POSITION DEPENDENT LINEAR INTRA PREDICTION FOR IMAGE CODING

Li Zhang, Siwei Ma, Wen Gao

Institute of Digital Media, Peking University, Beijing, China
Email: {li.zhang, swma, wgao}@pku.edu.cn

ABSTRACT

Intra prediction has been efficiently employed in block based image/video coding to remove the spatial redundancy. In this paper, we propose a position dependent linear intra prediction (PDLIP) approach to further improve the accuracy of prediction. Specifically, for each prediction direction, the prediction of each sample within a target block is calculated as a linear weighted summation of the surrounding spatial neighboring samples in a fixed window. And the prediction coefficients are adaptively derived in the encoding process using the least square estimation method. The derivation process utilizes the original data of the previous image with the same prediction direction after rate-distortion selection to adaptively adjust to the video contents and local context within one block. To alleviate the computation burden of decoding, these prediction coefficients are sent to decoder as side information. Experimental results demonstrate that the proposed PDLIP can achieve a significant coding gain with a slight increase in computational complexity at encoder.

Index Terms— Intra prediction, image coding, linear prediction, least square

1. INTRODUCTION

Intra prediction, which reduces spatial redundancy between the current block and its neighbors, is a key component of block based image/video coding techniques. Better intra prediction method can result in residual with less energy, which will reduce the number of bits needed to reconstruct the signal at decoder. A particularly noteworthy algorithm is the one defined in H.264/AVC [1], the sample predictor block of which is created by extrapolating the reconstructed pixels surrounding the target block along various directions. To better capture the local properties of video signals, H.264/AVC divides the block sizes for intra prediction (IP) into 4×4 (Intra4×4), 8×8 (Intra8×8) and 16x16 (Intra16×16). For Intra4×4 and Intra8×8 modes, nine prediction modes (i.e., eight directional modes plus one DC mode) are employed for luminance samples. Additionally, four prediction modes (vertical, horizontal, DC and plane modes) are utilized for Intra16×16 luma blocks. In each case, the encoder selects a directional spatial prediction mode [2] which governs the creation of a prediction of the complete block of samples.

To further improve the performance of IP modes, several new techniques have been proposed. An IP method using multiple reference lines in a reconstructed region is presented in [3], which makes use of the distant samples as well as the nearby ones. Liu et al. [4] introduced the use of two windows of reconstructed pixels from two previous frames to train linear prediction coefficients at both encoder and decoder. Intra prediction method using half-pel samples instead of integer-pel samples has been used in [5]. These methods are based on the assumption that all sequences exhibit similar content characteristics and all the samples in one block are on straight edges along the same direction of the prediction, in which case fixed extrapolation coefficients are efficient. However, video contents are often various and complex texture may exist in one block. As a consequence, if the prediction coefficients can be adaptive to the video contexts, further coding gains can be achieved.

In this paper, we propose a position dependent linear intra prediction (PDLIP) approach to further improve the intra prediction accuracy. In PDLIP, each pixel within the target block is predicted as a linear weighted summation of spatially neighboring samples. The prediction coefficients are derived by online training with the statistics of the previous frame. Utilizing the original samples of the previous frame with the same prediction direction after rate-distortion decision, we estimate the prediction coefficients of pixels located at each position within a target block and each direction. To avoid the increase of decoding complexity, these coefficients are sent as side information to the decoder, where the same training process can be omitted.

The rest of this paper is organized as follows. Section 2 presents the background of intra prediction. Section 3 gives a detailed description of the proposed PDLIP. Experimental results are reported in Section 4, followed by conclusions and future work in Section 5.

