Enhanced Line-based Intra Prediction with Fixed Interpolation Filtering

Li Zhang, Siwei Ma, Wen Gao
Institute of Digital Media
Peking University,
Beijing, China
{li.zhang, swma, wgao}@pku.edu.cn

Xin Zhao
Key Lab of Intelligence. Information Processing
Institute of Computing Technology
Beijing, China
xzhao@jdl.ac.cn

Abstract—**In this paper, a novel intra prediction algorithm, named enhanced line-based intra prediction (ELIP), is proposed to improve the traditional intra prediction methods for image/intra-frame coding. Different from the existing schemes where the interpolation filtering only depends on the intra prediction mode, in the proposed ELIP, the linear filtering coefficients are further refined by imposing both mode and position dependencies. For each intra prediction mode and each position within one block, a set of linear filtering coefficients are obtained from off-line training using the least square method. The proposed algorithm has been implemented in the latest ITU-T VCEG KTA software. Experimental results demonstrate that, compared with the intra coding scheme of KTA with new intra coding tool enabled, up to 0.44dB additional coding gain is achieved by the proposed method, while keeping applicable computational complexity for practical video codecs.**

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

I. INTRODUCTION

Reducing the spatial redundancies presented in an image is the basis for image/video compression standards including MPEG-2 [1] and H.264/AVC [2]. Intra prediction (IP) is such a tool to represent image using prediction residual with much lower energy, and is of great importance for compression tasks. The development of intra prediction has passed several decades. In MPEG-2, transform-domain intra prediction is employed to code DC coefficients with differential prediction. After that, researches are mainly focused on spatial-domain intra prediction with multiple prediction modes and directions selection [3][4][5], and a representative one is the scheme defined in H.264/AVC [6].

In H.264/AVC, the line-based intra prediction is employed, where the sample predictor block is created by extrapolating the reconstructed pixels surrounding the target block along the dominating direction. Moreover, to better capture the local properties of video signals, for intra prediction, H.264/AVC divides the block sizes into 4×4 (Intra4×4), 8×8 (Intra8×8) and 16x16 (Intra16×16). For both 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. For each case, the encoder measures the efficiency of each candidate based on a certain criteria, and selects the optimal one for the actual coding. Furthermore, efforts have been made to further improve the efficiency of intra prediction. Liu *et al.* [7] utilized two windows of reconstructed pixels from two previous frames to train the linear prediction coefficients at both encoder and decoder. In [8], Wang proposed a distance-based weighted prediction method to improve the prediction efficiency of DC mode. Moreover, a weighted cross prediction mode is also proposed in [9]. With the modified interpolation method, the decorrelation ability is enhanced and higher coding gain can be achieved. In [10], Marta proposed a new IP scheme using different sets of prediction weight for each pixel within one block, which is based on a two-pass scheme and introduces high coding complexity and large delay.

In this paper, an enhanced line-based intra prediction (ELIP) approach is proposed to further improve the intra prediction accuracy. In ELIP, each pixel within the target block is predicted using a linear weighted sum of spatially neighboring samples. To derive the linear filtering coefficients, the least square method is used for off-line training and the training database is constructed with classified original samples. Furthermore, the filtering coefficients are kept as constant throughout the encoding or decoding process under the consideration of computational complexity. It is noted that different from the intra prediction method in H.264/AVC, the linear filtering coefficients are both mode- and position-dependent. That is, for different IP mode, samples located at different positions within one block employ different linear filtering coefficients. Therefore, higher prediction accuracy can be achieved.

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 ELIP. Experimental

results are reported in Section 4, followed by conclusions and future work in Section 5.
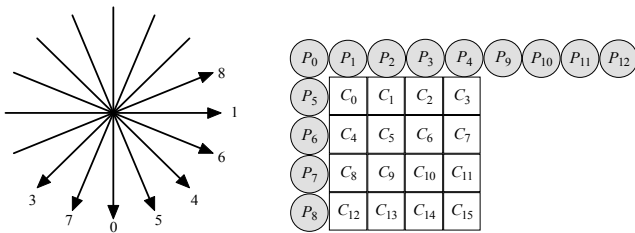
II. SPATIAL INTRA PREDICTION IN H.264/AVC

In H.264/AVC, nine different spatial intra prediction modes for 4×4 luma blocks are available, including one DC mode and eight directional modes as illustrated in Fig. 1(a). The prediction of a target block is derived by extrapolating the neighboring reconstructed pixels along one specific direction with fixed coefficients. Fig.1 illustrates 16 samples (labeled as C_i ($i = 0~15$)) of one 4×4 block which are predicted by the reconstructed samples of upper and left-hand reference pixels of current block (labeled as P_j ($j = 0~13$)). For DC mode, the prediction values of all the samples are the same. When the upper and left-hand neighbor blocks exist, the prediction value of C_i can be calculated by

$$\hat{C}_i = (\sum_{j=1}^{8} P_j + 4) \gg 3 . \qquad (1)$$

For the vertical prediction mode labeled as mode 0 in Fig. 1(a), the predictor of samples 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.

