LEAST-SQUARES APPROXIMATION SURFACES FOR HIGH QUALITY INTRA-FRAME PREDICTION IN FUTURE VIDEO STANDARDS

[1]Ramon Fernandes,[1]Gustavo Sanchez,[1]Rodrigo Cataldo,[2]Luciano Agostini,[1]César Marcon

[1]Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)
[2]Universidade Federal de Pelotas (UFPeL)

ABSTRACT

This paper proposes the Least-Squares Approximation Surfaces (LSAS) for achieving a high intra-frame prediction quality in future video coding standards. LSAS approximates the encoding block by a mathematical expression, whose coefficients are transmitted together with each frame of the video. Software evaluations demonstrate that using a second order polynomial obtained the best tradeoff between PSNR and encoding effort. Its evaluation shows an average PSNR gain from 0.8 dB to 5.0 dB higher than the High-Efficiency Video Coding (HEVC) intra-frame prediction, according to the encoded block size. As a counterpart, the LSAS encoding time varies between -47% and 108% compared to HEVC. LSAS can also be combined with HEVC intra-frame prediction, allowing an average increase in PSNR of 2.9 dB with an increase in the computational effort of 215% compared to the original HEVC intra-frame prediction.

***Index Terms*—Intra-frame prediction, HEVC, Video Coding, Image Approximation.**

I. INTRODUCTION

Several digital video applications have arisen in the past few years. Among these applications, there is a demand for high quality and high definition videos for streaming over the internet, storage on disk, and others. Therefore, a high effort was spent in the standardization of modern video encoders such as the High Efficiency Video Coding (HEVC) [1], VP9 [2], Audio Video Coding Standard 2 (AVS2) [3], and the next emergent video coding technologies [4][5].

These standards employ intra-frame and inter-frame predictions. The intra-frame prediction explores only the spatial information of neighboring blocks [6]. Meanwhile, the inter-frame prediction is based on redundant information in past-encoded frames [7]. After generating the prediction, these encoders transform and quantize the prediction error (referred to as the prediction residuals) of the encoding process. Finally, the quantized residuals and the high-level syntax elements are entropy encoded and packed into the bitstream, finishing the encoding process.

The prediction process plays a crucial role in the encoded video quality and encoded stream size; therefore, it is essential to investigate efficient prediction approaches for the next generation video coding standards. Since this work focus on intra-frame prediction, the remaining of this paper is strictly related to this subject. Moreover, we considered the HEVC structure as the baseline for the remaining explanations.

Some works already proposed new prediction methods. In [8], the authors proposed Sparse Least Squares Prediction (SLSP), a prediction algorithm based on linear equations, which serves as an additional directional intra-frame prediction mode for HEVC. Their algorithm specializes in complex textured areas, adopting a training window to adjust prediction coefficients, present in both encoder and decoder, to predict sample values. Results illustrate Peak Signal-to-Noise Ratio (PSNR) gains of up to 1 dB across all test sequences, with a bitrate reduction of at most 15%, at the cost of an order of magnitude increase in computational complexity. The work of [9] uses neighboring blocks to predict sample values through a linear equation, based on a Markovian model. A training window utilizes previously decoded samples to predict new samples. The training window varies according to the sample location in the image, predicting samples through least-squares computation. Their results obtained an average increase of 0.41 dB and an average decrease in bitrate of 7.34% when compared to H.264 [10], albeit at nearly 150% increase in the encoding and decoding times.

Although these works can obtain reliable predictions of the encoding blocks, and consequently a high encoding efficiency, better results can be achieved if a more sophisticated prediction is designed. Therefore, this paper proposes the Least-Squares Approximation Surfaces (LSAS) intra-frame prediction for allowing a higher prediction quality. The remaining of this paper is divided as follows. Details of the HEVC intra-frame prediction structure are discussed in Section II. Section III explains our algorithm proposal. The evaluation results are displayed and discussed in Section IV, and Section V renders the conclusion of this paper.

II. HEVC INTRA-FRAME PREDICTION

The HEVC intra-frame prediction exploits the spatial redundancy to compress video information using the high correlation of nearby samples (pixels) in a frame [11]. To efficiently predict different kinds of content, HEVC employs a range of sample prediction methods, divided into two classes [6]: (i) angular prediction, for modeling images with directional edges; and (ii) planar and DC prediction, used for smooth image content.

978-1-5386-9562-3/18/$31.00 ©2018 IEEE

HEVC intra prediction uses reference samples from adjacent reconstructed blocks (above and left). Angular prediction employs 33 directions, shown in Fig. 1, using reference samples to interpolate pixel values in a predicted block. Although it is suitable for approximating blocks with several edges, the angular prediction may result in blockiness or visible contouring in smooth image areas. Planar prediction overcomes these issues by creating a prediction surface without discontinuities at the block boundaries, while DC prediction averages the sample values to populate the predicted block with constant values.

