JOINTLY OPTIMIZED TRANSFORM DOMAIN TEMPORAL PREDICTION AND SUB-PIXEL INTERPOLATION

Shunyao Li, Tejaswi Nanjundaswamy, and Kenneth Rose

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

ABSTRACT

Conventional pixel-domain block matching temporal (inter) prediction is suboptimal, since it ignores the underlying spatial correlation. Hence in our recent research we proposed transform domain temporal prediction (TDTP), wherein spatially decorrelated transform coefficients are individually predicted. Later we proposed extended block TDTP (EB-TDTP), which fully exploits spatial correlation around reference block boundaries. However, the transform domain temporal correlation exploited by (EB-)TDTP interferes with the frequency response of sub-pixel interpolation filters. Thus, in this paper, we propose to replace the standard sub-pixel interpolation with filters which are jointly designed with EB-TDTP based on statistics of the data, for either separable or non-separable interpolation structures. We also employ a two-loop asymptotic closed-loop (ACL) approach for statistically stable off-line design. Experiments show that our framework can achieve up to 1dB gain in PSNR over HEVC.

Index Terms— Temporal prediction, sub-pixel interpolation, spatial correlation, DCT, video coding

1. INTRODUCTION

In modern video coding standards, such as HEVC [1], compression efficiency is largely achieved by removing temporal redundancies between frames, which is usually referred to as inter prediction. Instead of directly encoding the raw pixel values, the encoder looks for a similar reference block in previously reconstructed frames through block matching. The difference between the reference block and the original block is then encoded and sent to the decoder. This one-to-one pixel domain block matching is suboptimal for inter prediction mainly because it ignores the underlying spatial correlation between pixels. Various approaches have been proposed to account for such spatial correlation in inter prediction, e.g., multi-tap filtering [2–4] and three-dimensional subband coding [5, 6]. However, these approaches usually suffer from high encoder complexity.

In our past work [7, 8], we proposed transform domain temporal prediction (TDTP), which tackles this problem from a different perspective. Since DCT achieves spatial decorrelation, it allows for an optimal one-to-one transform coefficient prediction. Therefore, for each DCT frequency i, we estimate the temporal correlation ρ_i and employ a first-order autoregressive (AR) prediction model with ρ_i as the scaling factor. In Table 1, we can see that the temporal correlation ρ_i varies considerably with frequency, and generally decreases as frequency increases. TDTP captures such variation in temporal correlation across frequencies, which is otherwise masked in the pixel domain by the dominant low frequencies, and the resulting $\rho \approx 1$

0.999	0.998	0.997	0.970	0.944	0.930	0.842	0.808
0.996	0.978	0.979	0.963	0.957	0.884	0.900	0.797
0.983	0.984	0.975	0.944	0.978	0.931	0.857	0.794
0.967	0.980	0.977	0.965	0.958	0.920	0.930	0.768
0.960	0.950	0.962	0.964	0.942	0.889	0.904	0.756
0.927	0.938	0.934	0.922	0.919	0.882	0.831	0.748
0.898	0.881	0.919	0.906	0.869	0.815	0.700	0.512
0.835	0.760	0.826	0.769	0.717	0.640	0.470	0.339

Table 1. Transform domain temporal correlations for 8x8 DCT coefficients of *mobile* sequence at QP=22

leads to the prevalence of block matching and copying techniques in current video codecs.

However, the sub-pixel interpolation filters (employed to achieve higher motion precision) in standard video codecs interfere with TDTP. Since the sub-pixel interpolation filters are designed to pass low frequency signals and attenuate high frequency signals, they have a similar scaling effect as the temporal correlation ρ_i across frequencies, which interferes with TDTP and thus needs to be properly accounted for in the TDTP design. These filters also exploit some extra spatial correlation by using the information from neighboring pixels around reference block boundaries.

