasonable R, we can obtain a reasonable residual covariance function.

Since most non-stationarity in the residual covariance function is reflected in a drastic change of the variance function, the prediction inaccuracy is relatively stationary. Therefore, we relate the prediction inaccuracy with the first-order Markov process. This model is extensively used in image processing applications to model stationary processes. To be specific, we choose in this paper

$$E[\theta(m_1,n_1)\theta(m_2,n_2)] = \rho_1^{|m_1-m_2|}\rho_2^{|n_1-n_2|} \quad (11)$$

With the choice of the function in Equation (11), we can see that when $m_1 = m_2$ and $n_1 = n_2$,

$$E[\theta^2(m_1,n_1)] = E[\theta^2(m_2,n_2)] = 1 \quad (12)$$

.

With Equations (10), (11) and (12),

$$R = \rho_1^{|m_1-m_2|}\rho_2^{|n_1-n_2|} \quad (13)$$

Therefore, the residual covariance function is:

$$\begin{aligned} &Cov\left[r(m_1,n_1)r(m_2,n_2)\right] \\ &= \sigma(m_1,n_1)\sigma(m_2,n_2)\rho_1^{|m_1-m_2|}\rho_2^{|n_1-n_2|} \end{aligned} \quad (14)$$

We note that the covariance function with the same form is proposed in [27] for encoding motion-compensated residuals.

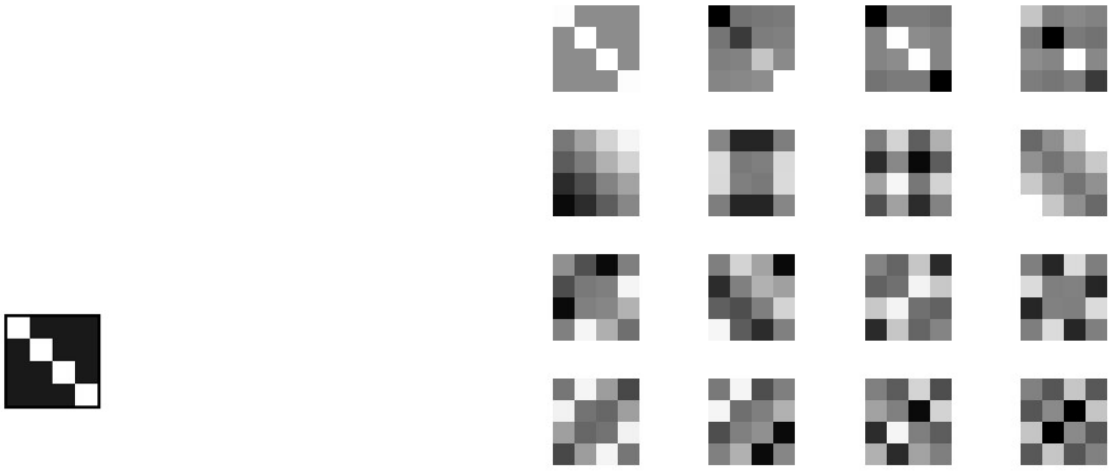
D. Transforms based on the proposed covariance function

From the covariance function in Equation (14), we would like to compute the KLT basis functions. The KLT is used to encode the current residual block. In general, it is very difficult to obtain a closed-form solution of the transform basis functions based on the proposed covariance function. To study the characteristics of the transform basis functions, we consider two examples.

We first consider a simplified 1-D example. Suppose that a zero-mean signal is denoted as $x(n)$, where $0 \le n \le 3$. The variance of this signal is given by $\sigma^2(0) = \sigma^2(1) = 0$ and $\sigma^2(2) = \sigma^2(3) = 1$. A typical transform that ignores the variance information, such as the DCT, will in general result in transform coefficients of length 4. However, if the given variance information is considered, we can easily see that $x(0)$ and $x(1)$ are almost surely to be zero. Therefore, the covariance function proposed in Equation (14) will result in a transform with the first two basis functions supported only on $x(2)$ and $x(3)$. This leads to significant transform coefficients of length at most 2. In other words, by considering the variance information, we are effectively adapting the transform to the non-stationarity of the signal. Therefore, the resulting transform tends to achieve much better energy compaction in this example.

As another example, we consider the example shown in Figure 5.[2] In this example, we show the variance function in a 4x4 block on the left side. The variance of the brighter pixels is 0.9 while the variance of the darker pixels is 0.1. This variance function is used to construct a covariance function in Equation (14), with $\rho_1 = \rho_2 = 0.99$. On the right side, we show the KLT basis functions from this covariance function. From this figure, we observe that the region of support for the first several basis functions is mostly within the region

[2]The transform basis functions are shown with an offset 0.5 to illustrate negative values.