2. SPATIAL INTRA PREDICTION IN H.264/AVC

In H.264/AVC, nine different spatial IP modes for 4×4 luma blocks are available, including one DC mode and eight directional modes as illustrated in Fig. 1(a). The predictor of a target block can be derived by extrapolating the

This work was supported in part by National Science Foundation (608330 13) and National Basic Research Program of China (973 Program, 2009CB320903).

neighboring reconstructed pixels along some direction with fixed coefficients. As an example, using the vertical prediction mode marked as mode 0 in Fig. 1(a), to predict a target 4×4 block to be coded as shown in Fig. 1(b), the predictor of each sample C_i (i = 0, 4, 8, 12) in the column of the block next to the neighboring row sample with value P_1, the prediction is the value P_1. These fixed extrapolations can achieve good results for the sequences with simple textures. However, they can not adapt to the varying content of sequences and contexts of different positions within one block. Furthermore, good extrapolation methods to generate predictors should also take the distortion degree of neighboring samples into consideration. Consequently, it is very desirable to devise a more effective IP method to better capture the varying local contexts of the target block.

3. PROPOSED POSITION DEPENDENT LINEAR INTRA PREDICTION METHOD

3.1. Overview of PDLIP

Considering that natural textures often exhibit various features, a simple approach to get more accurate prediction is making full use of the sequence characteristics and the previously coded frames. We employ classification and least square method to derive the adaptive linear weighting coefficients on the previous original frame. Fig. 2 illustrates the coding process of our proposed PDLIP. After one frame has been coded, the reconstructed sub-blocks are classified into several different sets based on the IP modes with rate distortion decision. At the same time, the co-located sub-blocks in the corresponding original picture are also merged into several sets. With this classification process, the uncorrelated or less correlated samples will be excluded which will be conducive to accurate linear coefficients derivation. Based on the classification, the sub-blocks and their neighboring samples of one set in the original frame form one training database to derive the linear prediction coefficients of some prediction mode. Please note that for each prediction mode, the linear coefficients are position dependent, that is, each sample to be coded at some coordinate within one block has its own linear coefficients while samples located at the same coordinates within one sub-block own the same prediction coefficients. The derived coefficients will be used for the following intra coding. Since the linear prediction coefficients, derived from previously coded results, implicitly embed the local texture characteristics, the IP mode can be adaptively adjusted according to the sequence characteristics, frame quality and local context. The following subsections will explain the linear intra prediction and coefficients derivation processes in detail.

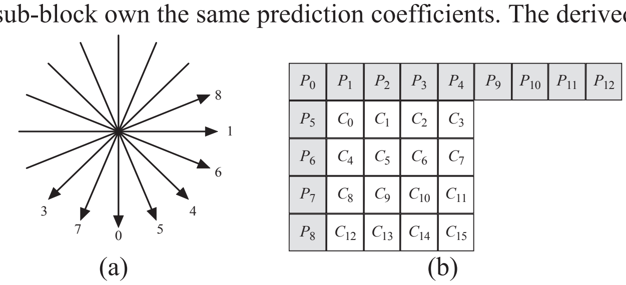

Fig.1. Spatial intra 4×4 prediction in H.264/AVC. (a) Directional intra modes. (b) Boundary and inside samples.

3.2. Linear Intra Prediction Process

In the proposed PDLIP, each pixel within the target sub-block under some IP mode k is predicted as the linear weighted summation of the reconstructed pixels in the left column and the above row relative to the target block. For each IP mode k, the predictor of the ith pixel C_i ($0 \le i \le 15$ for Intra4×4, and $0 \le i \le 63$ for Intra8×8) within the target sub-block can be obtained from its neighboring samples as:

$$\hat{C}_{k,i} = \sum_{j=0}^{N-1} W_{k,i}(j) \times P_j , \qquad (1)$$

where P_j ($0 \le j < N$) represents the jth neighboring intensity value of the target pixel C_i in the original frame. $W_{k,i}(j)$ is the derived linear prediction coefficient of the jth neighboring samples for the target pixel C_i under prediction mode k. N represents the filter tap, that is the neighboring sample number utilized for prediction. For the Intra4×4 mode, N is set to be 9 in our experiments and the corresponding 9 neighboring samples for prediction are those within the left column and the above row of the target block, as P_0~P_8 depicted in Fig. 1(b). For Intra8×8 modes, N is set to be 17. Note that the linear coefficient is mode and position dependent. Each sample to be coded at different positions within one sub-block or under different IP mode may own variable linear coefficients.