To better account for the low-pass filter effect as well as the neighboring information used in interpolation, we proposed extended blocks TDTP (EB-TDTP) [9]. As shown in Fig. 1, we first perform DCT on an extended block (including reference block and its neighboring pixels) to get spatially (mostly) uncorrelated frequency components, then apply individually designed prediction coefficients per frequency, then inverse transform to the pixel domain, and linearly map (i.e., interpolate) to the smaller original block size for prediction. Clearly, the prediction coefficients and the interpolation depend on each other. Thus in this paper we propose to replace the standard sub-pixel interpolation with filters *jointly designed* with prediction coefficients based on the statistics of the data. In [9], we introduced optimization of prediction coefficients given a known separable interpolation filters' linear mapping. In this paper, we significantly expand it to joint optimization, where the linear mapping can be either separable or non-separable filters, depending on the local video statistics. Moreover, since the design of any module in a predictive compression system suffers from instability due to quantization error propagation [10], we also extend our asymptotic closed-loop (ACL) [10, 11] approach of stable design to the joint optimization of prediction coefficients and interpolation filters. Experimental results demonstrate the effectiveness of the proposed framework and optimization approach with up to 1dB gain in PSNR over the baseline HEVC.
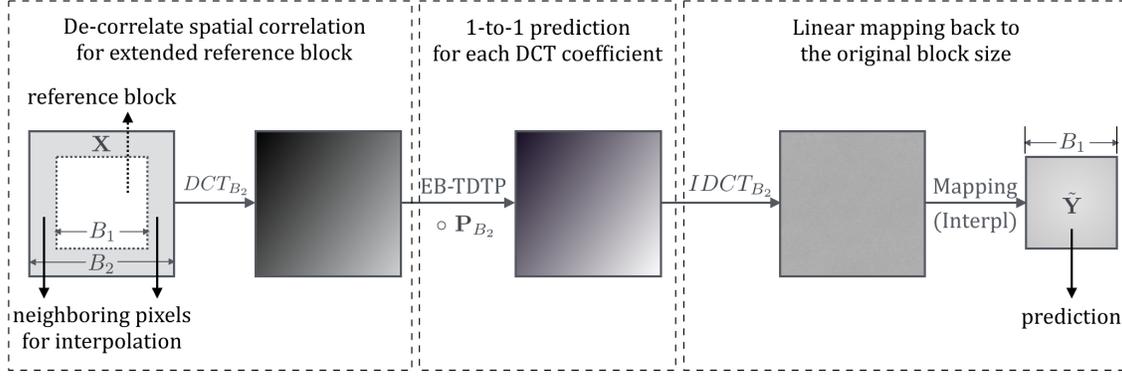

Fig. 1. Block diagram of the EB-TDTP framework

2. BACKGROUND

Similar to the pixel domain motion compensation, in EB-TDTP [9], we assume a first-order AR process for each DCT coefficient of blocks along a motion trajectory. We denote by x_n the DCT coefficient at a particular frequency of an inter-coded block in frame n, and by x_{n-1} the corresponding DCT coefficient of its reference block in frame $n - 1$, then the AR process is given as,

$$x_n = \rho x_{n-1} + z_n \tag{1}$$

where ρ is the prediction coefficient and z_n is the innovation. In a closed-loop operation, we use the reconstructed DCT coefficient, $\hat{x}_{n-1}$, as the reference. Thus the optimal prediction for each DCT coefficient is

$$\tilde{x}_n = \rho \hat{x}_{n-1}. \tag{2}$$

In the matrix form, this is equivalent to employing an element-by-element multiplication (denoted by $\circ$) on the extended reference block in DCT domain. We define $\mathbf{P}_b$ as a $b \times b$ matrix with elements as the prediction coefficients, ρ, corresponding to each frequency.

The complete block diagram for EB-TDTP is shown in Fig. 1. To predict the current block $\mathbf{Y}$ of size $B_1 \times B_1$, we take a larger block $\mathbf{X}$ of size $B_2 \times B_2$ centered around the motion compensated reference block, so that the information in neighboring pixels can be used effectively. Here we use $B_2 = B_1 + b$, where b is the number of taps in the separable interpolation filter. The vertical and horizontal interpolation filters in the matrix form are denoted as $\mathbf{F}_1$ and $\mathbf{F}_2$. Specifically, if the b-tap 1D vertical and horizontal interpolation filters are denoted as $\mathbf{f}_1$ and $\mathbf{f}_2$ (column vectors), then,

