A RECURSIVE EXTRAPOLATION APPROACH TO INTRA PREDICTION IN VIDEO CODING

Yue Chen, Jingning Han, and Kenneth Rose

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

ABSTRACT

A novel intra prediction scheme, based on recursive extrapolation filters, is introduced. Standard intra prediction largely consists of copying boundary pixels (or linear combinations thereof) along certain directions, which reflects an overly simplistic model for the underlying spatial correlations. As an alternative, we view the image signal as a 2-D *non-separable* Markov model, whose corresponding correlation model better captures the nuanced directionality effects within blocks. This viewpoint motivates the design of a set of prediction modes represented by three-tap extrapolation filters, which replace the standard "pixel-copying" prediction modes, and provide efficient prediction at modest complexity. The Markov property is exploited by recursive predictions from nearest neighbors without recourse to simplistic separability assumptions, and while effectively accounting for correlation decay with distance from available boundary pixels. Coefficients for the set of mode filters are first trained by an efficient "k-modes" iterative technique designed to monotonically decrease the mean squared prediction error, and are then adjusted to directly optimize the overall rate-distortion objective. This prediction scheme complements the hybrid (cosine and sine) transform coding approach developed by our group, to achieve consistent coding gains, as shown for standard and commercial intra coders such as H.264/AVC and VP8.

Index Terms— Spatial prediction, prediction filter, transform coding, video coding

1. INTRODUCTION

The Intra modes of modern video codecs exploit spatial inter-block redundancies by predicting from previously reconstructed boundary pixels. This typically involves a set of modes, each generating the prediction by copying boundary pixels (or their linear combination) along a certain angle, to imitate the directionality of the texture content [1]. The encoder then selects, amongst these prediction modes, the one that minimizes the rate-distortion cost, per block, to adapt to local statistics. It effectively assumes an extreme separable Markov model for the image signal in the block, which is oriented according to the prediction direction, namely, where the inter-pixel correlation coefficient equals one along the selected direction, and zero along the perpendicular direction.

The simplistic and ad hoc nature of the above scheme ignores the nuanced variations in correlation across the block, and in particular, how correlation decays with distance from the boundary, as well as the strong possibility that the signal is not perfectly separable, which implies significant underutilization of information from neighboring pixels. This motivates the proposed recursive extrapolation approach to intra prediction. It assumes a 2-D non-separable Markov model for the image, which is known to capture image statistics significantly better than the special case of the conventional 2-D separable model. However, such generalization is notorious for its added complexity, a fact that made separable models popular despite their acknowledged suboptimality. The proposed scheme attempts to recoup the benefits of the non-separable model for the specific prediction task, while circumventing the complexity cost, by employing a unified 3-tap extrapolation filter, whose coefficients vary across different prediction modes. These coefficient values are determined off-line using a "k-mode" iterative technique (a distant relative of the known k-means clustering method), to directly minimize the mean squared prediction error, or ultimately the overall rate-distortion cost, over the training data. Exploiting the Markov property, the prediction can be effectively implemented by recursively applying the 3-tap filter throughout the block, thereby maintaining fairly low computational complexity.

Highly relevant prior work includes [2], where in addition to the standard directional prediction modes, a complementary filter-based prediction mode is introduced, whose coefficients are adaptively estimated according to the motion-compensated reference blocks in the previous frames, i.e., it essentially utilizes inter-frame information within the intra mode. Another approach proposed in [3] formulates the per-pixel prediction as a linear combination of all the available boundary pixels of the block of interest, where the coefficients are updated on-the-fly. For instance, an $N \times N$ block will require N^2 predictors per mode, each a linear combina-

This work was supported by Google, Inc.

tion of $(2N+1)$ boundary pixels, and obviously $(2N+1)$ multiplications per pixel prediction. Moreover the number of coefficients, all estimated in a spatially adaptive manner, is $N^2(2N+1)$ and increases dramatically with the block dimension. Hence both the codec complexity and the needed locally stationary training data to adapt the coefficients grow fast and pose considerable limitation on practical use.

The proposed intra prediction approach resolves such difficulties, by appealing to the Markov property of the 2-D non-separable model, which ensures that all the available information to predict a pixel is captured by its nearest neighbors, either previously reconstructed *or predicted*. Therefore, the prediction can be recursively calculated as the outcome of a 3-tap filter, starting at the known boundaries and ending at the far end of the block, each pixel requires three multiplications; and every mode is completely characterized by only three coefficients, i.e., the computational cost per pixel does not increase with block size. We note that this property is particularly relevant to the growing importance of compression efficiency at high definition resolutions where larger block sizes are expected, and such complexity considerations are critical. It is experimentally shown that the scheme provides substantially improved prediction compared to the conventional directional modes. Moreover, it complements the hybrid ADST-DCT transform for coding the prediction residual, which was developed by our group [4, 5] from related observations regarding the decay in correlation with distance from the boundary, and achieves significant coding performance gains.