Different from the traditional IP modes with fixed extrapolation weights, the PDLIP provides the adaptive prediction weights for each prediction mode, which can be adjusted to the sequence content and coded frame quality. Furthermore, it can be noted that each pixel within the target block has unique prediction weights to capture the local context. Therefore, PDLIP can better capture the spatial varying properties of the target block, which can improve the prediction accuracy.

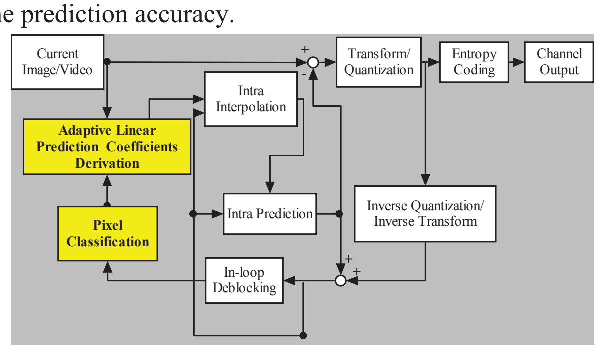

Fig.2. Flowchart of PDLIP.

3.3. Linear Coefficients Derivation

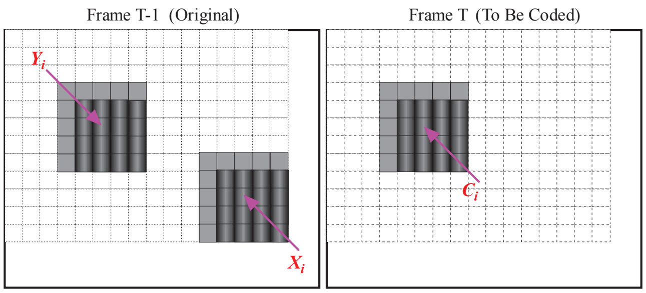

Fig. 3. Illustration of training samples

Linear prediction coefficients $W_{k,\ i}$ (j) can be derived through a few different ways. For example, we can apply either the average or median filters to the neighboring samples. These methods can be easily implemented but have the downside of low accuracy. As we know, the more accurate the linear prediction coefficients are, the higher the coding efficiency will be. To accurately derive $W_{k\ ,i}$ (j), the least square method is employed in our proposed PDLIP. After one frame has been fully coded, the derivation process of the linear prediction coefficients for each IP mode and each relative position within one sub-block will start.

Firstly, training samples collection with **sub-block classification** to find the training samples in the previous coded frame. Sub-blocks with the same IP mode based on the rate-distortion decision will be treated as one category. Samples located in one category at the original frame constitute one training database. Note that we use the original samples instead of the reconstructed ones in order to obtain more accurate prediction coefficients. To calculate the linear prediction coefficients for each coordinate related to one sub-block under each IP mode in the next to be coded frame, we have to find all the valid training samples in the training database. If the sample in a training database can be treated as valid, two more conditions should be satisfied besides the same IP mode:

1) It has the same coordinate relative to the sub-block with the pixel to be predicted to make sure the linear prediction is position dependent.
2) All the samples within the left column and the above row of the sub-block must exist.

Two examples are shown in Fig. 3 to further interpret what valid samples indicate here. The two sub-blocks and their neighbors are depicted in texture region in frame T-1. Take sample X_i for example, if we want to predict pixel C_i in frame T with IP mode k, which locates at the second row and the second column of the current 4×4 block, we first check whether the sub-block where sample X_i is located has chosen mode k. If the answer is positive, we merge this sub-block into the training database of mode k. Then, we will test whether the left column and up row of the current sub-block exist. If the answer is also positive, sample X_i will be defined to be valid, since X_i is also located at the second row and the second column. The rest samples in the current sub-block are set to be invalid for predicting C_i while valid for the same coordinate samples of intra mode k. The same testing process is applied to sample Y_i and other training samples.

