ENHANCED INTRA PREDICTION AND TRANSFORM FOR VIDEO CODING

Kai Zhang[1,2,3*], *Shawmin Lei*[3] *and Wen Gao*[1,2]

[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100080, China
[2]Graduate University of the Chinese Academy of Sciences, Beijing 100039, China
[3]MediaTek (Beijing) Inc., Beijing, 100080, China
{kzhang, wgao}@jdl.ac.cn, shawmin.lei@mediatek.com

ABSTRACT

Intra-coding plays an important role in video coding schemes and it has become a hot research topic of the next generation video coding standard. In this paper, we propose two techniques to improve the efficiency of the prediction and the transform, respectively, for intra-coding. First, we introduce an overlapped block intra-prediction (OBIP) method, which makes use of the intra directional prediction modes of neighboring blocks as well as the mode of the current block. A position dependent weighted sum of several possible predictors will be treated as the final prediction. Second, we propose a multiple-model Karhunen–Loève transform (MMKLT) technique to further improve the mode dependent directional transform (MDDT) method. More than one KLT will be trained offline for each intra-prediction mode based on residual variances. Experimental results indicate that the proposed two methods together can achieve an average bit-rate reduction about 11% compared with H.264 or 5% compared with MDDT in all-intra coding.

Index Terms—H.264/AVC, intra-coding, intra-prediction, KLT, MDDT

1. INTRODUCTION

Intra-coding plays an important role in hybrid video coding schemes, especially in applications such as error resilience, random access, low-complexity encoding and professional video editing. H.264/AVC [1], which is a novel international video coding standard, introduces directional intra prediction in the spatial domain. Several predefined intra-prediction modes are served as candidates, and then the rate distortion optimization (RDO) [2] procedure will choose the most appropriate one for each block. In this way, H.264/AVC outperforms previous video/image coding standards dramatically on intra-coding efficiency.

Many researchers have made lots of efforts to further improve the intra-coding performance of H.264/AVC. Yang *et al.* [3] borrowed the motion compensation concept and utilized template matching both on the encoder side and the decoder side. Zhang *et al.* [4] and Shiodera *et al.* [5] introduced additional prediction modes by means of exploring new texture features [4] or linear combination of the existing modes [5] respectively. It is also suggested to alternate the coding order of blocks adaptively in a macro-block (MB) [5]. Recently, Ye *et al.* [6] proposed the mode dependent directional transform (MDDT) method, which showed a good performance thus has been adopted into the VCEG/KTA software [7]. In general, MDDT is based on the observation that residues generated by different intra-prediction modes tend to possess different statistic characteristic. Therefore, if one block is predicted by a certain intra-prediction mode, one specific transform corresponding to this mode should be utilized. The optimal transform for each mode is proved to be the KLT theoretically [8], which can be obtained by offline training with several sequences. Moreover, MDDT allows the scanning order of coefficients after the transform to adjust adaptively online according to statistic.

To improve the efficiency of intra-coding further, we propose two novel methods. First, we introduce the overlapped block intra-prediction (OBIP) technique, which was inspired by the well-known overlapped block motion compensation (OBMC) [9] idea. In OBIP, the prediction for one position is a linear combination of three possible predictors, which are generated by using the prediction modes of the current block, the left block and the up block. The weights of predictors can be estimated offline by applying a linear regression process [10]. Second, we enhance MDDT with the multiple-model Karhunen–Loève transform (MMKLT) approach. It is found that the optimal transform of a block depends on not only the prediction mode but also the residual variance. Therefore, we predefine three classes of KLTs for each intra-prediction mode, which are trained offline according to different variances. The encoder can select the most suitable transform class based on RDO criterion at MB-level, and transmit this information to the decoder.

The rest of this paper is organized as follows. Section 2 and section 3 describe the proposed OBIP and MMKLT methods respectively. Experimental results are presented in section 4 and conclusions are drawn in section 5.

* This work is done when the first author is with MediaTek (Beijing) Inc. as an intern.