2. THE RECURSIVE PREDICTOR PARADIGM

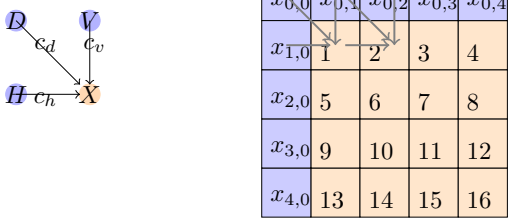
Consider a 2-D non-separable Markov model with zero-mean and unit variance, whose evolution recursion can be written as (see Fig. 1(a)):

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

where V, H, and D are the specified neighbors of X, and ϵ denotes the innovation term. The coefficients c_v, c_h, and c_d effectively capture the correlation gradients in the 2-D space, or the 'directionality' of the image signal. Let $\hat{V}$, $\hat{H}$, $\hat{D}$ denote the codec reconstruction of V, H, and D, respectively. At high bit-rates, the reconstructions closely approximate the original values, and hence the optimal predictor of X is closely approximated by

$$\tilde{X} = c_v \hat{V} + c_h \hat{H} + c_d \hat{D} \,. \tag{2}$$

This predictor is often also used in moderate to low bit-rates. Note that the standard directional prediction modes are degenerate and extreme special cases. For example, when $c_v = 1$, $c_h = 0$, and $c_d = 0$, it degenerates to the Vertical mode and simply copies the above pixel as prediction. Similarly, $c_d = 1$, $c_v = 0$, $c_h = 0$ corresponds to the standard Diagonal Down-Right mode. Clearly, the values of these coefficients effectively control the evolution patterns of the 2-D stochastic process. The proposed predictor will thus employ this unified 3-tap extrapolation filter, with an optimized set of coefficients tailored for each prediction modes, and hence it replaces and generalizes the conventional 'pixel-copying' approach.



(a) "2-D Markov unit" (b) Prediction order

Fig. 1. Recursive extrapolation with a 3-tap filter

Recall that our focus is on block-based video and image compression where, unlike the above differential pulse code modulation (DPCM) setting which predicts each sample from available neighboring reconstructions, the codec must predict the entire block given a set of boundary pixels. As illustrated in Fig. 1(b), the prediction of an inner pixel does not have access to the reconstructions of its immediate neighbors. Further note that unlike the horizontal-vertical separable model, where the optimal predictor of $x_{i,j}$ takes the simple form $c_v^i \hat{x}_{0,j} + c_h^j \hat{x}_{i,0} - c_v^i c_h^j \hat{x}_{0,0}$, the predictor for the general non-separable model potentially involves all available boundary pixels, weighted by coefficients that are fairly complicated to calculate. We circumvent this difficulty by recursive extrapolation. The predicted content of the block is defined as the outcome of the 3-tap filter, starting from the known boundary and recursively applied throughout to the far end of the block. The filter uses reconstructed pixel values when available, and predicted values otherwise. A valid prediction scanning order is specified in Fig. 1(b). How to optimize the filter coefficients is the topic of the next section.

Note that standard intra coders include infrequently used prediction modes that predict from top-right or bottom-left, i.e., the Horizontal-Up, Vertical-Left, and Diagonal-Down-Left modes. To simplify the presentation and implementation, we retained such modes unchanged from the standard (while noting that they can be generalized as well, an extension that would require more space and is hence beyond the scope of this paper).

3. RECURSIVE PREDICTOR DESIGN

We first consider optimal linear predictor design in the simple setting of a known 2-D Gauss-Markov model. Image signals, however, are often better characterized by a mixture of

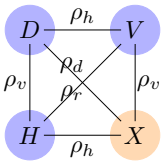


Fig. 2. Neighboring samples' correlation coefficients

unknown Markov models, instead of a single stationary process. Therefore, a variant of K-means clustering is needed to iteratively partition a training set of blocks into clusters (or equivalently, modes), and redesign an optimal filter per cluster/mode, namely, a "K-modes" iterative approach. We then extend the optimization for the ultimate overall rate-distortion criterion.