Variance Block Transform basis functions

Fig. 5: An example of the proposed transform basis functions

where the variance is large. This observation indicates that the proposed transform is adapted well to the non-stationarity of the signal. Specifically, the proposed transform first considers encoding the pixels with large intensities and compresses most of their energy into a small number of transform coefficients.

Finally, we note that the covariance function is estimated only from the coded boundaries. Therefore, the same covariance function can be estimated both at the encoder and the decoder. We do not have to transmit any side information associated with the transform coefficients.

To summarize, the proposed method consists of the following steps:

Step A: For each pixel in the current block, estimate the variance function according to Equation (8).

Step B: Using the variance function in Step A, construct the covariance function according to Equation (14).

Step C: Compute the KLT of the covariance function in Step B. Use this KLT to encode the current block.

E. Gradient computation on a discrete grid

In the proposed method, we derive the residual covariance as a function of the boundary gradient. In an ideal situation, the boundary gradient at any given location can be computed, if coded boundary samples are dense enough. In practice, the density of available samples is limited by the density of the sampling grid. This limitation requires the boundary gradient to be estimated from a small number of boundary pixels. In this section, we discuss the gradient computation on a discrete sampling grid.

Consider estimating the variance function in Equation (8). In this equation, the boundary gradient is evaluated at location n'. The value of n' can be computed from the location of the current pixel and the given prediction direction. The geometry is shown in Figure 4.[3] While the coordinates of the current pixel are always integers, n' may not necessarily be an integer. To compute the gradient for different possible values of n', we consider three typical cases.

1) n' is a positive integer: A positive integer n' implies that we are interested in evaluating the gradient on the sampling grid. In this case, we consider estimating the gradient from three reference samples. Suppose we predict from the left boundary and we consider the block shown in Figure 6. To evaluate the gradient at location $(0, n')$, we can estimate the gradient as either $f(0, n') - f(0, n' - 1)$ or $f(0, n' + 1) - f(0, n')$. From the proposed model, the prediction inaccuracy is not biased towards the positive or the negative side of n'. Therefore, the contribution of two estimations is likely to be equal. Since the variance is proportional to the square of the gradient, we can estimate the square of the gradient effectively as the mean square of two estimations. In other words, when n' is a positive integer:

$$\left[\frac{\partial f(0,n)}{\partial n}\right]^2\Bigg|_{n'} = \frac{1}{2}\left[f(0,n') - f(0,n'-1)\right]^2 + \frac{1}{2}\left[f(0,n'+1) - f(0,n')\right]^2 \quad (15)$$

2) n' is not a integer: When n' is not an integer, we wish to evaluate the gradient in between two boundary pixels $f(0, \lceil n' \rceil)$ and $f(0, \lfloor n' \rfloor)$. This is illustrated in Figure 7. In this case, the squared gradient is simply given by:

$$\left[\frac{\partial f(0,n)}{\partial n}\right]^2\Bigg|_{n'} = \left[f(0,\lceil n' \rceil) - f(0,\lfloor n' \rfloor)\right]^2 \quad (16)$$

[3]We discuss the case when the left boundary is used, the upper boundary case can be generalized by symmetry.

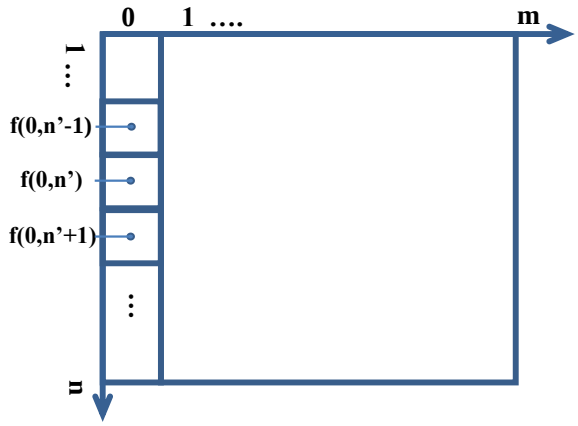


Fig. 6: Gradient computation for positive integer n'