One advantage associated with such line-based intra prediction is that it is easy for practical implementation. However, it lacks of sufficient consideration for accurate prediction. As pointed out in [11], the correlation between two samples in natural images is inversely relative to their distance. Therefore, samples at different positions within one block should employ different prediction methods to generate the predictor. Based on the above observation, it is motivated to design the prediction filtering coefficients with both mode and position adaptation.

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

III. ENHANCED LINE-BASED INTRA PREDICTION

In this section, a brief overview of our proposed ELIP method is presented first, and the technical details are discussed in the following subsections, including the training process and implementation.

A. Overview of ELIP

To restrain the increase of computational complexity as much as possible, fixed interpolation filtering in the conventional methods is used in the proposed ELIP. Each sample within a target sub-block under some intra prediction mode is predicted as a 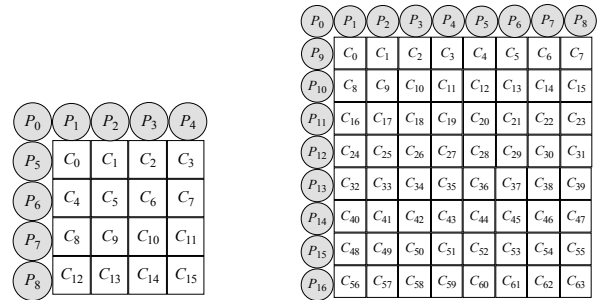linear weighted summation of the reconstructed samples in the left column and the above row relative to the target block. However, there are two key differences between ELIP and the conventional method:

1) ELIP employs different filtering coefficients for samples located at different position within one coding block.
2) Different reconstructed samples are used as reference samples.

In the proposed ELIP, the predictor of the ith pixel C_i ($0 \leq i \leq 15$) within the target sub-block can be obtained from its neighboring samples as:

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

where k denotes the IP mode index. P_j ($0 \leq j < N$) represents the jth neighboring reconstructed intensity value of the target pixel C_i. $W_{k,i}(j)$ 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 number of neighboring samples utilized for prediction. For vertical, horizontal and DC modes of Intra4×4, 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, labeled as $P_0 \sim P_8$ depicted in Fig. 2(a). The 17 neighboring samples for vertical, horizontal and DC modes of Intra8×8 are depicted in Fig. 2(b), labeled as $P_0 \sim P_{16}$. 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 own variable linear filtering coefficients. Therefore, the proposed ELIP is able to better reduce the spatial redundancies among the target samples and the reference samples, which can bring in higher prediction accuracy and further coding efficiency.

(a) Intra4×4   (b) Intra8×8
Figure 2. Reference sample definition in ELIP.

B. Off-line Training

Linear filtering coefficients $W_{k,i}(j)$ can be derived through different ways. For example, we can apply either the average or median filters to the neighboring samples. These methods are easy for practical implementations but have the downside of low accuracy. As it is well known that, the more accurate the linear filtering coefficients are, the higher the coding efficiency will be. To accurately derive $W_{k,i}(j)$, the least square method is employed in our off-line training process. The training process is described as follows:

Firstly, a training sequence is coded with the existing intra prediction method. To better capture the texture information, we utilize 8 widely used sequences in 1920×1080 format to construct the training sequence. The first four frames of each test sequence are extracted and construct one training sequence, which is intra coded with rate-distortion optimization on. To derive the linear filtering coefficients of both Intra4×4 and Intra8×8, adaptive block-size transform is also used in our simulation.

Secondly, sample databases for each position and each prediction mode with sample classification are set up. 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, including target sample C_i and its neighboring reference sample P_j, instead of the reconstructed ones, to train a set of more accurate filtering coefficients. To calculate the linear filtering coefficients for each position 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 including:

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.