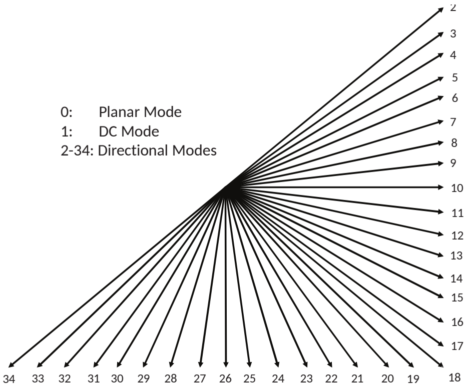

Fig. 1. Intra-frame predictions modes in HEVC.[6]

To choose the best prediction mode, HEVC recursively evaluates all levels inside a quadtree structure, where the block size ranges from 64×64 to 4×4. All 35 prediction modes are evaluated for each block size employing the Rough Mode Decision (RMD) algorithm [12], which uses the Sum of Absolute Transformed Differences (SATD) to predict without needing the complete Rate-Distortion (RD) cost. The modes with the lowest cost are inserted into an RD-list, to be later evaluated, along with the Most Probable Modes (MPM) chosen according to the neighboring blocks.

The HEVC intra-frame prediction has high computational effort since all modes are recursively tested to find the lowest approximation cost. Meanwhile, the approximation still incurs in residuals that are transformed and quantized, incurring signal degradation when decoded. Unlike HEVC intra-prediction, our proposal requires evaluating a block only once, allowing to improve the encoding performance.

III. LEAST-SQUARES POLYNOMIAL SURFACES APPROXIMATION

The LSAS prediction estimates the sample values in prediction blocks by deriving the coefficients of a mathematical function, which approximates the sample distribution in the block. Given an $n \times n$ block, where a pair of xy coordinates determines each sample in the block. Equation 1 shows the bivariate polynomial assumed by LSAS model as the surface equations. The polynomial order h can be adjusted to better approximate the function, although this incurs in more coefficients to be calculated.

$$f(x,y) = \sum_{i=0}^{h} \sum_{j=0}^{h} a_k x^i y^j \quad | \quad 0 \le k \le 2h \tag{1}$$

The coefficients of each bivariate polynomial are calculated employing a system of linear equations, whose approximation function has $Ax = b$ format regarding both xy coordinates. Besides, solving a linear equation requires defining an error function. Equation 2 illustrates an order 1 bivariate polynomial represented by three coefficients, where z_{ij} corresponds to the measured sample value and $(a_0 + a_1 x_i + a_2 y_j)$ to the surface prediction. Regarding an image, x and y correspond to the pixel coordinates, while z to the pixel value. Solving the coefficients produces a fitness function that approximates samples in an image region.

$$\begin{gathered} E(a_0, a_1, a_2) = \\ \sum_{i,j=0}^{n} \left[(a_0 + a_1 x_i + a_2 y_j) - z_{ij} \right]^2 \end{gathered} \tag{2}$$

Since the linear equations system has way more equations than coefficients due to the low order polynomials, we multiply A by its transpose A^T, and solve the linear equation $A^T A \hat{x} = A^T b$, with $\hat{x}$ as the least squares solution [13]. Applying Equation 2 to the linear system $A^T A \hat{x} = A^T b$ results in three linear equations which can be easily solved.

$$\begin{bmatrix} \sum_{i,j=0}^{n} 1 & \sum_{i,j=0}^{n} y_j & \sum_{i,j=0}^{n} x_i \\ \sum_{i,j=0}^{n} y_j & \sum_{i,j=0}^{n} y_j^2 & \sum_{i,j=0}^{n} x_i y_i \\ \sum_{i,j=0}^{n} x_i & \sum_{i,j=0}^{n} x_i y_i & \sum_{i,j=0}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum_{i,j=0}^{n} z_{i,j} \\ \sum_{i,j=0}^{n} y_i z_{i,j} \\ \sum_{i,j=0}^{n} x_i z_{i,j} \end{bmatrix} \tag{3}$$

In this work, the system of linear equations 3 is solved through the Gauss-Seidel [14] iterative method until achieving a satisfactory convergence for the coefficients. According to experimental analysis, no more than twenty iterations were necessary in all cases to compute a solution.