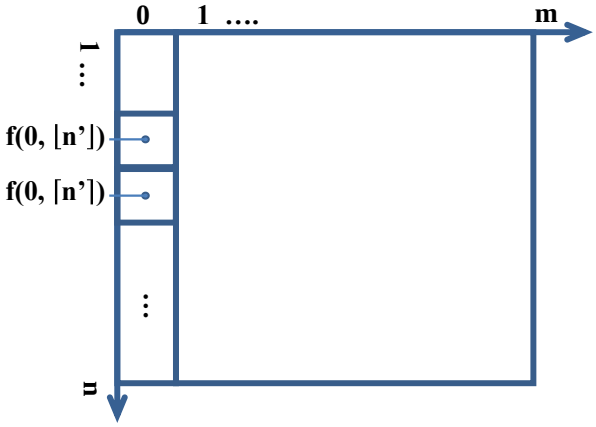


Fig. 7: Gradient computation for non-integer n'

3) n' is zero (the corner predictor is used): In the case when non horizontal/vertical prediction is chosen, the upper left corner predictor is used when $n' = 0$. This is shown in Figure 8. In this case, the gradient can be estimated as $f(1,0) - f(0,0)$ when the accurate prediction is from the upper boundary. On the other hand, the gradient can be estimated as $f(0,1) - f(0,0)$ when the accurate prediction comes from the left boundary. Both cases are equally likely to happen. As in the case when n' is a non-zero integer, we wish to estimate the gradient by averaging two cases.

In Equation (8), the variance is scaled by a factor related with the prediction angle α. The prediction angle is fixed when only one boundary is used. In the case when $n' = 0$, both the upper boundary and the left boundary are involved in the gradient computation. The prediction angle is different for the upper boundary and for the left boundary. Therefore, we chose to directly estimate the variance in this case. The variance is estimated as:

$$\sigma^2(m,n) = \frac{1}{2}\left[\frac{L}{\cos\alpha_U}\right]^2 (f(1,0) - f(0,0))^2 + \frac{1}{2}\left[\frac{L}{\cos\alpha_L}\right]^2 (f(0,1) - f(0,0))^2 \quad (17)$$

where α_U is the prediction angle from the upper boundary and α_L is the prediction angle from the left boundary.

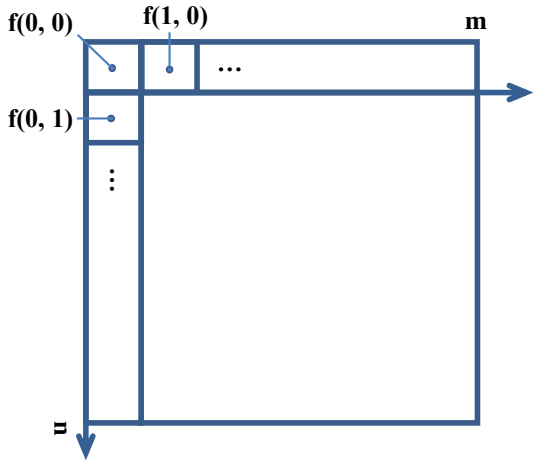


Fig. 8: Gradient computation for $n' = 0$

IV. Experimental Results

In this section, we investigate the performance of the proposed method. We discuss the experimental setup in Section IV-A. We then show that the proposed method can effectively estimate the residual statistics that reflect the characteristics of the residual signals in Section IV-B. Then we investigate the energy compaction property of the proposed method for the rest of this section.

A. Experimental Setup

In the experiments that we perform, we obtain the directional intra prediction residuals according to the H.264 prediction. The block size is fixed to 4x4 and all prediction modes are used. Original samples are used to construct the directional intra predictors and estimate the covariance function. The effect of quantized boundary predictors used in practice will be discussed in Section IV-G. For the proposed transforms, we estimate the covariance function as discussed in Section III. The parameter ρ is chosen to be 0.99.[4] In the covariance estimation process, the coded boundary gradient may become zero in boundary and smooth regions. For these cases, we use the DCT or the ADST instead.

The energy compaction property of the proposed transforms is investigated. Specifically, we use the proposed transforms in hybrid with the DCT or the ADST [12]. We compare the energy compaction of the hybrid transform to the DCT or the ADST. We compute the preserved energy given the total number of chosen coefficients. Transform coefficients with largest magnitudes within a frame are chosen. In the case of the hybrid transform, the transforms and transform coefficients are selected, for each block, utilizing the algorithms proposed in [28], [29]. We plot the preserved energy as a function of the total number of chosen coefficients. The preserved energy

[4]We note that in our experiments, changing ρ within a reasonable range does not significantly affect the results.

Estimated variance

Magnitude of residuals

Fig. 9: Comparison between the estimated variance and residual signal

is in terms of the percentage relative to the total energy. The total number of chosen coefficients is presented in terms of the percentage relative to the total number of coefficients. A larger preserved energy value at the same percentage of chosen coefficients indicates a higher performance in energy preservation. It is evident in [28], [29] that the energy compaction capability is a useful measure of performance in coding applications.