2. OVERLAPPED BLOCK INTRA-PREDICTION

2.1. Intra-prediction in H.264/AVC

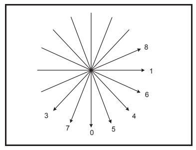

Fig. 1 Intra_4x4 prediction modes.

H.264/AVC introduces directional intra-prediction in the spatial domain [1]. Intra_4x4 and Intra_8x8 both have up to nine prediction modes, namely mode 0 - mode 8. Except the DC mode (mode 2), the other eight modes correspond to different prediction directions as illustrated in Fig. 1. In natural sequences, the optimal intra-prediction mode usually represents the texture direction in a block. Therefore, the directional prediction can reduce the texture redundancy significantly.

2.2. Beyond-block effect

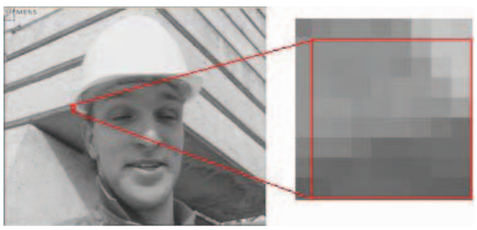

Fig. 2 Beyond-block effect.

Although directional intra-prediction works well in most cases, it has an inherent shortcoming. Intuitively, the prediction is conducted block by block due to the requirement of block-transform, but texture in natural images is not restricted by block edges. Therefore, only one prediction mode for a block may be not accurate enough to describe the texture characteristic in the region. We call this phenomenon 'beyond-block effect'. Fig. 2 demonstrates a typical example of beyond-block effect in an 8x8 luminance block of the CIF sequence 'foreman'. It is obvious that the texture in this individual block possesses both the right-up direction and the left-down direction. No matter mode 4 or mode 8 is chosen, the geometric properties cannot be expressed sufficiently for this block, thus the intra-prediction will be inefficient.

An important aspect of beyond-block effect is that the texture characteristic of a block is influenced by that of its neighboring blocks, which can be represented by their intra-prediction modes. In the example above, we can observe that the texture of the left block and up block just possess the right-up and left-down directions respectively. This effect leads us to develop the OBIP algorithm.

2.3. OBIP algorithm

Based on the beyond-block effect, we propose the OBIP method. The idea of OBIP is quite similar to the well-know OBMC technique [9] if we draw an analogy between the role of motion vectors in inter-prediction and that of directional modes in intra-prediction. Detailed description about OBIP is as follows.

When a block is predicted by OBIP, three intra-prediction modes will be considered: the current block mode ($Mode_C$), the left block mode ($Mode_A$) and the up block mode ($Mode_B$). For a pixel $S(i, j)$ where (i, j) is the coordinate in the block, three possible predictors $P_C(i, j)$, $P_A(i, j)$, and $P_B(i, j)$ can be generated with $Mode_C$, $Mode_A$ and $Mode_B$ respectively. Then the prediction $P(i, j)$ for $S(i, j)$ can be calculated as

$$P(i,j) = \sum_{X \in \{A,B,C\}} \omega_X^{ModeC}(i,j) \bullet P_X(i,j)\,, \qquad (1)$$

where $\omega_X^{ModeC}(i,j)$ is the weighting value on $P_X(i, j)$ when the current block mode is $Mode_C$. We can see from (1) that the possible predictors are weighted summed up, and the weighting values depend on pixel positions as well as the current block mode.

There are several ways to understand expression (1). In theory, $\omega_X^{ModeC}(i,j)$ can be regarded as the probability that $P_X(i, j)$ is the accurate prediction for $S(i, j)$, then $P(i, j)$ will be the optimal estimation for $S(i, j)$ [9]. In practice, we consider $P(i, j)$ as a linear estimation for $S(i, j)$, and the optimal $\omega_X^{ModeC}(i,j)$ in L^2 space can be obtained by a linear regression process [10].