3.1. Filter Design for a Known 2-D Markov Model

Consider the non-separable 2-D Gauss-Markov model of (1) with a known correlation matrix. The coefficients of the least mean squared predictor are given by

$$\underline{c}_{opt} = [R_{XV} R_{XH} R_{XD}] \begin{bmatrix} R_{VV} & R_{VH} & R_{VD} \\ R_{VH} & R_{HH} & R_{HD} \\ R_{VD} & R_{HD} & R_{DD} \end{bmatrix}^{-1}, \quad (3)$$

where R_{AB} denotes the cross correlation between A and B. Since we assume stationarity, it can be simplified to

$$\underline{c}_{opt} = [\rho_v \quad \rho_h \quad \rho_d] \begin{bmatrix} 1 & \rho_r & \rho_h \\ \rho_r & 1 & \rho_v \\ \rho_h & \rho_v & 1 \end{bmatrix}^{-1}, \quad (4)$$

where ρ_v, ρ_h, ρ_d and ρ_r are the correlation coefficients relating neighboring samples, as shown in Fig. 2.

3.2. K-Mode Iterative Clustering

Images are often well characterized spatially as a mixture of unknown 2-D Markov models. Proper analysis requires separating a training set of image blocks into subsets, each representable by a specific 2-D model, and hence with its optimal intra prediction mode. Therefore, we employ a variant of K-means clustering to iteratively partition the training data into clusters, and re-optimize the prediction filter for each cluster or mode, hereby named the "K-modes iterative approach".

The process can be initialized by partitioning the training blocks according to a standard intra-coder's mode decisions. It then runs the two step iterations: redesign the prediction filter for each subset of blocks, and then repartition the blocks to best utilize the prediction filters, monotonically reducing the mean squared prediction error until convergence.

$FilterDesign$: given the training set partition, for each cluster of blocks, estimate the correlation coefficients

$$\begin{aligned} \rho_v &= \frac{\sum (x_{i,j} - \overline{x})(x_{i+1,j} - \overline{x})}{\sum (x_{i,j} - \overline{x})^2}, \\ \rho_h &= \frac{\sum (x_{i,j} - \overline{x})(x_{i,j+1} - \overline{x})}{\sum (x_{i,j} - \overline{x})^2}, \\ \rho_d &= \frac{\sum (x_{i,j} - \overline{x})(x_{i+1,j+1} - \overline{x})}{\sum (x_{i,j} - \overline{x})^2}, \\ \rho_r &= \frac{\sum (x_{i,j} - \overline{x})(x_{i+1,j-1} - \overline{x})}{\sum (x_{i,j} - \overline{x})^2}. \end{aligned} \quad (5)$$

where $\overline{x}$ denotes the mean value over the block, and plug in (4) to obtain the filter coefficients.

$Partition$: Given a set of mode filters, assign each block to the mode/cluster whose filter minimizes the prediction error.

3.3. Adjustment for Overall Rate-Distortion Optimization

Motivated by the recognition that (i) the signal is not exactly a mixture of (even non-separable) Markov processes; and (ii) minimizing the prediction error is somewhat mismatched with optimizing the overall rate-distortion trade-off, we propose a final phase to adjust the filter coefficients obtained by minimizing the prediction error, while accounting now for the ultimate overall rate-distortion performance.

Integrate the above designed filters in the prediction module, and run the encoding process to obtain the total rate-distortion cost L_{opt}. Let c_i denote the coefficient set. Run the following iterations:

1. increase c_i by Δ, run the encoder to get the rate-distortion cost L.
2. if $L < L_{opt}$, update the minimum cost $L_{opt} = L$ and repeat Step 1. Otherwise, decrease c_i by Δ and continue to Step 3.
3. decrease c_i by Δ, run the encoder to get the rate-distortion cost L.
4. if $L < L_{opt}$, update the minimum cost $L_{opt} = L$ and repeat Step 3. Otherwise, increase c_i by Δ, increase i by one and go to step 1.

Repeat the iterations until convergence. In our experiments we set $\Delta = 0.01$ and the iterations converged after 3 loops over the coefficient set.

4. EXPERIMENTAL RESULTS

The proposed recursive extrapolation approach to intra prediction was implemented within the H.264/AVC and VP8 reference frameworks, to validate its performance gains in conjunction with established coders. We emphasize that it is directly extendable to block operations at larger block size at no additional complexity cost, and hence is applicable to

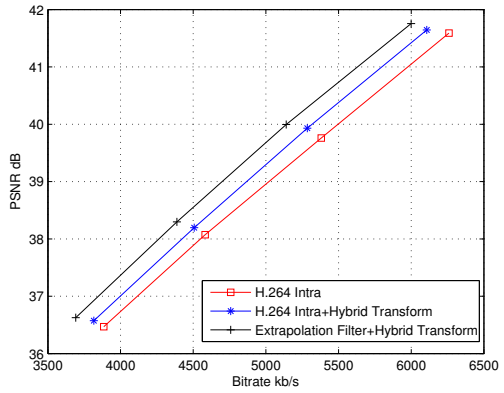


Fig. 3. Coding performance comparison for sequence bus at CIF resolution. The extrapolation filter based intra predictor is implemented in the H.264/AVC Intra Coder, followed by hybrid transform coding scheme for the prediction residuals.