B. Variance Estimation

In the proposed method, the non-stationarity of residual signals is reflected by the local change of the estimated variance function. An accurate estimation of the residual variance would result in more compact transform coefficients. The ideal estimated variance function should take large values precisely where the residual is large. On the other hand, transforms that do not consider the non-stationarity of residual signals, such as the DCT, implicitly assume a uniform variance function. The performance of the transform heavily depends on the consistency between the estimated variance function and the residual signal.

Figure 9 shows the estimated variance function and the magnitude of the residual signal, from the intra frame of the sequence "ice_qcif". We first observe that the estimated variance is visually consistent with the magnitude of the residual signal.

To quantify the consistency between the magnitude of the residual signal and the estimated variance function, we study the cumulative energy of the residual signal. In Figure 10, we show three cumulative energy curves. In the optimal cumulative energy, we rank order the residual magnitude and compute the cumulative energy from the largest residual pixels. In the cumulative energy from the estimated variance, we rank order the estimated variance and compute the cumulative energy from pixels with the largest estimated variance. In the randomized cumulative energy, we compute the cumulative energy from a randomly chosen set of pixels. The cumulative energy indicates how informative the estimated variance is in preserving the residual energy.

In the ideal case, suppose the estimated variance is very accurate and precisely reflects the rank order information of the residual magnitude. If we choose the residual pixels from the largest estimated variance, the preserved energy as a function of the number of preserved pixels is the largest. It is represented by the optimal cumulative energy. On the other hand, suppose the estimated variance is not related to the residual magnitude. In this case, the cumulative energy will be close to a randomized cumulative energy. The cumulative energy from the estimated variance in practice should lie between these two extremes. From Figure 10, we see that the cumulative energy from the estimated variance in practice is close to the optimal cumulative energy. This suggests that the estimated variance is correlated with the residual pixel magnitude. This observation implies that the estimated variance is informative in predicting the magnitude of the residual signals on a pixel-by-pixel basis. In other words, the estimated variance from the prediction inaccuracy model can estimate the non-stationarity of the residual signal. In a practical video coding system, residual signals are usually encoded with transforms. The estimated variance function can be used to design more effective transforms for the residual blocks.

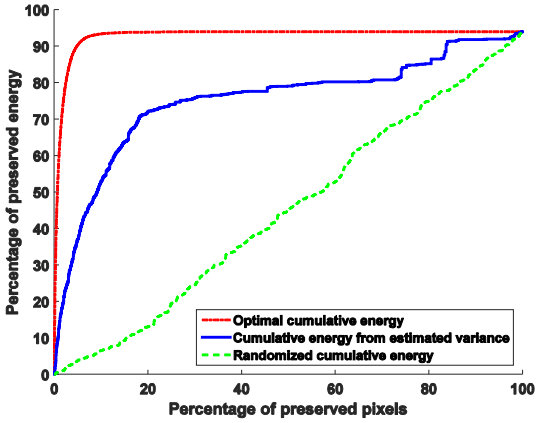


Fig. 10: Cumulative energy functions

For some blocks, the estimated variance function may not be consistent with the magnitude of the residual signals. An inconsistent estimation of the residual magnitude will significantly degrade the performance. In our experiments, we observe that replacing the DCT with the proposed KLT for every block only slightly improves the energy compaction

performance on average. Replacing the ADST with the proposed KLT for every block slightly degrades the performance on average. In addition, the performance of using only the KLT varies significantly for different sequences. For a more robust performance, therefore, we use the proposed method in hybrid with other robust transforms. For the rest of the paper, we consider using the proposed transforms in hybrid with the DCT or the ADST on a block-by-block basis.

C. Results: hybrid with the DCT

In this section, we compare the energy compaction performance of two transform settings: 1) The 2D-DCT. 2) The proposed transform in hybrid with the 2D-DCT.

Figure 11 shows the energy compaction performance of the hybrid transform and the DCT, for the intra frame in the sequence "carphone_QCIF". From this figure, we see that the same amount of energy can be preserved with a significantly smaller number of transform coefficients by using the KLT in addition to the DCT. The percentage of coefficient saving is summarized in Table I.