For robustness, the linear regression is done offline at various bit-rates on plenty of training sequences, including: 'foreman', 'paris' in CIF format, 'bigships', 'city', 'crew', 'night', 'jets', 'raven' in 720p format, and 'sunflower', 'traffic', 'toys_calendar' in 1080p format. Fig. 3 depicts the normalized weighting values on each position for Intra_4x4 mode 4. It is a reasonable result that prediction modes of neighboring blocks play a more important part on the pixels at edge, and the current block mode will dominate on the inner pixels.

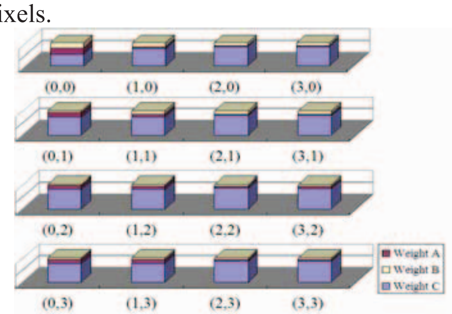

Fig. 3 Weighting values for Intra_4x4 mode 4.

In practice, we make OBIP an optional function for Intra_4x4 and Intra_8x8. Thus one MB can choose whether to use OBIP or not according to RDO criterion [2], and a switch flag must be coded and transmitted. Table 1 exposes the percentage of MBs that prefer OBIP and the average residual energy reduction (ΔE) in all-intra coding on some sequences outside the training set when QP = 32. The results indicate that OBIP is an efficient prediction method.

Table 1 Verification of OBIP

sequences	I4MB OBIP on	I8MB OBIP on	ΔE
SVGA\janine	55.27%	49.60%	-18.12%
720p\cyclists	62.14%	47.70%	-13.59%
1080p\riverbed	74.02%	60.90%	-16.80%

3. MULTIPLE-MODEL K-L TRANSFORM

3.1. Karhunen–Loève transform

KLT has very extensive applications in stochastic signal analysis and processing. Suppose C_x is the covariance matrix of a broad stationary stochastic signal vector $\vec{x}$, and then the eigen vectors of C_x will constitute the bases of the KLT for $\vec{x}$, denoted as K_x. Thus, the covariance matrix C_y of the transformed vector $\vec{y}$ can be formulated as

$$\begin{aligned} C_y &= E\left\{[\vec{y}-\vec{\mu}_y][\vec{y}-\vec{\mu}_y]^T\right\} \\ &= E\left\{[K_x\vec{x}-K_x\vec{\mu}_x][K_x\vec{x}-K_x\vec{\mu}_x]^T\right\} \\ &= K_x C_x K_x^T = \begin{bmatrix} \lambda_0 & & & \\ & \lambda_1 & & \\ & & \ddots & \\ & & & \lambda_{N-1} \end{bmatrix} \end{aligned} , \quad (2)$$

where μ_x and μ_y are means, and $\{\lambda_n\}$ is the eigen values. Thus KLT can transform $\vec{x}$ into another expression $\vec{y}$ which eliminates all correlations between its components. This property makes KLT the optimal transform for data compression. In practice, it is widely accepted that discrete cosine transform (DCT) is a good approximation to KLT with much less complexity in image and video coding [8].

It is argued that DCT may not work so well with the directional intra-prediction because the residues generated by different prediction modes usually possess different statistic characteristic. Therefore, MDDT [6] treats residues generated by different modes as different signals, and attain a distinguished KLT offline by (2) for each one.

3.2. Analysis of residual signals

Although MDDT can improve the intra-coding performance significantly, it has not explored the diversity of residual signals sufficiently. Actually, residues often appear quite different in statistic even if they are generated by the same intra-prediction mode. This is reasonable because many factors besides prediction modes affect residues, such as the regularity of texture and the quality of prediction. Fig. 4 shows residues in two Intra_4x4 block of the CIF sequence ‘salesman’ at a high bit-rate (QP = 22). Although they both choose mode 4 in the RDO process, their residual signals appear very different: one is quite flat while the other jumps violently. It is obvious that only one transform can hardly handle both of them efficiently.

