RATE-DISTORTION OPTIMIZATION AND ADAPTATION OF INTRA PREDICTION FILTER PARAMETERS

Shunyao Li, Yue Chen, Jingning Han[†]*, Tejaswi Nanjundaswamy, and Kenneth Rose*

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

ABSTRACT

Conventional "pixel copying" prediction used in current video standards was shown in previous work to be sub-optimal compared to 2-D non-separable Markov model based recursive extrapolation approaches. The premise of this paper is that in order to achieve the full potential of these approaches it is necessary to account for several requirements, namely, the design of prediction modes (and respective extrapolation filters) must optimize a rate-distortion cost rather than minimize the mean squared prediction error; the filters must be of sufficient complexity to cover all necessary directions; and the approach must include adaptation to available information indicative of local statistics. Hence, the proposed system employs four-tap recursive extrapolation filters that can predict from all standard directions, combined with a filter design method that accounts for the overall rate-distortion cost in conjunction with the codec decisions, along with adaptation of filter coefficients to relevant local information provided by encoder decisions on target bit rate and block size. Experimental evidence is provided for substantial coding gains over conventional intra coding.

***Index Terms*—** Spatial prediction, rate-distortion optimization, adaptive extrapolation filter, video coding

1. INTRODUCTION

Intra prediction is critical for image and video coders to exploit spatial correlations within a image/frame. In current block based video coders [1, 2], reconstructed boundary pixels (or their linear combination) are copied along a specific direction to generate prediction for the current block, wherein the directionality depends on local texture. This technique assumes an extreme separable Markov model with correlation coefficient of one along the selected direction and zero correlation along the perpendicular direction. This simplistic approach underutilizes the information available in the neighboring pixels, as it ignores the possibility that correlation is not perfectly and extremely separable, as well as the expected variations in correlation within a block, especially the correlation decay with distance from the boundary. Many approaches have been proposed to overcome these limitations. Predicting every pixel as a linear combination of all the boundary pixels was proposed in [3], and predicting blocks using weighted average of multiple decoded blocks was proposed in [4]. However, both these approaches suffer from considerable increase in computational complexity. Other related work includes [5] and [6], wherein both spatial and temporal correlations are jointly exploited, and hence not applicable when coding with no dependency on the past. Recursive extrapolation based on a separable Markov model was proposed in [7], however, this approach neglects the possibility that correlation is not perfectly separable. As an effective low computational complexity solution, our lab proposed in [8], a non-separable Markov model based recursive extrapolation approach with three-tap filters. However, the three-tap filters could not capture all relevant prediction directions. Recursive extrapolation filters with four-taps to cover more directionalities appeared in [9], where filters were designed to minimize the mean squared prediction error rather than the ultimate rate-distortion (RD) cost. None of the above recursive extrapolation techniques adapt the filters to account for variation in overall signal statistics.

In this paper, we propose a RD optimized, adaptive, recursive extrapolation filters with four taps. The fourth tap points at either the top-right pixel or the bottom-left pixel, depending on the prediction directionality. Moreover, as the design is within an overall RD framework, we also retain the original prediction modes to provide additional flexibility to the encoder when it makes RD cost-based prediction mode selections. A gradient descent approach (which accounts for the availability of the original modes) is employed to optimize the new mode filters with respect to the RD cost. Noting that the RD cost surface is highly complex and riddled with poor local optima, we employ, as reasonable initialization for the optimization procedure, a set of mode filters obtained by minimizing the simpler mean squared prediction error cost. We also make the filters adaptive to local information of target bit rate and block size to account for variation in statistics. We demonstrate the efficacy of the approach by implementing it within the VP9 framework. We also demonstrate its general applicability via a preliminary implementation within the HEVC framework. Evaluation results provide evidence of substantial coding gains over conventional intra coding.

2. RECURSIVE EXTRAPOLATION FILTERING

The basic building block of this proposal is the non-separable Markov model based recursive extrapolation filter to tackle the underutilization of available boundary information in conventional "pixel copying" based intra prediction. In [8], the image signals were modeled by a three-tap Markov process with evolution recursion given as:

$$X = c_v V + c_h H + c_d D + \epsilon \tag{1}$$

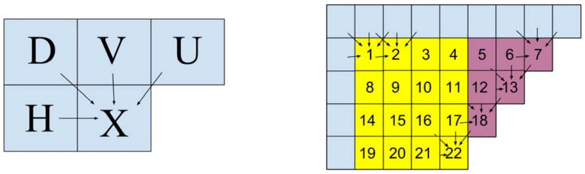
This three tap structure has a critical limitation of not being able to represent directions arising from top-right and bottom-left. Hence, to cover all directions we add a fourth tap of, top-right pixel with left-to-right prediction scan order, or, bottom-left pixel with top-to-bottom prediction scan order, similar to [9]. This modification requires previously reconstructed boundary pixels to be available at top-right of the current block, or at bottom-left of the current block. While the availability of the top-right reconstructed boundary pix-

---

This work was supported by Google Inc.