$$\mathbf{F}_1 = \begin{bmatrix} 0 & \mathbf{f}_1^T & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{f}_1^T & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \mathbf{f}_1^T \end{bmatrix} \tag{3}$$

$$\mathbf{F}_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \mathbf{f}_2 & 0 & \cdots & 0 \\ 0 & \mathbf{f}_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \mathbf{f}_2 \end{bmatrix} \tag{4}$$

at dimension $B_1 \times B_2$ and $B_2 \times B_1$, respectively. Therefore, the interpolated reference block is $\mathbf{F}_1 \mathbf{X} \mathbf{F}_2$. We also define $\mathbf{D}_{B_2}$ as the matrix operator of vertical 1D-DCT of size B_2. The final EB-TDTP

prediction $\tilde{\mathbf{Y}}$ can be formulated as,

$$\tilde{\mathbf{Y}} = \mathbf{F}_1 \mathbf{D}'_{B_2} ((\mathbf{D}_{B_2} \mathbf{X} \mathbf{D}'_{B_2}) \circ \mathbf{P}_{B_2}) \mathbf{D}_{B_2} \mathbf{F}_2 \tag{5}$$

where $\circ \mathbf{P}_{B_2}$ is the frequency-wise prediction in DCT domain introduced in (2).

3. JOINT OPTIMIZATION

As described in the Sec. 1, the sub-pixel interpolation filters interfere with EB-TDTP, and hence they need to be jointly designed with the prediction coefficients. Traditionally in video codecs, the interpolation is performed via separable filters [12–16]. However, these filters sometimes cannot perfectly capture the spatial correlation depending on the local statistics of the video sequence. Alternatively, non-separable filters [17, 18] can be more flexible, but cover smaller spatial area if we want to maintain the same complexity as separable filters. Therefore, we propose optimization approaches for both separable and non-separable filters, and choose one based on the statistics of the video sequence, with a sequence-level flag. We use $\mathbf{F}_I$ to denote the interpolation filter set in general, which in the following sections can be $\{\mathbf{F}_1, \mathbf{F}_2\}$ as two 1D separable filters or $\mathbf{F}$ as 2D non-separable filters.

3.1. Separable Filters

Our overall objective is to design $\{\mathbf{P}_{B_2}, \mathbf{F}_1, \mathbf{F}_2\}$ to minimize the mean squared prediction error (MSE),

$$J = \left\| \mathbf{Y} - \tilde{\mathbf{Y}} \right\|^2. \tag{6}$$

However, this is a non-linear multi-variate optimization problem. To make this problem tractable, we propose an iterative approach of optimizing one of $\{\mathbf{P}_{B_2}, \mathbf{F}_1, \mathbf{F}_2\}$, while fixing the other two. To simplify the cost expression, let us set

$$\mathbf{H}_1 = \mathbf{F}_1 \mathbf{D}'_{B_2}, \ \mathbf{H}_2 = \mathbf{D}_{B_2} \mathbf{F}_2, \tag{7}$$

$$\mathbf{X}_T = \mathbf{D}_{B_2} \mathbf{X} \mathbf{D}'_{B_2}, \tag{8}$$

so the cost is,

$$J = \left\| \mathbf{Y} - \mathbf{H}_1 (\mathbf{X}_T \circ \mathbf{P}_{B_2}) \mathbf{H}_2 \right\|^2. \tag{9}$$

In [9], we introduced the optimization approach for prediction coefficients, $\mathbf{P}_{B_2}$, with separable interpolation filters $\mathbf{F}_1, \mathbf{F}_2$ fixed. Given $\mathbf{P}_{B_2}$, we can optimize $\mathbf{F}_1, \mathbf{F}_2$ by optimizing $\mathbf{H}_1, \mathbf{H}_2$ (from

(7)). Since $\mathbf{H}_1$ and $\mathbf{H}_2$ are symmetric in (9), their optimization approaches are very similar.