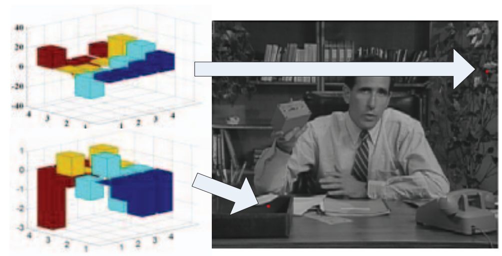

Fig. 4 Examples of residual signals.

3.3. MMKLT algorithm

Towards the problem above, we propose MMKLT method. Generally, multiple models are presumed for the residual signal generated by one intra-prediction mode, and one K-L transform is used for each model. The KLTs are obtained by offline training, and then they will be considered as candidate transform classes in the RDO process at MB level.

In the offline training procedure, residual signals with each prediction mode are classified according to their variances into three classes: low, medium and high. There are several reasons to select the variance as the criterion. First, the variance of residues is sensitive to many factors and can describe the characteristic of a block to some extent. Second, the variance is an unbiased estimation to the trace of C_x, thus it possesses an intrinsic relationship with KLT. Finally, the variance is very simple and easy to calculate. The same training set as OBIP is used. After training, three KLTs denoted as $T^L(m)$, $T^M(m)$ and $T^H(m)$, will be determined for mode m corresponding to the three residual classes. In other words, there will be three mode-dependent transform classes: $\{T^L\}$, $\{T^M\}$, and $\{T^H\}$.

In the encoding procedure, we apply MMKLT on Intra_4x4, Intra_8x8 and Intra_16x16 modes. The transform class is selected at MB level. For instance, if $\{T^L\}$ is used in an MB, a block with prediction mode m in the MB will utilize transform $T^L(m)$ fetched from $\{T^L\}$. The RDO process checks the three transform classes one by one, and chooses the one producing the least RD cost to be the optimal transform class for the MB. In addition, the adaptive scanning technique is inherited from MDDT.

In order to inform the decoder which transform class is used, overhead is needed at MB level. We adopt a predictive coding method to compress the overhead. First, the classes are indexed by symbols 0, 1, 2, in the order of their selection probabilities, from high to low. Then, the smaller class index of the left and up neighboring MBs is defined to be the predictor. If the transform class of the current MB is predicted correctly by this predictor, 1 is transmitted. Otherwise, 0 is transmitted and an additional bit which indicates the right type is coded. For efficiency, we let the selection probabilities adjust according to statistic online, so that the mapping between symbols and classes is adaptive.

4. EXPERIMENTAL RESULTS

In order to verify the efficiency of proposed algorithms, we have made extensive simulations on the reference software KTA 2.2r1 [7]. In our test, H.264 is chosen as the anchor. And the test conditions are shown in Table 2.

Table 2 Test conditions of proposed algorithms.

Profile	High, All-intra
CABAC	on
RDO	on
Transform 8x8	on
Coded frame number	100
QPI	22, 27, 32, 37

The experimental results are shown in Table 3, where ΔB is the bit-rate reduction [11] compared with the anchor. Fig. 5 depicts several RD curves. It can be seen that MDDT outperforms H.264 with an average bit-rate reduction more than 6%. MMKLT improves this gain about 3% to more than 9%, and OBIP contributes another 2%. As a whole, the proposed solution can achieve a bit-rate reduction about 11% compared with H.264. It should be noticed that the performance prevails on all test sequences with various resolutions, no matter whether the sequence is in the training set or not.

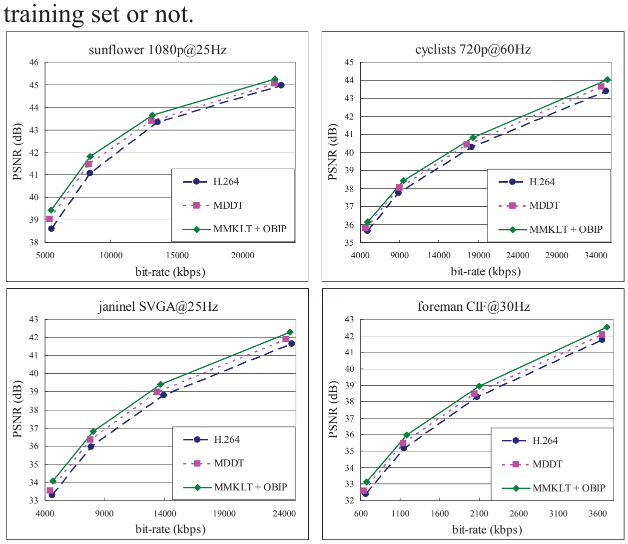

Fig. 5 RD curves in all-intra coding.