The proposed technique adapts well to image regions with curved edges inside a prediction block. Still, a side effect of using polynomial approximation surfaces is the smoothing of image areas in neighboring samples close to a high contrasting region. Fig. 2a illustrates an image region with significant contrast variation. Fig. 2b shows the approximation computed by HEVC combing all 35 modes to predict the image region better. Meanwhile, our proposed prediction technique is shown in Fig. 2c. The figures illustrate the HEVC prediction produces sharper and well-defined edges, while our approximation technique produces smoother surfaces.

IV. SETUP & RESULTS

We designed a framework containing HEVC intra and LSAS prediction to allow the evaluation and fair comparison of these techniques. In the designed framework, each encoding block size is evaluated alone, and the Sum of Absolute Differences (SAD) is responsible for selecting the best encoding prediction among the available ones. As a result, the framework computes the PSNR and the encoding time of the prediction technique.


(a) Original (b) HEVC Intra (c) Proposed

Fig. 2. Comparison of prediction techniques: a) original image; b) HEVC intra-frame prediction; c) proposed technique.

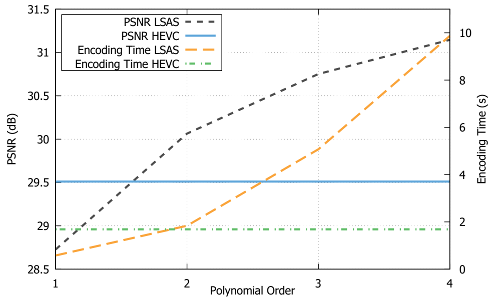

Fig. 3. Selection of best LSAS polynomial order for all block sizes, in average - HEVC values are exposed for comparison.

Section IV-A describes the LSAS training with five videos randomly selected among 17 sequences to be used in the polynomial order definition. Section IV-B evaluates the training quality employing the remaining 12 videos.

IV-A. Training

The five high definition videos (i.e., *Tennis*, *Rolling*, *Man_in Car*, *Basketball_Drive*, *BQ_Terrace*) randomly selected for training were evaluated according to the PSNR and encoding time using (i) LSAS algorithm, varying its polynomial order from 1 up to 4; and (ii) HEVC intra-frame prediction modes. Fig. 3 displays the average results encompassing block sizes from 4×4 to 64×64.

One can notice that for a first-order polynomial, the PSNR obtained by the solution is almost 1 dB lower than HEVC, while the encoding time is half of HEVC. For the second order polynomial, the PSNR gets more than a 0.5 dB increase over HEVC, while the encoding time is a little higher. When the polynomial order is still increased, it starts being saturated in the PSNR axis, while the encoding time grows exponentially. Therefore, we have selected the second order polynomial for block sizes as the best operation point, which is evaluated in the next subsection.

IV-B. Results

Using the best polynomial order (i.e., second order) obtained from our analysis in Section IV-A, we have encoded 12 videos employing LSAS and HEVC intra-frame prediction. The list of encoded videos is displayed in Table I. It is important to note that videos used in the training set were not evaluated here to avoid having over-trained results. The box plot of the PSNR results and the encoding time per frame are presented in Fig. 4 and Fig. 5, respectively.

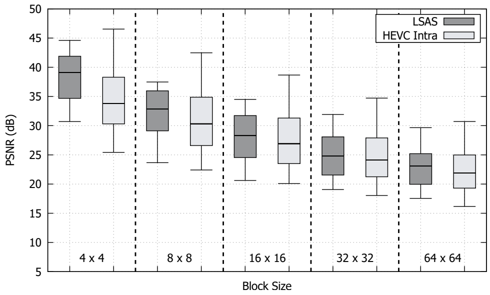

Fig. 4. PSNR comparison of LSAS and HEVC intra-frame prediction for all block sizes.

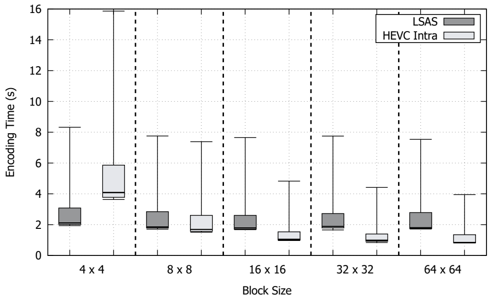

Fig. 5. Encoding time comparison of LSAS and HEVC intra-frame prediction for all block sizes.

Fig. 4 shows that LSAS obtains better PSNRs than HEVC intra-frame prediction for all available block sizes. LSAS achieves gains in the order of 5 dB for lower block sizes; however, for larger blocks, the gain starts getting smaller, reaching an average gain of 0.77 dB for 32×32 blocks, which is the worst result.