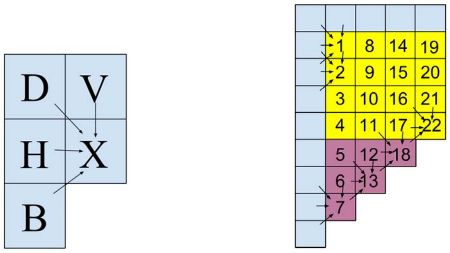
[†] Jingning Han is now with Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043.

U.S. Government work not protected by U.S. copyright



(a) 2-D Markov Process for top-right directions

(b) Prediction order for top-right directions



(c) 2-D Markov Process for bottom-left directions

(d) Prediction order for bottom-left directions

Fig. 1. Four-tap recursive extrapolation filter.

els is almost always ensured, newly introduced coding tools employing smaller prediction blocks (or prediction units, PU) within large coding blocks (or coding units, CU), also provide bottom-left reconstructed boundary pixels for some blocks. In case these additional boundary pixels are unavailable, the existing boundary pixel is copied to the unavailable pixel positions. The extended Markov process with four taps for directions originating from top-right is given as:

$$X = c_v V + c_h H + c_d D + c_u U + \epsilon \tag{2}$$

where V, H, D, U are neighboring pixels of X, as illustrated in Fig. 1(a). The four coefficients, c_v, c_h, c_d, and c_u, together capture the texture directionality. For medium to high bit rates, the reference pixels can be approximated by their reconstructions, to obtain the optimal predictor for X as

$$\tilde{X} = c_v \hat{V} + c_h \hat{H} + c_d \hat{D} + c_u \hat{U}. \tag{3}$$

The recursive prediction order is enumerated from 1 to 22 in Fig. 1(b) for a 4 × 4 block, and we note that a temporary triangular region is estimated to provide top-right references to the right boundary pixels of the current block. While the discussion hereafter focuses, for concreteness, on four-tap filters for directions originating from the top-right, the transposed versions of the Markov process, prediction filter, and prediction order, as illustrated in Fig. 1(c) and (d), are used for corresponding directions originating from the bottom-left.

3. FILTER DESIGN TO MINIMIZE PREDICTION ERROR

The overall prediction error to be minimized for the initial filter design is,

$$J = \sum_k J_k = \sum_k \sum_{\substack{\forall \text{ blocks} \\ \text{in Mode } k}} \sum_{i,j} (\tilde{x}_{i,j} - x_{i,j})^2. \tag{4}$$

To start the design, all the blocks are first classified into original "pixel copying" modes based on the prediction error, then filters for each of these block subsets are initialized with the optimal linear predictor coefficients calculated as below, using the open-loop statistics of the 2-D non-separable Markov model of (2),

$$\begin{bmatrix} c_v \\ c_h \\ c_d \\ c_u \end{bmatrix} = \begin{bmatrix} R_{VV} & R_{VH} & R_{VD} & R_{VU} \\ R_{VH} & R_{HH} & R_{HD} & R_{HU} \\ R_{VD} & R_{HD} & R_{DD} & R_{DU} \\ R_{VU} & R_{HU} & R_{DU} & R_{UU} \end{bmatrix}^{-1} \begin{bmatrix} R_{XV} \\ R_{XH} \\ R_{XD} \\ R_{XU} \end{bmatrix}, \tag{5}$$

where R_{XV} denotes the cross correlation between pixel X and its upper pixel V, and so forth.

