MULTIPLE LINEAR REGRESSION FOR HIGH EFFICIENCY VIDEO INTRA CODING

Zhaobin Zhang* *Yue Li*† *Li Li** *Zhu Li** *Shan Liu*†*

* University of Missouri Kansas City
†University of Science and Technology of China
†*Tencent America

ABSTRACT

In video coding frameworks, the essence of intra coding is leveraging the spatial correlation within a frame to remove redundancy thus achieving compact transmitting data. With modern video acquisition devices improvement, more high-definition videos emerge into people's lives which has set a new challenge for high efficiency video coding. In this paper, we propose a novel intra video coding scheme based on Multiple Linear Regression (MLR), named Multiple linear regression Intra Prediction (MIP). Instead of predicting pixel values by extrapolating, we try to exploit the potential capability of homogeneous regression method. The proposed method has a very concise and neat design yet achieves better performance compared with High Efficiency Video Coding (HEVC) reference software anchor. The experimental results demonstrate the effectiveness of the proposed method and provide interesting insights for further exploiting the capability of conventional algorithms for video coding when many people favor deep learning-based approaches.

Index Terms— Multiple Linear Regression, Intra Prediction, Video Coding

1. INTRODUCTION

Images and videos contain huge information to maintain satisfactory visual quality. It is impossible to transmit such huge multimedia data through the current Internet and wireless networks with limited bandwidth. Due to the redundancy existing in these pixel values, researchers have proposed tons of algorithms to remove the redundant information. High Efficiency Video Coding (HEVC) [1] standards have been state-of-the-art video coding framework for recent years. Techniques like more prediction modes and Rate Distortion Optimization (RDO) enable HEVC a great improvement compared with H.264/AVC [2], the previous generation video coding standards. HEVC supports 35 intra prediction modes including 33 angular prediction modes, a Planar mode and another DC mode. Planar mode is devised for large-area flatten blocks, DC mode is for slow-changing areas and angular modes are designed for areas with directional patterns. Boundary pixels are utilized to estimate current block pixel values. The essence of HEVC intra prediction is how to make a better prediction given already known pixels. Computational complexity and estimation performance have to be appropriately balanced during this procedure.

On top of HEVC intra prediction schemes, various algorithms have been proposed to make a better prediction. Related work can be divided into two categories, conventional methods and deep learning-based methods. For the first category, Y. Li [3] proposed to jointly predict current block by combining Intra Block Copy [4] and current directional prediction scheme. C. Noel [5] developed a regression-based intra prediction scheme by iteratively refining predictors through regularized regression. Meanwhile, motivated by the great success of deep learning in computer vision tasks, some researchers start to apply deep learning to video coding. J. Li [6] designed a fully-connected network to further exploit pixel correlations for intra prediction. Y. Li [7] proposed a Convolutional Neural Network (CNN) for HEVC intra coding following a down-sample up-sample pipeline.

However, neither conventional methods nor deep learning-based methods are the ideal solutions. Conventional methods often involve complicated arithmetic derivation. Deep learning-based methods are often accompanied by higher time complexity. Current HEVC intra prediction solution is not able to capture the rich texture information by one-time interpolation. In this paper, we propose a concise design based on multiple linear regression which allows for further exploiting the correlation between known pixels and the current block. The proposed model takes both the boundary reference pixels and the best intra prediction after Rate-distortion optimization (RDO) as inputs. It is expected to make a better estimation by a secondary exploration. To refine the model, the prediction mode information is also taken into consideration. The manifold structure of this high dimensional input is hierarchically partitioned by grouping the input into several modes according to the dominant texture orientation. In addition, instead of predicting in the transform domain which could induce information loss during dimensionality reduction, we apply MLR directly in the pixel domain. The proposed method is integrated into HEVC reference software. The experimental results yield interesting coding gain and provide possible research directions in the future.

The remainder of this paper is organized as follows. A brief review of HEVC intra prediction framework and a theoretical introduction of MLR are given in Section II. The detailed description of the proposed method is described in Section III. Experimental validation and analysis are presented in Section IV. Finally, Section V makes the conclusion.

2. RELATED WORK

In this section, we will briefly review HEVC current intra coding schemes as well as giving the theoretical description of Multiple Linear Regression.