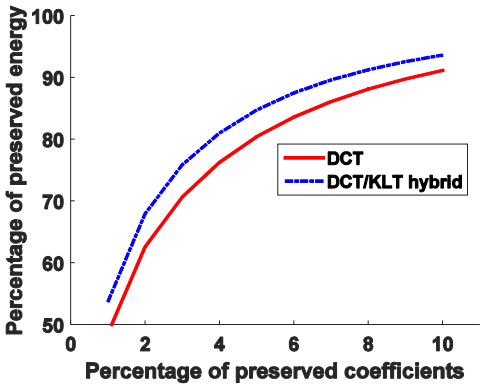


Fig. 11: Energy compaction performance of the hybrid transform (vs DCT). Sequence: Carphone_QCIF

The same experiment was repeated for around fifty test sequences with QCIF and CIF resolutions from the "derf" set [30]. The average coefficient savings relative to the DCT is summarized in Table II. In this table, the percentage of coefficient saving in the second row is measured when the same amount of energy is preserved by two transforms. The percentage of preserved coefficients for the DCT is specified in the first row.

In addition, we investigate the frequency of choosing the KLT in our experiments. For the sequences in our tests, the KLT is chosen for 45.67% of the non-zero blocks, when around 5% of the coefficients is preserved. In these blocks, the proposed transform is more effective than the DCT.

In another experiment, we show the scatter plot between the frequency of choosing the KLT and the coefficient saving. This is shown in Figure 12. In this figure, each circle represents a test sequence. From this figure, a positive correlation appears between the frequency of choosing the KLT and the coefficient saving. This indicates that more coefficients tend to be saved when the proposed model is effective in capturing more non-stationarity of the residual signal. This is consistent with the insights of the proposed method.

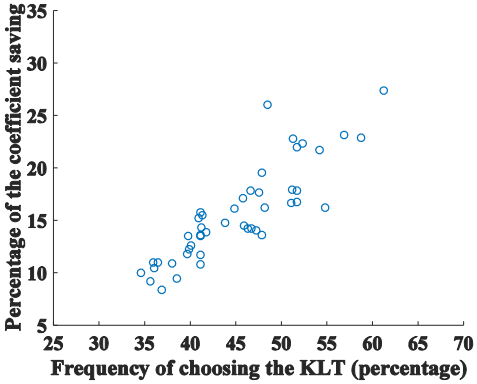


Fig. 12: Coefficient saving to the frequency of choosing the KLT

D. Results: hybrid with the ADST

In this section, we compare the energy compaction performance of two transform settings: 1) The ADST. 2) The proposed transform in hybrid with the ADST.

Figure 13 shows the energy compaction performance of two transform settings, for the intra frame in the sequence "carphone_QCIF". Similar to the case of the DCT, we observe a significant amount of coefficient saving when the proposed method is used in addition to the ADST. The percentage of coefficient saving is summarized in Table III.

The same experiment was repeated for the same sequence set. The average coefficient savings relative to the ADST is summarized in Table IV.

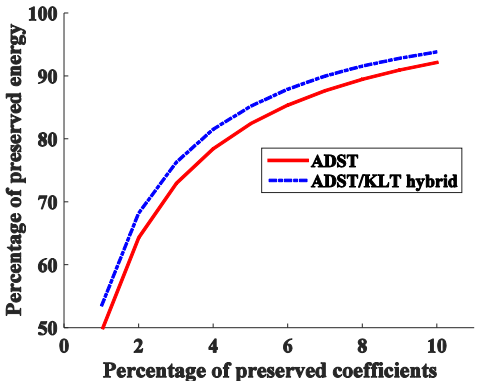


Fig. 13: Energy compaction performance of the hybrid transform (vs ADST). Sequence: Carphone_QCIF

We also investigate the frequency of choosing the KLT vs the ADST. For the sequences in our tests, the KLT is chosen for 38.71% of the non-zero blocks, when around 5% of the coefficients is preserved. Compared to the case when the KLT is used in hybrid with the DCT, the frequency of choosing the

Percentage of preserved energy (roughly)	60%	70%	80%	90%
Percentage of coefficient saving	18.6%	21.9%	23.1%	23.5%

TABLE I: Coefficient saving of the hybrid transform relative to the DCT. Sequence: Carphone_QCIF

Percentage of preserved coefficients (roughly)	3%	5%	7%	9%
Percentage of coefficient saving	15.6%	16.1%	16.0%	15.4%

TABLE II: Average coefficient saving of the hybrid transform relative to the DCT.

Percentage of preserved energy (roughly)	60%	70%	80%	90%
Percentage of coefficient saving	13.1%	13.9%	15.1%	16.4%

TABLE III: Coefficient saving of the hybrid transform relative to the ADST. Sequence: Carphone_QCIF

Percentage of preserved coefficients (roughly)	3%	5%	7%	9%
Percentage of coefficient saving	12.8%	12.7%	12.8%	12.6%

TABLE IV: Average coefficient saving of the hybrid transform relative to the ADST.