These filter coefficients are then optimized to minimize the overall prediction error of (4) via the "K-modes clustering", wherein the two steps of re-clustering the blocks, and re-estimating filter coefficients for each cluster, are repeated until convergence. Noting that the above procedure implicitly assumed that the filter operates on original values and not on predicted values as the recursive operation actually proceeds, a gradient descent approach is then employed to properly re-estimate filter coefficients for each cluster, with the gradient analytically calculated by taking a partial derivative of the per-mode cost J_k with respect to each of its coefficients, similar to [6]. For example, partial derivative with respect to $c_{v,k}$ is,

$$\frac{\partial J_k}{\partial c_{v,k}} = \sum_{\substack{\forall \text{ blocks} \\ \text{in Mode } k}} \sum_{i,j} 2(\tilde{x}_{i,j} - x_{i,j}) \frac{\partial \tilde{x}_{i,j}}{\partial c_{v,k}}, \tag{6}$$

where $\frac{\partial \tilde{x}_{i,j}}{\partial c_{v,k}}$, derived using (3), has the following recursive relationship,

$$\begin{aligned} \frac{\partial \tilde{x}_{i,j}}{\partial c_{v,k}} &= \tilde{x}_{i-1,j} + c_{v,k} \frac{\partial \tilde{x}_{i-1,j}}{\partial c_{v,k}} + c_{h,k} \frac{\partial \tilde{x}_{i,j-1}}{\partial c_{v,k}} \\ &\quad + c_{d,k} \frac{\partial \tilde{x}_{i-1,j-1}}{\partial c_{v,k}} + c_{u,k} \frac{\partial \tilde{x}_{i-1,j+1}}{\partial c_{v,k}}. \end{aligned} \tag{7}$$

Other partial derivatives can be derived similarly to calculate the required gradient.

4. RD OPTIMIZED ADAPTIVE FILTERS

Designing filter coefficients to minimize prediction error is mismatched with the ultimate RD cost that coders optimize. In other words, the resulting prediction error reduction will not necessarily or fully translate into RD performance improvement. This fundamental shortcoming motivates us to propose recursive extrapolation filters that are designed by direct optimization of the RD cost, and are further adaptive to the target bit rate normalized to the spatial resolution of the current frame, so as to match the filters to relevant signal statistics. In another deviation from previous recursive extrapolation approaches, we retain the original "pixel copying" prediction modes in addition to the filtering prediction modes so as to provide the encoder with additional flexibility as it optimizes the RD cost for prediction mode selection. We correspondingly adjust the filter design to account for the presence of original modes. As reasonable initialization for optimizing the complex RD cost surface, we use the filter coefficients obtained by directly minimizing the prediction error cost. We must however modify the basic prediction error minimizing design presented in Section 3 to account for the presence of original modes. The two steps of "K-modes clustering" now include, re-clustering the blocks into both filtering prediction modes and original modes, and re-estimating filter coefficients for only the filtering prediction modes. Then every set of filter coefficients obtained during each iteration is evaluated in terms of RD to select the best initialization for the RD cost optimizing filter design procedure.

The actual encoder with both filtering prediction modes and original prediction modes is used to obtain the rate and distortion, i.e., after taking into account the prediction, transformation, quantization and entropy coding. We operate the encoder in the target bit rate mode, thus simplifying the filter coefficient optimization problem to minimizing distortion, D, for the given rate, R. This optimization is done via gradient descent, with coefficients of one mode optimized at a time. The partial derivatives required for the gradient are calculated empirically, e.g., partial derivative with respect to the vertical coefficient of the kth mode is calculated as,

$$\frac{\partial D(c_{v,k})}{\partial c_{v,k}} = \frac{D(c_{v,k}+\Delta) - D(c_{v,k}-\Delta)}{2\Delta}. \tag{8}$$

During this optimization of filter coefficients of a mode, distortion can improve due to both reduction in prediction error for a block that already used this mode, or reclassification of a block from another mode to the current mode.

We also make filter coefficients adaptive to target bit rate normalized to the spatial resolution so that statistics while operating at around a bit rate are taken into account for filter design. We first divide all possible normalized target bit rates into N_R regions, and design filter coefficients for each of this region by encoding the training data at the centroid bit rate of a given region. As normalized target bit rate information is not available to the decoder, we add that as an additional frame-level side information, and account for this very minor cost in overall rate. We also make the filter coefficients adaptive to different block sizes so as to account for the relevant statistics. The overall filter design technique is summarized in Algorithm 1.