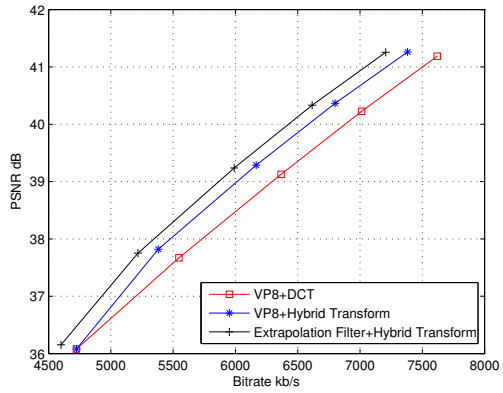


Fig. 4. Coding performance comparison for sequence $tempete$ at CIF resolution. The proposed predictor is implemented in the framework of VP8 Intra Coder, with hybrid transform coding scheme enabled for the residuals.

other emerging codecs for high definition sequences, including HEVC and VP-Next.

The H.264/AVC Intra coder reference had all directional prediction modes of block dimension 4×4 enabled, which were selected in a rate-distortion optimization framework. The residuals were then transformed using a 2-D discrete cosine transform (DCT), quantized and entropy coded via context-adaptive binary arithmetic coding. This reference codec was further modified to employ a hybrid transform coding scheme, earlier developed by our group [5], where it was assumed that regular directional intra prediction roughly approximated optimal prediction, and hence shown that the Karhaunen-Loeve transform for the residuals is a close relative of the discrete sine transform. The transform coding scheme of [5] was experimentally demonstrated to achieve significant coding performance gains, and was essentially adopted as a central component of intra coding in upcoming standard codecs. Our proposed recursive extrapolation filter effectively accounts for the variations of correlations across the pixel block and specifically for non-separability and correlation decay with distance, thereby achieving efficient prediction, which in turn complements the above transform coding approach to approach joint optimality. We evaluate the coding performance of the proposed prediction-transform scheme against the H.264/AVC Intra coder and against the upgraded version of H.264 that includes hybrid transform coding, in Fig. 3. Clearly, the proposed approach achieves consistent performance gains over its competitors. Experiments on other test sequences corroborate this observation with similar comparison results (Table 1). Note also that all simulation results in this paper are for test sequences that were excluded from the predictor training data.

The proposed approach was also tested in the VP8 framework, where only 4×4 block prediction modes were enabled, and replaced with our proposed extrapolation filters. Similar performance gains were obtained as shown in Fig. 4.

Table 1. Reduction in bit-rate due to the recursive extrapolation approach, relative to the H.264 Intra coder with hybrid transform coding. The test sequences are at CIF resolution.

bit savings (%)	PSNR (dB)		
Test Sequence	42	38	34
bus	3.26	3.67	4.53
$harbour$	2.42	3.19	3.08
$coastguard$	3.38	3.43	3.53
$bridge-close$	1.85	2.41	3.38
$tempete$	2.79	2.17	3.12

5. CONCLUSION

A novel prediction scheme, based on recursive extrapolation filters, is proposed for intra coding. It recoups the benefits of a 2-D non-separable model, while circumventing the complexity cost, by employing a unified 3-tap extrapolation filter. The requisite coefficients are obtained off-line using a "K-modes" iterative approach, to directly minimize the overall rate-distortion cost. Exploiting the Markov property, the prediction is effectively implemented by recursively applying the 3-tap filter throughout the block, thereby maintaining fairly low computational complexity. The prediction scheme complements our hybrid transform coding approach, to achieve consistent coding performance gains.

6. REFERENCES

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding stan-

dard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[2] L. Liu, Y. Liu, and E. Delp, "Enhanced intra prediction using context-adaptive linear prediction," *Proceedings of the Picture Coding Symposium*, Nov 2007.

[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, pp. 687–696, Nov 2011.

[4] J. Han, A. Saxena, and K. Rose, "Towards jointly optimal spatial prediction and adaptive transform in video/image coding," *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 726–729, 2010.

[5] J. Han, A. Saxena, V. Melkote, and K. Rose, "Jointly optimized spatial prediction and block transform for video and image coding," *IEEE Transactions on Image Processing*, vol. 21, pp. 1874–1884, April 2012.