Secondly, **linear coefficients calculation** with the least square method for each prediction mode and relative position within one sub-block. For all the valid training samples, we approximate them as the linear weighted summation of the neighboring samples. Take X_i for example, its approximated value under IP mode k can be calculated as

$$\hat{X}_i(k) = \sum_{j=0}^{N-1} W_{k,i}(j) \times P_j \ , \tag{2}$$

where P_j represents the jth neighboring pixel values of X_i, as depicted in Fig. 1(b). The distortion between the actual and the approximated value of the valid training sample X_i under IP mode k can be computed as

$$D\left(X_i(k)\right) = \left(X_i(k) - \hat{X}_i(k)\right)^2 , \tag{3}$$

where $X_i(k)$ denotes the actual value of sample X_i under IP mode k and $\hat{X}_i(k)$ represents the approximated values obtained by Eq. (2). Actually, $X_i(k)$ equals to X_i under each intra mode. Based on the above definitions, the optimal prediction coefficients for intra mode k and position i will correspond to those which minimize the distortion between all the actual valid training sample values and the approximated ones, which can be expressed as

$$\vec{W}^*_{N\times 1}(k,i) = \underset{\vec{W}(k,i)}{\operatorname{argmin}} \sum_{s_m \in S_{valid}} D\left(s_m\right) , \tag{4}$$

where S_{valid} represents the set of all the valid training samples and s_m represents the mth sample within S_{valid}. According to the least square method, the optimal prediction coefficient vector $\vec{W}_{N\times 1}{}^*(k,i)$ can be derived as

$$\vec{W}_{N\times 1}{}^*(k,i) = \left(P^T P\right)^{-1} \left(P^T \vec{S}\right), \tag{5}$$

where $\vec{S}$ denotes a column vector of valid training samples with length of M. P denotes the neighboring sample matrix for each of the pixels in $\vec{S}$ and it is a $M\times N$ matrix, with M representing the number of valid training samples and N representing the number of neighboring pixels to be utilized for each training sample.

With the above two process modules, the derivation of linear prediction coefficients for each intra mode and each relative position has been done. Since the linear coefficients derivation process is performed on the original pixel domain, these coefficients should be sent to the decoder side with fixed length in integer accuracy to avoid the mismatch between the encoder and the decoder. On the other hand, with these coefficients sent to decoder, the derivation process will not need to be done, which can decrease the computational complexity greatly.

4. EXPERIMENTAL RESULTS

To verify the performance of the proposed intra prediction method, the proposed PDLIP is implemented on the latest KTA software with version 2.6 [6]. Experiments are conducted on H.264 "High Profile" and all frames are coded as I frames. In addition, the recently adopted mode-dependent directional transform (MDDT) is used in our test. The coded frame numbers for sequences in 1280×720, 1920×1080 and 2560×1600 formats are 100, 60 and 30, respectively. CABAC is used for entropy coding, the high complexity rate-distortion optimization is set. Only three intra prediction modes (vertical, horizontal and DC) are replaced in our test. Based on the assumption that the whole sequence exhibits stationary Markov characteristics, these linear prediction coefficients are derived only once after the first frame is coded, which can avoid high complexity increase at encoder.

Six test sequences with different resolutions are tested and listed in Table 1, where coding efficiency is compared based on the methodology given in [7]. These results are derived under four different *QP*s, including 22, 27, 32 and 37, which are used as common test conditions in VCEG. From the third column in Table 1, it can be seen that over a wide range of test set, our proposed PDLIP achieves 0.707 dB coding gain on average for these QPs. These results are derived with only Intra4×4 enabled to check the coding gain of Intra4×4 modes. The corresponding rate-distortion performance for sequence *Raven* is also shown in Fig. 4. Here "anchor" represents the result achieved by the default intra prediction method in the KTA 2.6 software. It is easy to observe that the proposed PDLIP outperforms the anchor from low to high bitrates. The fourth column in Table 1 lists the coding gain over KTA with all intra prediction modes (Intra4×4, Intra8×8 and Intra16×16) on. Further coding gain of 0.21dB on average can be achieved compared with current KTA's status.