5. CONCULSION

In this paper, we present two enhanced intra-coding methods. We first describe the motivation and realization of OBIP. Then the details of MMKLT are provided including analysis of residual signals and construction of the transform classes. Experimental results show that the proposed algorithms together can improve intra-coding performance significantly.

Table 3 Experimental results in all-intra coding

ΔB vs H.264		MDDT	MMKLT	MMKLT+OBIP
CIF	carphone	-6.07%	-9.62%	-11.05%
	foreman	-5.64%	-10.33%	-11.12%
	rush	-6.81%	-9.10%	-10.92%
	salesman	-4.29%	-7.24%	-8.82%
	tempete	-5.29%	-6.30%	-7.91%
CIF ave		-5.62%	-8.52 %	-9.96%
SVGA\ janine		-7.65%	-10.83%	-13.44%
720p	city	-5.90%	-8.60%	-10.19%
	cyclists	-7.21%	-9.91%	-12.00%
	harbour	-5.89%	-7.91%	-9.89%
	raven	-7.93%	-9.30%	-12.03%
	sailormen	-5.44%	-7.11%	-8.88%
	sheriff	-6.60%	-9.00%	-10.57%
720p ave		-6.50%	-8.64%	-10.59%
1080p	riverbed	-6.82%	-9.50%	-12.29%
	rolling_tomatoes	-6.78%	-10.75%	-11.99%
	sunflower	-7.09%	-10.97%	-12.76%
	toys_calendar	-6.31%	-9.14%	-11.56%
	traffic	-5.73%	-8.02%	-10.79%
	tractor	-7.10%	-10.20%	-11.95%
1080p ave		-6.64%	-9.76%	-11.89%
AVERAGE		-6.36%	-9.10%	-11.01%

6. REFERENCES

[1] ITU-T, "Recommendation H.264," http://www.itu.int/rec/T-REC-H.264-200503-I/en, Mar. 2005.

[2] G. Sullivan, T. Wiegand, "Rate-Distortion Optimization for Video Compression," IEEE Signal Processing Magazine, vol. 15, pp. 74-90, Nov. 1998.

[3] J. Yang, B. Yin, Y. Sun and N. Zhang, "A Block-matching Based Intra Frame Prediction for H.264/AVC," International Conference on Multimedia and Expo, pp 705-708, Jul. 2006.

[4] P. Zhang, D. Zhao, S. Ma, Y. Lu, W. Gao, "Multiple Modes Intra-prediction in Intra Coding," International Conference on Multimedia and Expo, pp 419-422, Jun. 2004.

[5] T. Shiodera, A. Tanizawa and T. Chujoh, "Simulation Results of Bidirectional Intra Prediction on KTA Software Version 1.3," VCEG-AF06, Apr. 2007.

[6] Y. Ye and M. Karczewicz, "Complexity Analysis of Improved Intra Coding," VCEG-AH20, Jan. 2008

[7] VCEG, reference software KTA 2.2r1, http://iphome.hhi.de/suehring/tml/download/KTA/jm11.0kta2.2r1.zip.

[8] N. S. Jayant, P. Noll, Digital Coding of Waveforms: Principles and Applications to Speech and Video, Prentice Hall, March, 1984.

[9] M. Orchard, G. Sullivan, "Overlapped Block Motion Compensation: an Estimation-Theoretic Approach," IEEE Trans. on Image Processing, Vol. 3, pp. 693-699, Sep. 1994.

[10] S. Weisberg, Applied linear regression, John Wiley & Sons, Apr. 1985.

[11] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-curves," ITU-T/SG 16, VCEG-M33, Apr. 2001.