Thirdly, derive the linear filtering coefficients with the least square method for each prediction mode and relative position within one block. For all the valid training samples, we approximate them as the linear weighted summation of the neighboring samples. Take any sample C_i for example, its approximated value using IP mode k can be calculated as

$$\hat{C}_i = (\sum_{j=1}^{8} P_j + 4) \gg 3 .\tag{3}$$

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 C_i under IP mode k can be computed as

$$D\big(C_i(k)\big) = \big(C_i(k) - \hat{C}_i(k)\big)^2 .\tag{4}$$

where $C_i(k)$ denotes the actual value of sample X_i under IP mode k and $\hat{C}_i(k)$ represents the approximated values obtained by Eq. (3). Actually, $C_i(k)$ equals to C_i under every intra mode. Based on the above definitions, the optimal filtering coefficients for intra mode k and position i correspond to those which minimize the distortion between all the actual valid training sample values and the approximated ones, which can be expressed as

$$\vec{W}^*_{N\times1}(k,i) = \underset{\vec{W}(k,i)}{\mathrm{argmin}} \sum_{s_m \in S_{valid}} D\big(s_m\big) .\tag{5}$$

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 $\vec{W}_{N\times1}^*(k,i)$ can be derived as

$$\vec{W}^*_{N\times1}(k,i) = \big(P^T P\big)^{-1} \big(P^T \vec{S}\big) .\tag{6}$$

where $\vec{S}$ denotes a column vector of valid training samples with length of M. P denotes the neighboring sample matrix for each pixel 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 three process modules, the derivation of linear filtering coefficients for each intra mode and each relative position has been done. The derived filtering coefficients will be used in the intra prediction process, and they will not change throughout the encoding or decoding process.

C. Implementation

Since the linear filtering coefficients derivation process is performed with the least square method, they are not always integers. To avoid floating-point operations, $W_{k, i}(j)$ is quantized to a 16-bit integer. Equation (2) can then be rewritten as

$$\hat{C}_i(k) = (\sum_{j=0}^{N-1} W_{k,i}(j) \times P_j + (1 << 15)) \gg 16 .\tag{7}$$

In our simulations, it is observed that for vertical, horizontal and DC modes of Intra4×4, 9 neighboring samples are good enough, and 17 neighboring samples are needed to calculate the predictor for Intra8×8. Furthermore, to better cooperate with the proposed ELIP, we update the transform matrices for these three IP modes of Intra4×4 and Intra8×8 in mode-dependent directional transform (MDDT) [12]. The calculation process is the same as that in MDDT, which utilizes the singular value decomposition (SVD) method for off-line training.

IV. EXPERIMENTAL RESULTS

To verify the performance of the proposed intra prediction method, the proposed ELIP is implemented on the latest KTA software with version 2.6 [13]. Experiments are conducted on H.264/AVC "High Profile" and all frames are coded as I frames. In addition, the MDDT technique is enabled in both the anchor and the proposed method in our test. The test set covers a wide range of resolutions including QCIF (176×144), CIF (352×288), 720p (1280×720), 1080p (1920×1080) and 1600p (2560×1600) formats. The coded frame number is 100, 60 and 30 for 720p (and below), 1080p and 1600p, respectively. CABAC is turned on for entropy coding. The

high complexity rate-distortion optimization is set. Although the proposed method works only on three IP modes (vertical, horizontal and DC) of 4×4 and 8×8 sub-block units, Intra4×4, Intra8×8 and Intra16×16 modes are all enabled and adaptive block-size transform is used in our test.

Eighteen test sequences covering different resolutions are tested and listed in Table 1, where coding efficiency is compared based on the methodology given in [13]. These results are derived under four different *QP*s, including 20, 24, 28 and 32. From Table 1, it can be seen that over a wide range of test set, our proposed ELIP achieves up to 0.44 dB coding gain on average for these QPs. The rate-distortion performance for sequence *Harbour* in 720p format is also shown in Fig. 3. Here "anchor" represents the result achieved by the default intra prediction method in the KTA 2.6 software with MDDT on. It is observed that the proposed ELIP outperforms the anchor in the full range of bitrates.