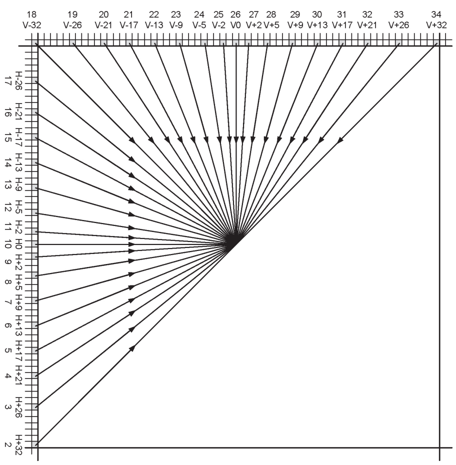

Fig. 1. HEVC angular intra prediction modes numbered from 2 to 34 and the associated displacement parameters.

2.1. HEVC Intra Coding

HEVC inherits block-based scheme from previous video coding frameworks with the main difference of Coding Tree Units (CTU) concept. HEVC intra coding is based on spatial interpolation of samples from previously decoded image blocks. It supports up to 35 intra prediction modes named planar, DC and 33 angular prediction modes as shown in Table 1.

The left column and top row neighboring pixels are utilized to predict the current block. The planar prediction mode first interpolates the bottom right pixel and the other pixels are interpolated by the bilinear method. The DC mode makes the prediction by averaging the top row and the left column. Figure 1 depicts the 33 angular prediction modes numbered from 2 to 34. The current block is predicted with the boundary pixels through interpolation. Motivated by the success in [8] by incorporating piecewise linear projection, this paper aims at investigating the potentiality of combining interpolation and multiple linear regression.

Table 1. Specification of intra prediction modes and associated names

Intra Prediction Mode	Associated Names
0	Planar
1	DC
2..34	Angular (N), N = 2..34

2.2. Multiple Linear Regression

Multiple linear regression (MLR) is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression (MLR) is to model the relationship between the explanatory and response variables. The following model is a multiple linear regression model with k predictor variables, $x_1, \ldots, x_k$.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_k + \epsilon \tag{1}$$

where ϵ is the error term, β_0 is the intercept, β_1-β_k are partial regression coefficients, e.g., β_i when $1 \leq i \leq k$ represents the change in the mean response corresponding to a unit change in x_i when the other variables are held constant. The objective of MLR is to solve for the coefficient set $\Theta = \{\beta_0, \beta_1, \ldots, \beta_k\}$ given observations X and targets Y.

Lease squares is often used to solve the MLR problem.Suppose each predictor variable $x_1, x_2, \ldots, x_k$ has n observations. Then x_{ij} represents the ith observation of the jth predictor variable x_j. For example, x_{51} represents the first value of the fifth observation. Specifically, the previous equation can be expressed as:

$$y_j = \beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \cdots + \beta_k x_{jk} + \epsilon_j \tag{2}$$

where $1 \leq j \leq n$, y_j is the jth target value. The system of n equations can then be represented in matrix notation as follows:

$$y = X\beta + \epsilon \tag{3}$$

The matrix is referred to as the design matrix. It contains information about the levels of the predictor variables at which the observations are obtained. The vector β contains all the regression coefficients. To obtain the regression model, β is estimated using the least square estimates as expressed below.

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{4}$$

Then the estimated value of y can be calculated as follows after $\hat{\beta}$ is obtained.

$$\begin{aligned} \hat{y} &= X\hat{\beta} \\ \epsilon &= y - \hat{y} \end{aligned} \tag{5}$$

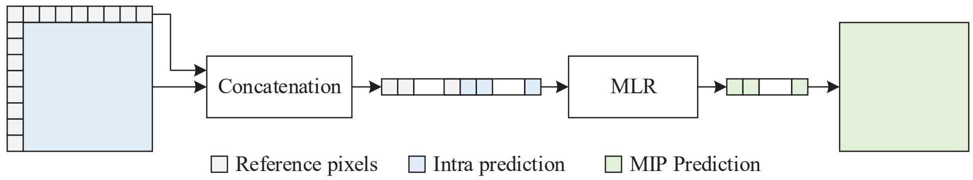

Fig. 2. Framework of proposed Multiple linear regression Intra Prediction (MIP) scheme. Left and top boundary pixels as well as best intra prediction block are utilized to fit the MLR model.

3. MLR FOR INTRA CODING

In this section, we first introduce the proposed scheme along with our motivation. Then the way of how the proposed prediction scheme is integrated into HEVC reference software is detailed.

3.1. Framework of MIP

The proposed scheme architecture is depicted in Fig. 2. As the best intra prediction block is obtained by Rate Distortion Optimization (RDO), it contains abundant detailed information of the current block. However, there is still some information that interpolation is not able to capture. Multiple linear regression is trying to predict the current block on top of best intra prediction by combining interpolating and linear regression.