Algorithm 1 Overall filter design algorithm

```
Collect blocks of different sizes and modes by encoding the training data with the actual encoder.
for each block size do
    Design of filter coefficients of all modes to minimize prediction error.
    for each normalized target bit rate region do
        Selection of minimum distortion resulting set of filter coefficients amongst those obtained during prediction error minimizing design, as initialization, using the actual encoder.
        for each filter mode do
            Gradient descent optimization of filter coefficients of the current mode to minimize distortion for the target bit rate (which is centroid of the current bit rate region), using the actual encoder to calculate the required partial derivatives as in (8).
        end for
    end for
end for
```

5. EXPERIMENTAL RESULTS

We implement the proposed technique within the VP9 framework to demonstrate its efficacy. We then demonstrate its general applicability with a preliminary implementation in the HEVC framework.

5.1. Experimental Results for VP9

We compare the following three coders in this experiment:

- The standard reference VP9 codec.
- VP9 codec with its 10 prediction modes replaced with filters designed to minimize prediction error.
- VP9 codec with our proposed RD optimized adaptive filtering 10 prediction modes in addition to the 10 original modes.

The filtering mode being active is indicated with one additional bit, which is encoded using the arithmetic coding framework of VP9, and the probability table required is estimated using the training data. Note that in all our experiments training data is excluded from the test data.

Various clips from the derf dataset were encoded in intra-only settings. The performance gains of prediction error optimizing filters and our proposed RD optimizing filters in terms of percentage bit rate reduction over standard reference VP9 codec are presented in Table 1 for various PSNR and test clips. These results show that the gains obtained by our proposed technique is clearly superior to that of the prediction error optimizing technique.

Table 1. Comparison of reduction in bit rate over reference VP9 codec for using prediction error optimized filters and RD optimized filters.

	Bit rate reduction (%)					
Methods	Prediction error optimized filters			RD optimized filters		
PSNR (in dB)	32	35	38	32	35	38
bus(CIF)	0.78	0.93	1.01	3.57	3.40	3.26
tempete(CIF)	1.59	1.65	1.45	3.83	3.86	3.66
waterfall(CIF)	1.77	2.27	2.21	3.93	4.33	3.95
stockholm(720p)	0.78	2.25	2.09	4.88	5.75	5.11
park_run(720p)	2.13	1.74	1.32	3.54	3.15	2.74
ducks take off (720p)	0.80	1.10	0.86	4.37	4.07	3.58
old town cross (1080p)	-0.91	1.11	1.18	3.23	4.91	4.02
crowd_run(1080p)	1.18	1.44	1.30	3.52	3.71	3.56
Average for above 8 clips	1.02	1.56	1.43	3.86	4.15	3.74

We also plot the RD curves comparing standard VP9 codec, prediction error optimized filters and RD optimized filters for the test sequences *bus* in Fig. 2 and *ducks take off* in Fig. 3. We can clearly see that RD optimization of the filters provides considerable gains at a wide range of bit rates.

5.2. Preliminary Experimental Results for HEVC

We demonstrate the general applicability of the proposed technique with a preliminary implementation within HEVC, where we replace the 9 least used intra modes in HEVC with filtering based modes for all but 32×32 blocks, we do not make the filter coefficients adaptive to target bit rate, and we introduce a frame level flag to selectively disable using filtering based modes when they do not offer gains. We don't add the filtering modes as new modes since HEVC already has a large number of intra prediction modes, and properly adding modes would requires careful handling of side information. Making filters adaptive to target bit rate and side information optimization is part of current work.

Various clips from the derf dataset were encoded in intra main settings within the HEVC's HM12.1 framework with 32×32 blocks

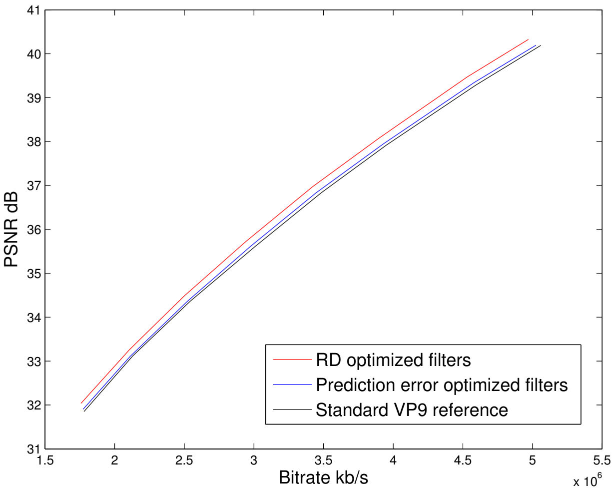


Fig. 2. Coding performance comparison for sequence *bus* at *CIF* resolution in VP9

disabled. The percentage bit rate reduction obtained for this preliminary implementation within HEVC when compared to the standard HEVC coder are given in Table 2 for various PSNR and test clips. The preliminary results are promising as they demonstrate the proposed technique's potential to obtain substantial gains when fully optimized.