First we assume $\mathbf{H}_1$ and $\mathbf{P}_{B_2}$ are fixed, then $\mathbf{H}_2$ only depends on $\mathbf{f}_2$, which reduces to the overall problem to linear optimization. That is, we convert J to $\|\mathbf{A}\mathbf{f}_2 - \mathbf{b}\|^2$, where,

$$\mathbf{A}(v,i) = \sum_{k=0}^{B_2-1}\sum_{l=0}^{B_2-1} \mathbf{D}_{B_2}(l, i+n+1)\mathbf{H}_1(m,k)\mathbf{X}_T(k,l)\mathbf{P}_{B_2}(k,l),$$
(10)

$$\mathbf{b}(v) = \mathbf{Y}(m,n),$$
(11)

$$v = mB_1 + n \ (m,n = 0...B_1 - 1), i = 0...b - 1.$$
(12)

The optimal solution for $\mathbf{f}_2$, given $\mathbf{P}_{B_2}$ and $\mathbf{f}_1$, is

$$\mathbf{f}_2 = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b},$$
(13)

and $\mathbf{H}_2$ can be derived from (4) and (7). Similarly, for optimizing $\mathbf{H}_1$ (or $\mathbf{f}_1$) fixing $\mathbf{H}_2$ and $\mathbf{P}_{B_2}$, (10) and (11) change to,

$$\mathbf{A}(v,i) = \sum_{k=0}^{B_2-1}\sum_{l=0}^{B_2-1} \mathbf{D}_{B_2}(k, i+m+1)\mathbf{H}_2(l,n)\mathbf{X}_T(k,l)\mathbf{P}_{B_2}(k,l),$$
(14)

$$\mathbf{b}(v) = \mathbf{Y}(m,n).$$
(15)

3.2. Non-Separable Filters

In the previous sub-section, we presented the joint optimization for prediction coefficients and separable filters. As mentioned, we also investigate the performance of non-separable filters as an alternative for more complex local statistics. We maintain the same complexity by using a 2D 4x4 non-separable filters, denoted as $\mathbf{F}$, to keep the same number of multiplications as using two 1D 8-tap filters (the current HEVC interpolation filters). The prediction $\tilde{\mathbf{Y}}$ from (5) becomes

$$\tilde{\mathbf{Y}} = (\mathbf{D}'_{B_2}((\mathbf{D}_{B_2}\mathbf{X}\mathbf{D}'_{B_2}) \circ \mathbf{P}_{B_2})\mathbf{D}_{B_2}) \circledast \mathbf{F} \quad (16)$$

$$= (\mathbf{D}'_{B_2}(\mathbf{X}_T \circ \mathbf{P}_{B_2})\mathbf{D}_{B_2}) \circledast \mathbf{F} \quad (17)$$

where $\circledast$ denotes the 2D convolution, which returns $\tilde{\mathbf{Y}}$ as the valid $B_1 \times B_1$ region at the center. We set $\mathbf{G} = \mathbf{D}'_{B_2}(\mathbf{X}_T \circ \mathbf{P}_{B_2})\mathbf{D}_{B_2}$, and p as the size of 2D $p \times p$ non-separable filter, then

$$\tilde{\mathbf{Y}} = \mathbf{G} \circledast \mathbf{F}, \quad (18)$$

$$\tilde{\mathbf{Y}}(m,n) = \sum_{i=-p/2}^{p/2-1}\sum_{j=-p/2}^{p/2-1} \mathbf{F}(i+\frac{p}{2}, j+\frac{p}{2})\mathbf{G}(m+i+1, n+j+1),$$
(19)

$$\mathbf{G}(s,t) = \sum_{k=0}^{B_2-1}\sum_{l=0}^{B_2-1} \mathbf{D}_{B_2}(k,s)\mathbf{X}_T(k,l)\mathbf{P}_{B_2}(k,l)\mathbf{D}_{B_2}(l,t).$$
(20)

$$(m,n = 0...B_1 - 1, \ s,t = 0...B_2 - 1)$$

We again use an iterative approach to jointly optimize $\mathbf{F}$ and $\mathbf{P}_{B_2}$. Given $\mathbf{P}_{B_2}$, the optimal $\mathbf{F}$ would be the Wiener filter. Given $\mathbf{F}$, to estimate $\mathbf{P}_{B_2}$, we convert it to a least square estimation problem of minimizing $\|\mathbf{A}\mathbf{p}_{B_2} - \mathbf{b}\|^2$, where $\mathbf{p}_{B_2}$ is the vector form (of size
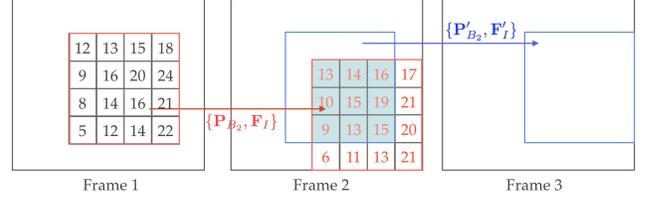