5. CONCLUSIONS AND FUTURE WORK

In this paper, PDLIP is proposed to improve the intra prediction accuracy. For each intra prediction mode, each sample within a target block is predicted as the linear weighted summation of the reconstructed pixels surrounding the target block. The linear prediction coefficients derivation process utilizes the previous original samples, the rate-distortion decision results and the relative coordinate within one sub-block. To avoid the increase of decoding complexity, the linear prediction coefficients are computed after one frame being coded and sent to decoder as side information. Experimental results show that the proposed PDLIP is able to improve the intra prediction accuracy greatly with coding gain up to 1.06 dB. In the near future, we will extend it for other intra modes. Besides, robust strategy to detect the scene change and re-calculate the linear coefficients will also be studied.

6. REFERENCES

[1] ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4) AVC: "Advanced Video Coding for Generic Audiovisual Services," Mar. 2005.

[2] T. Wiegand, G. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," IEEE Trans. on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560-576, Jul. 2003.

[3] S. Matsuo, S. Takamura, K. Kamikura and Y. Yashima, "Extension of Intra Prediction Using Multiple Reference Lines," ITU-T Q.6/SG16 VCEG, VCEG-AF05, Apr. 2007.

[4] L. Liu, Y. Liu, and E. J. Delp, "Enhanced Intra Prediction Using Context-Adaptive Linear Prediction," Proc. of Picture Coding Symposium, Nov. 2007.

[5] J. Park, S. Park and B. Jeon, "Intra Prediction with Subpel Samples," ITU-T Q.6/SG16 VCEG, VCEG-AG09, Oct. 2007.

[6] "jm11.0kta2.6," "http://iphome.hhi.de/suehring/tml/download/KTA," Oct. 2009.

[7] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-Curves," ITU-T Q.6/SG16 VCEG, VCEG-M33, Apr. 2001.

Table 1. Coding gain of PDLIP over KTA with MDDT

Resolution	Sequence	△PSNR (dB)	
		Only Intra4×4	Intra4×4/8×8 Intra16×16
1280×720 @60Hz	City	0.402	0.191
	Harbour	0.666	0.338
	Raven	0.610	0. 097
1920×1080 @24Hz	Kimo1	0..218	0.187
	Sunflower	1.063	0.153
2560×1600	PeopleOnStreet	0.796	0.180
Average		**0.707**	**0.210**

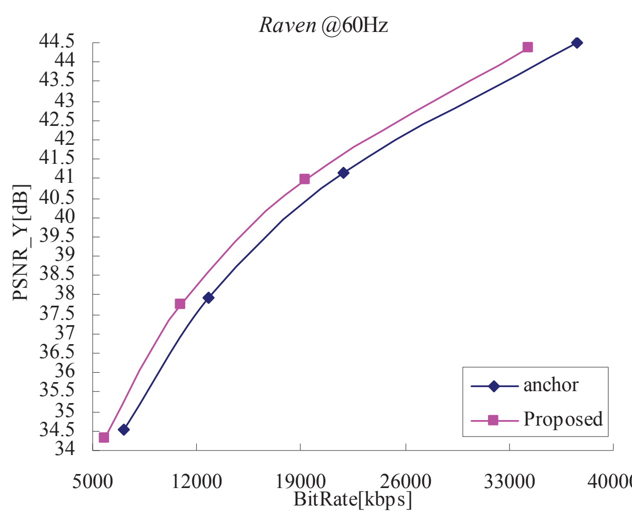

Fig.4. RD curves of *Raven* (HD@60HZ)