KLT is slightly smaller. Still, the frequency of choosing the KLT is significant. This indicates that the proposed transforms are still effective when the ADST replaces the DCT.

The same scatter plot between the frequency of choosing the KLT and the coefficient saving is shown for the case of the ADST in Figure 14. A similar positive correlation appears between the two factors.

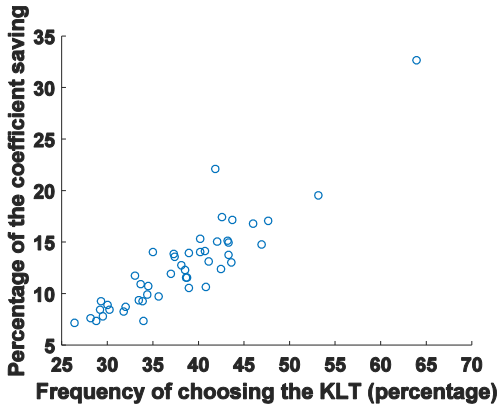


Fig. 14: Coefficient saving to the frequency of choosing the KLT

E. Summary of the energy compaction performance

In this section, we summarize the energy compaction performance of the proposed transform. Specifically, we compare the energy compaction performance of the four transforms discussed in previous sections. They are 1) DCT, 2) ADST, 3) KLT hybrid with DCT and 4) KLT hybrid with DST. We measure the performance in terms of the percentage of coefficients used to preserve the same amount of energy relative to the DCT. The coefficient saving is measured when the same energy is preserved with 5% DCT coefficients, averaged over the sequences that we tested. The result is shown in Figure 15.

From the figure, we see that the DCT on average results in the worst performance. Replacing the DCT with the ADST slightly improves the performance as expected. When the KLT is used in addition to either the DCT or the ADST, the performance significantly improves. This is because the prediction inaccuracy model is effective in many typical residual blocks. The covariance estimated from this model captures the non-stationarity of residual signals that neither the DCT nor the ADST can capture. In fact, when the KLT is used, whether it is hybrid with DCT or the ADST only makes a small difference. This implies that much non-stationarity in the residual signals is captured by the proposed KLT. The remaining stationary blocks can be encoded with a reasonable stationary transform and the choice of such transform is not as important.

F. Effectiveness of the proposed KLT

In our experiments, the KLT is used in hybrid with the ADST or the DCT. From the choice of the transform in each block, we can investigate how effective the KLT is for a typical frame.

Figure 16 shows the choice of transform for an intra frame "ice_CIF" when the KLT is used in hybrid with the ADST. We show the intra frame, blocks that select the KLT and blocks that select the ADST. [5] We first observe that the KLT is chosen in those regions with sharp discontinuities and directional structures. We can visually reconstruct the contours of the objects from those blocks that choose the KLT. The proposed model is expected to be effective in these regions. On the other hand, the KLT is less effective in the regions where the ADST is chosen. From the figure, we see that these blocks are

[5]For the blocks with very small residual energy, all transform coefficients are below the threshold and they are treated as zero blocks. We only show the intra blocks that have significant residual transform coefficients

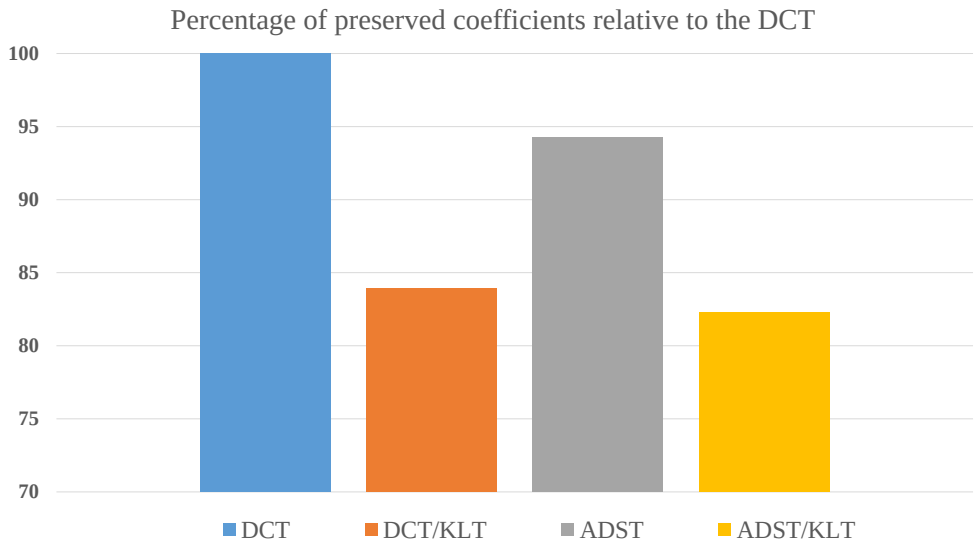


Fig. 15: Percentage of preserved coefficients relative to the DCT

distributed more randomly than the KLT blocks. Most of these blocks are relatively stationary, have complicated textures or discontinuities with less regular directionality. We observe similar patterns for other intra frames in our experiments.