Fig. 5 displays that LSAS implies in lower encoding time for 4×4 blocks, similar encoding time for 8×8 blocks and higher encoding time for the remaining available blocks.

Further PSNR enhancement is achieved with L-HEVC, a mixed system that selects LSAS or HEVC intra predictions according to the SADs results per encoded block. To illustrate these enhancements, we employed both algorithms for evaluating the encoding of 12 videos. Table I shows results of the PSNR variation and the percentage of increase in encoding time for all block sizes when comparing LSAS to L-HEVC - results are presented in percentage normalized to the basic HEVC intra-frame prediction.

Notice that L-HEVC gains are higher for smaller block sizes, reaching up to 6.16 dB more than HEVC intra-frame prediction. The lowest gain is measured for 64×64 blocks when 1.32 dB gain is obtained. Although these gains seem small, it is important to emphasize that the PSNR is measured on a logarithmic scale. Therefore, more than 1 dB, as obtained by L-HEVC for all available block sizes, is very significant. In

Table I. PSNR and encoding time results for LSAS and L-HEVC compared to the base HEVC intra-frame prediction.

Video	Block Size																			
	4 × 4				8 × 8				16 × 16				32 × 32				64 × 64			
	PSNR (dB)		Encoding time		PSNR (dB)		Encoding time		PSNR (dB)		Encoding time		PSNR (dB)		Encoding time		PSNR (dB)		Encoding time	
	LSAS	L-HEVC	LSAS	L-HEVC	LSAS	L-HEVC	LSAS	L-HEVC	LSAS	L-HEVC	LSAS	L-HEVC	LSAS	L-HEVC	LSAS	L-HEVC	LSAS	L-HEVC	LSAS	L-HEVC
Traffic	+8.35	+8.61	-47.92%	+87.99%	+4.21	+4.70	+13.93%	+185.55%	+2.15	+2.57	+70.28%	+261.35%	+1.33	+1.60	+92.32%	+287.56%	+1.29	+1.44	+110.77%	+326.16%
Tractor	+6.28	+6.65	-48.17%	+83.04%	+3.59	+4.35	+10.86%	+167.14%	+2.18	+2.82	+71.40%	+250.51%	+1.45	+1.86	+90.72%	+275.20%	+1.18	+1.41	+105.85%	+304.87%
Sunflower	+8.20	+8.43	-47.03%	+85.56%	+5.83	+6.45	+8.82%	+172.20%	+3.90	+4.44	+73.43%	+269.41%	+2.17	+2.45	+90.64%	+279.10%	+1.06	+1.06	+113.05%	+292.90%
Steam Locomotive Train	+5.41	+5.62	-47.17%	+90.42%	+1.76	+2.41	+9.72%	+174.12%	+1.18	+1.95	+70.66%	+264.95%	+1.27	+1.99	+99.13%	+292.35%	+1.59	+2.13	+106.76%	+304.63%
Riverbed	+5.51	+5.89	-47.04%	+87.91%	+2.86	+3.56	+10.43%	+171.34%	+1.60	+2.20	+71.76%	+256.99%	+1.00	+1.38	+93.89%	+280.74%	+1.02	+1.20	+113.29%	+322.87%
Park Joy	+5.36	+5.41	-47.96%	+84.50%	+2.39	+2.45	+15.37%	+184.69%	+1.45	+1.48	+69.87%	+258.08%	+1.14	+1.11	+97.86%	+291.11%	+1.09	+1.02	+115.10%	+322.99%
Cactus	-0.65	+6.39	-46.68%	+89.41%	-2.79	+3.31	+8.61%	+166.23%	-2.25	+2.13	+58.62%	+231.50%	-2.24	+1.21	+72.95%	+238.78%	-1.24	+0.90	+90.64%	+270.31%
Coastguard	+3.30	+6.73	-47.54%	+88.32%	-0.52	+2.67	+5.92%	+160.55%	-1.13	+1.63	+58.92%	+215.74%	-1.30	+1.11	+79.89%	+222.98%	-0.72	+1.00	+93.84%	+238.95%
Crowd Run	+4.80	+5.02	-46.70%	+88.78%	+2.30	+2.50	+14.63%	+181.44%	+1.72	+1.77	+70.48%	+261.30%	+1.45	+1.38	+89.44%	+282.40%	+1.64	+1.51	+106.64%	+311.69%
Ducks take-off	+3.46	+3.92	-47.57%	+85.35%	+0.75	+1.74	+11.19%	+162.24%	+0.05	+0.99	+69.46%	+226.30%	+0.14	+0.67	+95.25%	+252.09%	+0.53	+0.64	+112.46%	+291.80%
Cactus	+4.37	+5.00	-46.74%	+85.98%	+1.94	+2.65	+13.05%	+174.96%	+1.44	+1.77	+72.94%	+248.08%	+1.31	+1.45	+93.28%	+257.68%	+1.73	+1.67	+107.42%	+280.99%
Kimono	+6.02	+6.22	-46.43%	+86.66%	+4.19	+4.76	+13.50%	+178.81%	+2.41	+3.07	+71.35%	+259.72%	+1.52	+2.13	+90.68%	+272.85%	+1.31	+1.85	+115.34%	+323.35%
Average	+5.03	+6.16	-47.25%	+86.99%	+2.21	+3.46	+11.34%	+173.27%	+1.22	+2.24	+69.10%	+250.33%	+0.77	+1.53	+90.50%	+269.40%	+0.87	+1.32	+107.60%	+299.29%