TABLE I.   CODING GAIN OF ELIP OVER KTA WITH MDDT

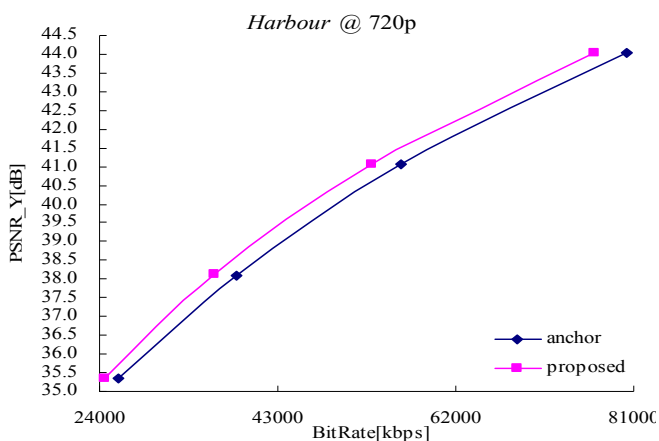| Resolution	Sequence	△PSNR (dB)	Resolution	Sequence	△PSNR (dB)
720p (60Hz)	City	0.30	QCIF (15Hz)	Bus	0.17
	Harbour	0.44		City	0.25
	Raven	0.18		Harbour	0.40
	Night	0.17	CIF (30Hz)	Bus	0.24
1080p (50Hz)	Crowdrun	0.18		Bike	0.27
	Parkjoy	0.17		City	0.23
	Wisley2	0.18		Harbour	0.40
1600p (30Hz)	Peopleon Street	0.21		Flower	0.27
	Traffic	0.16		Tempete	0.24
Average		**0.22**			**0.27**

Figure 3.   RD curves of Harbour (720p @ 60HZ)

V.   CONCLUSIONS AND FUTURE WORK

In this paper, an enhanced line-based intra prediction algorithm is proposed to further explore the spatial redundancies. For each IP mode, each sample within a target block is predicted as the linear weighted sum of the reconstructed pixels surrounding the target block. The linear filtering coefficients are derived off-line with the least square

method and classified original samples are used to construct the database. Experimental results demonstrate that the proposed ELIP presents better prediction accuracy and achieves superior coding gain with applicable computational complexity.

In the current ELIP, only three IP modes (vertical, horizontal, DC) of Intra 4×4 and Intra8×8 are replaced. In the near future, extensions to the remaining intra prediction directions will also be investigated as well as Intra16×16. Moreover, reduction of the filter tap for lower algorithm complexity will also be further investigated.

REFERENCES

[1] ITU-T. Information Technology - Generic Coding of Moving Pictures and Associated Audio Information: Video. ITU-T Recommendation H.262 | ISO/IEC 13818-2 (MPEG-2), 1995.

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

[3] G. Bjøntegaard, "Coding Improvement by Using 4×4 Blocks for Motion Vectors and Transform," in SG16/Q15. VCEG, 3rd Meeting, Document Q15-C23, Dec. 1997, ITU-T.

[4] D. Wang, "Improvement of H.26L Intra Block Prediction," ITU-T Q.6/SG16 VCEG, VCEG-L09, Jan. 2001.

[5] G. Conklin, "New Intra Prediction Modes," ITU-T Q.6/SG16 VCEG, VCEG-N54, Sep. 2001.

[6] 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.

[7] L. Liu, Y. Liu, and E. Delp, "Enhanced Intra Prediction Using Context-Adaptive Linear Prediction," Proc. of Picture Coding Symposium (PCS) 2007.

[8] S. Yu, Y. Gao, J. Chen and J. Zhou, "Distance-based Weighted Prediction for H.264 Intra Coding," IEEE IET International Conference on Audio, Language and Image Processing, pp. 1477-1480, July 2008.

[9] L. Wang, L. Po, Y. Uddin, K. Wong and S. Li, "A Novel Weighted Cross Prediction for H.264 Intra Coding," IEEE International Conference on Multimedia & Expo, pp. 165-168, May 2009.

[10] M. Karczewicz, "Improved Intra Prediction," ITU-T Q.6/SG16 VCEG, VCEG-AF15, April 2007

[11] J. O'Neal, JR. and T. Natarajan, "Coding Isotropic Images," IEEE Trans. Inform. Theory, vol. IT-23, pp.697-707, Nov. 1977.

[12] Y. Ye and M. Karczewicz, "Improved H.264 Intra Coding Based on Bi-directional Intra Prediction, Directional transform, and Adaptive Coefficient Scanning," Proceedings of the 2008 IEEE International Conference on Image Processing (ICIP), pp. 2116-2119, Oct. 2008.

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

[14] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-Curves," ITU-T Q.6/SG16 VCEG, VCEG-M33, Apr. 20.