G. Other comments

The proposed method is currently evaluated by the energy compaction property. When the proposed method is implemented in a video coding system, some practical issues arise.

First, we note that in a video coding system with hybrid transforms, we need to transmit 1-bit side information for each non-zero block, to indicate which transform to use. It is evident in the work reported in [14], [15], [17], [28], [29] that this small overhead is not likely to significantly affect the large positive gain from the better energy compaction.

Second, the entropy coding of the significant transform coefficients is ignored in the energy compaction analysis. In a practical video coding system, we may scan the transform coefficients in a specific order and entropy code the transform coefficients. The order of scanning can be determined by the expected magnitude of transform coefficients. This information is available when computing the KLT basis functions from the covariance function.

Third, the covariance function is estimated from coded boundaries in a video coding system. The coded boundaries may be distorted due to quantization in the coded blocks. This distortion may potentially affect the accuracy of the covariance estimation and hence the performance of the transform. To see the performance under the distorted estimation, we repeated the experiments by estimating the covariance function from the distorted boundary information. Specifically, we estimated the covariance function from boundaries of coded frames processed by the H.264 system, under a reasonable range of QP. We did not observe a significant amount of performance degradation in the experiments.

V. Conclusions

In this paper, we propose a class of transforms for directional intra prediction residuals based on prediction inaccuracy modeling. In this method, we first model the process that generates the directional intra prediction residuals. We then derive the covariance function by considering the prediction inaccuracy. The covariance function is estimated as a function of the gradient of coded boundaries. The KLT of the covariance function is used to encode the residual block. The proposed transforms can effectively estimate the residual covariance in many typical intra prediction residual blocks. The proposed transform can save a significant amount of transform coefficients while preserving the same amount of residual energy.

In addition to the proposed transforms for intra prediction residuals, the prediction inaccuracy modeling can be used as a robust estimation method for other transforms. The prediction inaccuracy analysis can be useful when other prediction methods are used. For example, we are interested in investigating transforms based on prediction inaccuracy for motion-compensated residuals, resolution-enhancement residuals and binocular prediction residuals.

References

[1] J. S. Lim, "Two-dimensional signal and image processing," *Englewood Cliffs, NJ, Prentice Hall, 1990, 710 p.*, vol. 1, 1990.

[2] H. Kitajima, "Energy packing efficiency of the Hadamard transform," *Communications, IEEE Transactions on*, vol. 24, no. 11, pp. 1256–1258, 1976.

[3] K. Karhunen, *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*. Universitat Helsinki, 1947, vol. 37.

[4] N. Ahmed, T. Natarajan, and K. Rao, "Discrete cosine transform," *IEEE Trans. Comput.*, vol. 100, no. 1, pp. 90–93, 1974.

[5] G. K. Wallace, "The JPEG still picture compression standard," *Consumer Electronics, IEEE Transactions on*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.

[6] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560–576, 2003.

[7] T. Wiegand, "Draft ITU-T recommendation and final draft international standard of joint video specification," *ITU-T rec. H. 264— ISO/IEC 14496-10 AVC*, 2003.

Intra frame

Blocks that choose the KLT

Blocks that choose the ADST

Fig. 16: Choice of transform for each block

[8] G. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *Circuits and Systems for Video Technology, IEEE Transactions on Vol. 22 Iss. 12*, 2012.

[9] B. Bross, W. Han, G. Sullivan, J. Ohm, and T. Wiegand, "High efficiency video coding (HEVC) text spec. draft 10 (for fdis&consent)," in *JCT-VC Doc. JCTVC-L1003, 12th Meeting: Geneva, Switzerland*, vol. 1, 2013.

[10] J. Bankoski, R. S. Bultje, A. Grange, Q. Gu, J. Han, J. Koleszar, D. Mukherjee, P. Wilkins, and Y. Xu, "Towards a next generation open-source video codec," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2013.

[11] M. Flickner and N. Ahmed, "A derivation for the discrete cosine transform," *Proc. IEEE*, vol. 70, no. 9, pp. 1132–1134, 1982.