fact, our solution is capable of achieving a gain of 2.9 dB, on average. However, L-HEVC consumes 215% higher encoding time, on average, since it requires evaluating both LSAS and HEVC intra-frame prediction.

V. CONCLUSION AND FUTURE WORK

This paper presented the Least-Squares Approximation Surfaces (LSAS) encoding technique aiming to enhance the intra-frame prediction quality in future video coding standards. The LSAS models the encoding block by a mathematical surface, requiring the transmission of its coefficients, along with smaller residues than HEVC intra-frame prediction. According to our evaluations, the second order polynomial obtained the best tradeoff between encoding effort and PSNR. Therefore, it was possible to achieve an average gain in PSNR of 2 dB with an increase in the computational effort of 46.3%, considering the average results of all block sizes. Besides, we designed a mixed system, called L-HEVC, which combines LSAS with HEVC intra-prediction, allowing an increase in PSNR of 2.9 dB with an encoding time increase of 215%. As future work, we plan to design LSAS inside the HEVC reference software and model a prediction for LSAS coefficients.

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

VI. REFERENCES

[1] G. 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] D. Mukherjee, J. Bankoski, A. Grange, J. Han, J. Koleszar, P. Wilkins, Y. Xu, and R. Bultje, “The latest open-source video codec vp9-an overview and preliminary results,” in *Picture Coding Symposium (PCS), 2013*. IEEE, 2013, pp. 390–393.

[3] Z. He, L. Yu, X. Zheng, S. Ma, and Y. He, “Framework of avs2-video coding,” in *Image Processing (ICIP), 2013 20th IEEE International Conference on*. IEEE, 2013, pp. 1515–1519.

[4] T. Ebrahimi, S. Foessel, F. Pereira, and P. Schelkens, “Jpeg pleno: Toward an efficient representation of visual reality,” *IEEE Multimedia*, vol. 23, no. 4, pp. 14–20, 2016.

[5] G. Sullivan J.-R. Ohm J. Boyce J. Chen, E. Alshina, “Algorithm description of joint exploration test model 6 (jem 6),” in *JVET-F1001*, 2017.

[6] J. Lainema, F. Bossen, W.-J. Han, J. Min, and K. Ugur, “Intra coding of the hevc standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1792–1801, 2012.

[7] C. Rhee, K. Lee, T. Kim, and H.-J. Lee, “A survey of fast mode decision algorithms for inter-prediction and their applications to high efficiency video coding,” *IEEE Transactions on Consumer Electronics*, vol. 58, no. 4, 2012.

[8] L. Lucas, N. Rodrigues, C. Pagliari, E. da Silva, and S. de Faria, “Sparse least-squares prediction for intra image coding,” in *2015 IEEE International Conference on Image Processing (ICIP)*, Sept 2015, pp. 1115–1119.

[9] D. Garcia and R. de Queiroz, “Least-squares directional intra prediction in h.264/avc,” *IEEE Signal Processing Letters*, vol. 17, no. 10, pp. 831–834, Oct 2010.

[10] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the h.264/avc video coding standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[11] G. Sullivan and T. Wiegand, “Rate-distortion optimization for video compression,” *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 74–90, Nov 1998.

[12] L. Zhao, L. Zhang, S. Ma, and D. Zhao, “Fast mode decision algorithm for intra prediction in hevc,” in *Visual Communications and Image Processing (VCIP), 2011 IEEE*. IEEE, 2011, pp. 1–4.

[13] C. Lawson and R. Hanson, *Solving least squares problems*, SIAM, 1995.

[14] S. Yoon and A. Jameson, “Lower-upper symmetric-gauss-seidel method for the euler and navier-stokes equations,” *AIAA journal*, vol. 26, no. 9, pp. 1025–1026, 1988.