In HEVC implementation, a total of $4N + 1$ reference pixels are incorporated to predict current $N \times N$ block, the top row, top right row, left column and left bottom column. Since the top right row and left bottom column might not be available sometimes and filling non-existing pixels will induce additional distortion to the MIP model, we only leverage the top row and the left column, making the reference pixels dimension of $2N + 1$. The reference pixels and the best intra prediction will be flattened and concatenated together and fed into the MLR model. The target is the ground truth block of $N \times N$. We denote the concatenated reference pixels and the intra prediction as X, the target block as Y, the MIP prediction as $\hat{Y}$. The training process is trying to minimize the following loss function.

$$L = \| Y - \hat{Y} \|_2^2, \quad \hat{Y} = XA + b \tag{6}$$

Due to the rich content in natural video sequences, it is not capable of capturing the structure with a single prediction mode. Therefore, HEVC developed multiple intra prediction modes, e.g., Planar, DC and 33 angular modes. The similar idea has been adopted in this work. Separate MIP model is trained for each mode. MIP modes are designed according to intra prediction modes. Other possible classification schemes are left for future investigation. Separate model is trained for Planar and DC mode due to their special texture characteristics. Since the neighboring angular modes share a lot of similarities, we combine 3 adjacent angular modes into a single MIP angular mode to avoid the singularity. To be specific, the relationship between MIP prediction mode and HEVC intra prediction mode is expressed as follows.

$$m = \begin{cases} 0, & \text{if } n = 0 \\ 1, & \text{if } n = 1 \\ floor(\frac{n-2}{3}) + 2, & \text{if } n > 1 \end{cases} \tag{7}$$

where n is HEVC intra prediction mode index and m is MIP mode index. Therefore, there are in total 13 MIP modes.

3.2. Integration into HEVC

As shown in Fig. 3, input video signal is performed by directional prediction first, e.g., HEVC existing intra prediction scheme which consists of 35 modes. Then it will be processed by the proposed method and RDO is followed to find the optimal prediction mode. An additional bit flag is encoded to indicate whether MIP is selected.

4. EXPERIMENTS

The proposed MIP method is implemented in HEVC reference software 16.0 [9] and its anchor is used for comparison. All the experiments are conducted under HEVC common test condition All Intra (AI) configuration. Quantization Parameter (QP) is set to {22, 27, 32, 37}. HEVC common test sequences are used to perform the evaluation, consisting of 5 classes with resolutions ranging from 720p to 4k. To save simulation time, only the first frame is used for each sequence.

4.1. Training data derivation

DIV2K 2K resolution high-quality image dataset is used to generate the training dataset. DIV2K contains 800 2K training images, 100 validation images and 100 testing images coving a wide range of contents.

To obtain the best intra prediction, we encode the training samples with HEVC reference software and extract the best intra prediction from the bitstream. The boundary reference pixels are extracted from the reconstructed data. To classify the training samples into local groups according to their directional information, the intra prediction direction is saved. Since we will train a separate set of MIP models for each QP

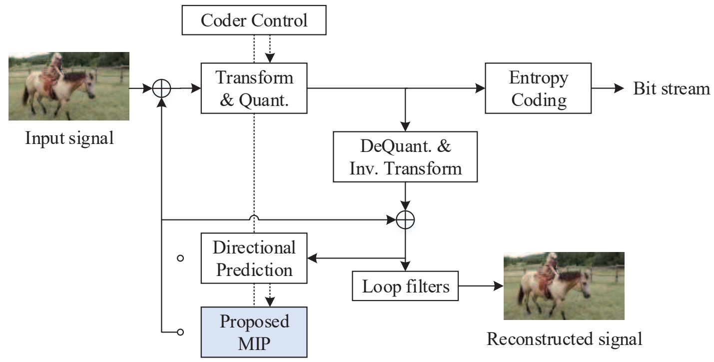

Fig. 3. The proposed MIP is added as additional intra prediction mode and plugged after existing HEVC intra prediction.

Table 2. The BD-Rate results of the proposed method.