[12] J. Han, A. Saxena, V. Melkote, and K. Rose, "Jointly optimized spatial prediction and block transform for video and image coding," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 1874–1884, 2012.

[13] A. Saxena and F. C. Fernandes, "DCT/DST-based transform coding for intra prediction in image/video coding," *Image Processing, IEEE Transactions on*, vol. 22, no. 10, pp. 3974–3981, 2013.

[14] F. Kamisli and J. Lim, "Video compression with 1-D directional transforms in H.264/AVC," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 738–741.

[15] F. Kamisli and J. S. Lim, "1-D transforms for the motion compensation residual," *Image Processing, IEEE Transactions on*, vol. 20, no. 4, pp. 1036–1046, 2011.

[16] H. Zhang and J. Lim, "Analysis of one-dimensional transforms in coding motion compensation prediction residuals for video applications," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, march 2012, pp. 1229 –1232.

[17] F. Kamisli and J. S. Lim, "Directional wavelet transforms for prediction residuals in video coding," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE, 2009, pp. 613–616.

[18] B. Karasoy and F. Kamisli, "Transforms for the disparity-compensated prediction residuals," in *Signal Processing and Communications Applications Conference (SIU), 2014 22nd*. IEEE, 2014, pp. 168–171.

[19] W. Niehsen and M. Brunig, "Covariance analysis of motion-compensated frame differences," *IEEE transactions on circuits and systems for video technology*, vol. 9, no. 4, pp. 536–539, 1999.

[20] K.-C. Hui and W.-C. Siu, "Extended analysis of motion-compensated frame difference for block-based motion prediction error," *Image Processing, IEEE Transactions on*, vol. 16, no. 5, pp. 1232–1245, 2007.

[21] C.-F. Chen and K. K. Pang, "The optimal transform of motion-compensated frame difference images in a hybrid coder," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 40, no. 6, pp. 393–397, 1993.

[22] Y. Ye and M. Karczewicz, "Improved H.264 intra coding based on bi-directional intra prediction, directional transform, and adaptive coefficient scanning," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008, pp. 2116–2119.

[23] X. Zhao, L. Zhang, S. Ma, and W. Gao, "Video coding with rate-distortion optimized transform," *IEEE Trans. Circuits Syst. Video Technol*, vol. 22, no. 1, pp. 138–151, 2012.

[24] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311–4322, 2006.

[25] C. Lan, J. Xu, G. Shi, and F. Wu, "Exploiting non-local correlation via signal-dependent transform (SDT)," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 7, pp. 1298–1308, 2011.

[26] M. Wang, K. N. Ngan, and L. Xu, "Efficient H.264/AVC video coding with adaptive transforms," *IEEE TRANSACTIONS ON MULTIMEDIA*, vol. 16, no. 4, p. 933, 2014.

[27] B. Tao and M. T. Orchard, "Prediction of second-order statistics in motion-compensated video coding," in *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*. IEEE, 1998, pp. 910–914.

[28] X. Cai and J. S. Lim, "Algorithms for transform selection in multiple-transform video compression," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*. IEEE, 2012, pp. 2481–2484.

[29] X. Cai and J. Lim, "Algorithms for transform selection in multiple-transform video compression." *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, vol. 22, no. 12, p. 5395, 2013.

[30] "Derf sequences," https://media.xiph.org/video/derf/.

Xun Cai (S'12) is a Ph.D. candidate in the Advanced Telecommunications and Signal Processing group at MIT. He received the B.S. degree in the electrical engineering from the University of Science and Technology of China in 2010, and the S.M. degree in electrical engineering and computer science from the Massachusetts Institute of Technology in 2012. His current research interests include video compression, image processing and digital signal processing.

Jae Lim (S'76-M'78-SM'83-F'86) received the S.B., S.M., E.E., and Sc.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 1974,1975,1978 and 1978, respectively.

He joined the MIT faculty in 1978 and is currently a Professor in the Electrical Engineering and Computer Science Department. His research interests include digital signal processing and its applications to image and speech processing. He has contributed more than one hundred articles to journals and conference proceedings. He is a holder of more than 40 patents in the areas of advanced television systems and signal compression. He is the author of a textbook, *Two-Dimensional Signal and Image Processing* (Prentice-Hall).

Dr. Lim is the recipient of many awards including the Senior Award from the IEEE ASSP Society and the H.E. Edgerton Faculty Achievement Award from MIT. He is a fellow of the IEEE and a member of the Academy of Digital Television Pioneers.