Table 2. Reduction in bit rate over HM12.1 reference encoder for using preliminary RD optimized filters.

Bit Savings (%)			
PSNR (dB)	32	35	38
waterfall(CIF)	0.97	0.62	1.21
ducks take off(720p)	1.88	1.11	0.63
tractor(1080p)	1.20	1.62	1.92
riverbed(1080p)	-0.11	0.58	1.69

Finally, the average BD-rate [10] improvements for using RD optimized filters compared to standard coders for 37 clips from the derf dataset subdivided based on clip resolution are given in Table 3 for both VP9 and HEVC. All the evaluation results clearly demonstrate that consistent gains are provided by the proposed RD optimized adaptive recursive extrapolation filters.

Table 3. BD-rate improvement for our proposed RD optimized filters over standard VP9 and HEVC coders.

Bit Savings (%)		
Resolution	VP9	HEVC
CIF	2.90	0.20
720p	3.55	0.71
1080p	2.89	1.02

6. CONCLUSION

This paper demonstrates that a RD optimized adaptive recursive extrapolation filter based on 2-D non-separable Markov model with four taps can effectively exploit spatial correlation. The proposed technique aligns the filter design to the ultimate RD cost of the encoder, while leveraging initialization from design that minimizes prediction error; employs four tap filters to cover all directionalities; adapts the filter coefficients to variations in local statistics by exploiting information conveyed by target bit rate and block size; and retain the original prediction modes to provide additional encoder flexibility within the RD framework. Consistent performance gains were demonstrated by evaluation results.

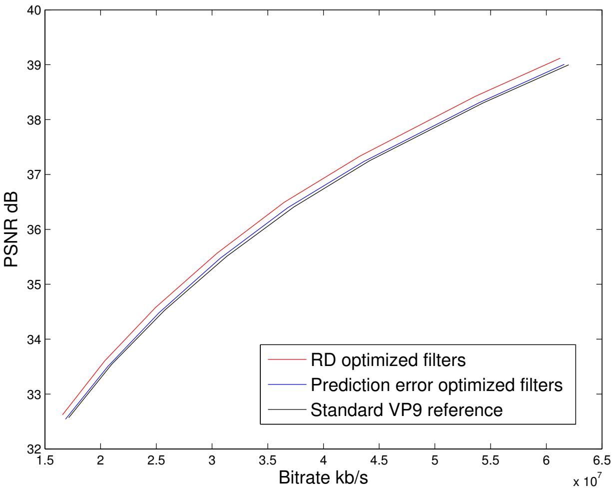


Fig. 3. Coding performance comparison for sequence *ducks take off* at *720p* resolution in VP9

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. 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, Feb 2013, p. 866606.

[3] L. Zhang, X. Zhao, S. Ma, Q. Wang, and W. Gao, "Novel intra prediction via position-dependent filtering," *Journal of Visual Communication and Image Representation*, vol. 22, no. 8, pp. 687–696, Nov 2011.

[4] H. Kimata, M. Kitahara, and Y. Yashima, "Recursively weighting pixel domain intra prediction on H.264," in *Visual Communications and Image Processing 2003*. International Society for Optics and Photonics, Nov 2003, pp. 2035–2042.

[5] L. Liu, Y. Liu, and E. J. Delp, "Enhanced intra prediction using context-adaptive linear prediction," in *Proceeding of Picture Coding Symposium*, Dec 2007, pp. 1–4.

[6] Y. Chen, J. Han, T. Nanjundaswamy, and K. Rose, "A joint spatio-temporal filtering approach to efficient prediction in video compression," in *Proceeding of Picture Coding Symposium*, Dec 2013.

[7] F. Kamisli, "Intra prediction based on statistical modeling of images," in *IEEE Visual Communications and Image Processing (VCIP)*, Nov 2012, pp. 1–6.

[8] Y. Chen, J. Han, and K. Rose, "A recursive extrapolation approach to intra prediction in video coding," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, May 2013, pp. 1734–1738.

[9] F. Kamisli, "Intra prediction based on markov process modeling of images," *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 3916–3925, Aug 2013.

[10] G. Bjontegard, "Calculation of average psnr differences between RD-curves," *VCEG Meeting (ITU-T SG16 Q.6), Austin, Texas, USA*, April 2001.