Fig. 2. The instability problem in closed-loop predictor design

$B_2^2 \times 1$) of $\mathbf{P}_{B_2}$, with $\mathbf{A}$ and $\mathbf{b}$ as,

$$\mathbf{A}(v,u) = \sum_{i=1}^{p}\sum_{j=1}^{p} \mathbf{F}(i,j)\mathbf{D}_{B_2}(k, m-\frac{p}{2}+i)$$
$$\mathbf{X}_T(k,l)\mathbf{D}_{B_2}(l, n-\frac{p}{2}+j) \quad (21)$$

$$\mathbf{b}(v) = \mathbf{Y}(m,n) \quad (22)$$

$$v = mB_1 + n \ (m,n = 0...B_1 - 1) \quad (23)$$

$$u = kB_2 + l \ (k,l = 0...B_2 - 1) \quad (24)$$

The optimal solution for prediction coefficients $\mathbf{P}_{B_2}$ given 2D non-separable interpolation filter $\mathbf{F}$ is

$$\mathbf{p}_{B_2} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}, \quad (25)$$

$$\mathbf{P}_{B_2}(k,l) = \rho_{k,l} = \mathbf{p}_{B_2}(u). \quad (26)$$

We iteratively optimize $\mathbf{F}$ and $\mathbf{P}_{B_2}$, until the prediction error converges.

4. OVERALL TRAINING PROCESS

We incorporate this joint optimization into a two-loop asymptotic-closed loop (ACL) design scheme we proposed in [10] to solve the design instability problem in off-line training. In closed-loop video coding, reconstructed frames are used as reference for future frames. As shown in Fig. 2, the $\{\mathbf{P}_{B_2}, \mathbf{F}_I\}$ are trained for statistics of the given matching blocks (the two red and blue pairs). However, as we employ them in closed-loop, the reconstructed frames are updated, which changes the reference frames for the future frames (the shaded region in frame 2 is updated by the red block pair, and is used as reference for the blue block in frame 3). This results in deviation from the statistics we designed for in future frames, which grows over time and leads to substantial ineffectiveness of the designed prediction parameters.

In ACL, we use an iterative open-loop (hence stable) design procedure that approximates closed-loop operation on convergence. Specifically, we use a two loop ACL where in the inner loop we fix the encoder decisions and update prediction parameters $\{\mathbf{P}_{B_2}, \mathbf{F}_I\}$, and in the outer loop encoder decisions are updated with $\{\mathbf{P}_{B_2}, \mathbf{F}_I\}$ fixed. The detailed overall training algorithm is summarized in Algorithm 1.

5. RESULTS

The proposed framework is implemented in HM 14.0 to compare the coding gain over the baseline HEVC of TDTP, EB-TDTP, and jointly optimized EB-TDTP with interpolation filters. Without loss of generality, all sequences are coded in IPPP format, and only previous frame is allowed as reference. Note that the approach is also applicable to bi-prediction and long-term reference motion compensation. We allow the sub-pixel interpolation to be at quarter-pel precision. To simplify the experiments, both prediction and transform block sizes are restricted to 8×8, and the sample adaptive offset (SAO) is temporarily disabled.