Sequence		BD-Rate Y	BD-Rate U	BD-Rate V
Class A	Traffic	-0.9%	-0.6%	-1.3%
	PeopleOnStreet	-0.5%	0.0%	0.1%
	Nebuta	-0.9%	-0.8%	-0.7%
	SteamLocomotive	-0.6%	-0.3%	-0.3%
Class B	Kimono	-0.6%	-1.5%	-1.1%
	ParkScene	-0.6%	-1.1%	-1.0%
	Cactus	-0.7%	-0.3%	-1.1%
	BQTerrace	-0.4%	-1.3%	-0.6%
	BasketballDrive	-1.0%	0.2%	-1.0%
Class C	BasketballDrill	-0.1%	-2.0%	-0.8%
	BQMall	-0.2%	0.1%	-0.4%
	PartyScene	-0.1%	-0.2%	-0.7%
	RaceHorsesC	-0.4%	-0.2%	-0.7%
Class D	BasketballPass	0.1%	1.1%	-0.8%
	BQSquare	0.1%	0.1%	-1.1%
	BlowingBubbles	0.4%	-1.3%	-1.4%
	RaceHorses	-0.3%	-1.6%	0.6%
Class E	FourPeople	-0.4%	0.3%	-1.6%
	Johnny	-0.3%	-1.1%	-1.3%
	KristenAndSara	-0.5%	-0.4%	-0.6%
Class A		-0.7%	-0.5%	-0.5%
Class B		-0.6%	-0.8%	-1.0%
Class C		-0.2%	-0.6%	-0.8%
Class D		0.1%	-0.4%	-0.7%
Class E		-0.4%	-0.4%	-1.2%
Average		-0.4%	-0.6%	-0.8%
Enc Time		487%		
Dec Time		154%		

and block size combination, and there are 13 transformations in each set. Therefore, there are in total $4 \times 4 \times 13 = 208$ MIP models.

4.2. Results and Analysis

The BD-Rate reduction of all the test sequences is summarized in Table 2. An average of -0.4%, -0.6%, and -0.8% BD-Rate saving is achieved for Luma and two Chroma components respectively by the proposed method. It is noticeable that MIP performs better on high-resolution sequences than on low-resolution sequences. As high as -0.9% BD-Rate reduction is observed on *Traffic* from Class A while only -0.2% BD-Rate saving is achieved on Class C. It can also be seen that the proposed method performs consistently better on chroma components across different resolution test sequences. As the training is off-line and all the transformations and bias matrices are saved thus not much encoding time is increased. In contrast, encoding time criterion is not as crucial as decoding time. In addition, the intra coding only occupies only a small fraction of the total encoding time in video coding, and the increase of encoding complexity is less significant in the whole video coding system.

5. CONCLUSION

This paper proposes a new method based on multiple linear regression (MLR) for high efficiency video intra coding. The proposed method accepts both the boundary reference pixels and the best intra prediction block as inputs and trys to learn an end-to-end projection using a linear regression model. The model is constructed directly in the pixel domain instead of in the transformation domain which requires more arithmetic computation. To refine the model, separate models are trained according to their intra prediction directional information. This guarantees that the linear regression model can capture more useful correlations from the directional patterns. It shares a very neat and concise design yet achieves promising performance. The proposed method is integrated into HEVC reference software and performs consistently better than its anchor with all intra configuration. These observations provide useful evidence on exploiting applicable techniques for next generation video coding standards.

6. REFERENCES

[1] G. J. Sullivan, J. R. 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, Dec 2012.

[2] T. Wiegand, G. J. 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, July 2003.

[3] Y. Li, L. Li, D. Liu, H. Li, and F. Wu, "Combining directional intra prediction and intra block copy with block partition for hevc," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sept 2016, pp. 524–528.

[4] X. Xu, S. Liu, T. D. Chuang, Y. W. Huang, S. M. Lei, K. Rapaka, C. Pang, V. Seregin, Y. K. Wang, and M. Karczewicz, "Intra block copy in hevc screen content coding extensions," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 6, no. 4, pp. 409–419, Dec 2016.

[5] Carlo Noel Ochotorena and Yukihiko Yamashita, "Regression-based intra-prediction for image and video coding," *CoRR*, vol. abs/1605.03754, 2016.

[6] J. Li, B. Li, J. Xu, R. Xiong, and W. Gao, "Fully connected network-based intra prediction for image coding," *IEEE Transactions on Image Processing*, vol. 27, no. 7, pp. 3236–3247, July 2018.

[7] Y. Li, D. Liu, H. Li, L. Li, F. Wu, H. Zhang, and H. Yang, "Convolutional neural network-based block up-sampling for intra frame coding," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2017.

[8] Y. Li, L. Li, Z. Li, and H. Li, "Hierarchical piece-wise linear projections for efficient intra-prediction coding," in *2017 IEEE Visual Communications and Image Processing (VCIP)*, Dec 2017, pp. 1–4.

[9] HEVC, "Hevc reference software 16.0," 2018.