Initialize $\mathbf{P}_{B_2}$ and $\mathbf{F}_I$;
Run the codec in closed-loop, store the encoder decision d and reconstructed video file rec;
while $iter < max\_iter$ **do**
  while *MSE decreases* **do**
    Run the codec in open-loop using the same encoder decision d, but with rec as the motion compensation reference; Extract the reference and original blocks;
    for *each sub-pixel position* **do**
      (a) Optimize EB-TDTP coefficients $\mathbf{P}_{B_2}$;
      (b) Optimize interpolation filters $\mathbf{F}_I$;
      Repeat (a) and (b) until MSE converges;
    end
    Run the codec in open-loop again, using the same d, rec, and the optimized $\mathbf{P}_{B_2}$ and $\mathbf{F}_I$; Update rec as the newly reconstructed video file; Update MSE;
  end
  Run the codec in closed-loop, update the encoder decision d and reconstructed video file rec, store the RD-cost into array $\{cost\}$. $iter = iter + 1$;
end
Find the minimum RD cost in $\{cost\}$, set the corresponding $\mathbf{P}_{B_2}$ and $\mathbf{F}_I$ as the final trained predictors;

Algorithm 1: Joint design algorithm for EB-TDTP and interpolation filters

We first examine the full potential of jointly optimized EB-TDTP system by designing a specific set of coefficients for each sequence using the off-line training method described above. Each sequence is tested at various bitrates with QP ranging from 22 to 37. BD-rate improvement over the baseline HEVC is provided in Table 2. Generally, the jointly optimized EB-TDTP with interpolation filters outperform the other two with an average of 7.54% in BD-rate reduction. The RD curves for the sequence *BQSquare* is shown in Fig. 3, with up to 1 dB PSNR improvements over the baseline HEVC.

While the above results are applicable for video storage applications where encoding is performed off-line, allowing use of individually optimized predictor and interpolator, we also evaluate the performance outside the training set, which is useful in other applications. We provide a choice of fixed 8 sets of predictors for the encoder to choose at each frame, with a negligible overhead of 3 bits per frame. For simplicity, we use the 8 most distinct sets of predictors from the training set as choices for this evaluation. BD-rate improvement over the baseline HEVC for a test dataset is shown in Table 3 with an reasonable gain of 3.76% in BD-rate reduction. Future research directions include optimal training of these choices of predictors, developing an efficiently adaptive EB-TDTP framework and design approaches.

6. CONCLUSION

This paper proposed a framework of extended block transform domain temporal prediction which is jointly designed with the sub-pixel interpolation filters. Joint optimization of both separable and non-separable filters (at the same complexity) is proposed. Moreover, the design employs the asymptotic-closed loop approach to avoid instability in off-line training. Experimental results with substantial gains (up to 1dB gain in PSNR) demonstrate the effectiveness of the proposed framework.
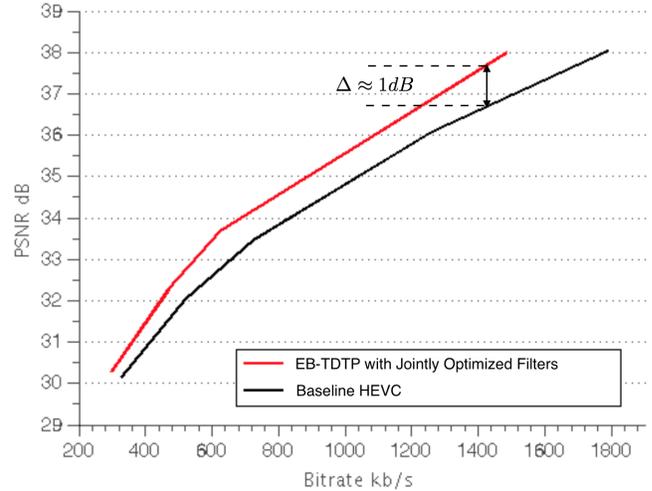

Fig. 3. Coding performance comparison for sequence *BQSquare*

	TDTP	EB-TDTP	JointOpt
coastguard (CIF)	8.61	10.03	10.63
bridge-far (CIF)	9.20	10.58	11.35
mobile(CIF)	4.60	7.44	8.58
highway (CIF)	3.94	6.10	7.73
stefan (CIF)	3.67	3.90	4.67
BQSquare (240p)	0.74	1.90	14.44
BlowingBubbles (240p)	1.06	1.01	1.20
BQMall (480p)	2.04	1.83	3.43
PartyScene (480p)	1.21	1.41	5.72
Keiba (480p)	4.04	4.52	4.48
vidyo1 (720p)	1.53	2.51	2.51
BQTerrace (1080p)	12.78	15.03	20.14
ParkScene (1080p)	2.53	2.57	2.57
Kimono (1080p)	7.21	6.91	8.16
AVERAGE	**4.51**	**5.41**	**7.54**

Table 2. Comparison of reduction in bitrate over reference encoder for training set (Note that although in terms of prediction error, JointOpt is guaranteed to be better than EB-TDTP, this does not always translate to RD performance improvement.)

	JointOpt
container (CIF)	9.16
bridge-close (CIF)	6.26
bus (CIF)	3.77
tempete (CIF)	3.67
waterfall (CIF)	1.81
flower (CIF)	0.21
city (CIF)	0.92
FourPeople (720p)	6.27
vidyo3 (720p)	4.18
vidyo4 (720p)	3.59
BasketballDrive (1080p)	4.71
Cactus (1080p)	3.21
Tennis (1080p)	1.09
AVERAGE	**3.76**

Table 3. Reduction in bitrate over reference encoder for test set

7. REFERENCES

[1] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[2] J. Kim and J. W. Woods, "Spatio-temporal adaptive 3-D kalman filter for video," *IEEE Transactions on Image Processing*, vol. 6, no. 3, pp. 414–424, 1997.

[3] T. Wedi, "Adaptive interpolation filter for motion and aliasing compensated prediction," in *Electronic Imaging 2002*. International Society for Optics and Photonics, 2002, pp. 415–422.

[4] S. Li, O. G Guleryuz, and S. Yea, "Reduced-rank condensed filter dictionaries for inter-picture prediction," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 1428–1432.

[5] S.-J. Choi and J. W. Woods, "Motion-compensated 3-D subband coding of video," *IEEE Transactions on Image Processing*, vol. 8, no. 2, pp. 155–167, Feb 1999.

[6] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 559–571, Sep 1994.

[7] J. Han, V. Melkote, and K. Rose, "Transform-domain temporal prediction in video coding: exploiting correlation variation across coefficients," in *IEEE International Conference on Image Processing (ICIP)*, 2010, pp. 953–956.

[8] J. Han, V. Melkote, and K. Rose, "Transform-domain temporal prediction in video coding with spatially adaptive spectral correlations," in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, 2011, pp. 1–6.

[9] S. Li, T. Nanjundaswamy, and K. Rose, "Transform domain temporal prediction with extended blocks," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 1476–1480.

[10] S. Li, T. Nanjundaswamy, Y. Chen, and K. Rose, "Asymptotic closed-loop design for transform domain temporal prediction," in *IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 4907–4911.

[11] H. Khalil, K. Rose, and S. Regunathan, "The asymptotic closed-loop approach to predictive vector quantizer design with application in video coding," in *IEEE transactions on image processing*, 2001, pp. 15–23.

[12] S. Wittmann and T. Wedi, "Separable adaptive interpolation filter for video coding," in *IEEE International Conference on Image Processing (ICIP)*, 2008, pp. 2500–2503.

[13] T. Wedi and H. G. Musmann, "Motion-and aliasing-compensated prediction for hybrid video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 577–586, 2003.

[14] K. Ugur, J. Lainema, and M. Gabbouj, "Adaptive interpolation filter with flexible symmetry for coding high resolution high quality video," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2007, vol. 1, pp. I–1013.

[15] D. Rusanovskyy, K. Ugur, and M. Gabbouj, "Adaptive interpolation with flexible filter structures for video coding," in *IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 1025–1028.

[16] Y. Ye, G. Motta, and M. Karczewicz, "Enhanced adaptive interpolation filters for video coding," in *IEEE Data Compression Conference (DCC)*, 2010, pp. 435–444.

[17] Y. Vatis and J. Ostermann, "Locally adaptive non-separable interpolation filter for H. 264/AVC," in *IEEE International Conference on Image Processing (ICIP)*, 2006, pp. 33–36.

[18] Y. Vatis, B. Edler, D. T. Nguyen, and J. Ostermann, "Motion-and aliasing-compensated prediction using a two-dimensional non-separable adaptive wiener interpolation filter," in *IEEE International Conference on Image Processing (ICIP)*, 2005, vol. 2, pp. II–894.