An Efficient Combined Inter and Intra Prediction Scheme for Video Coding

Run Cha, Oscar C. Au, Xiaopeng Fan, Feng Zou
Department of Electronic and Computer Engineering,
Hong Kong University of Science and Technology

***Abstract*—Inter prediction and intra prediction are utilized by video coding standard H.264/AVC to exploit the temporal and spatial redundancy respectively. To further improve coding efficiency, combined inter and intra prediction was proposed to produce more accurate prediction signal. However, these methods suffer from either high computation or limited improvement. In this paper we present an efficient inter and intra prediction scheme, which decides intra prediction mode in the combination mode pair adaptively and combines the inter- and intra-prediction signal using fixed spatial-variant weighted coefficient. Experimental results show that the proposed method achieves additional coding gain of up to 0.68% compared to H.264/AVC.**

I. Introduction

The current video coding standard H.264/AVC represents a number of advances in standard video coding technology, in terms of both coding efficiency improvement and flexibility for effective use over various network types and application domains. Based on conventional block-based motion-compensated hybrid video coding concepts, H.264/AVC[1] adopts many new technologies, such as directional spatial prediction for intra coding, in-loop deblocking filtering, quarter-sample-accurate motion compensation and so on.

Fig.1 shows the basic coding structure for H.264/AVC for a macroblock. Input video signal is first splitted into macroblocks and then each macroblock is predicted either spatially or temporally. By subtracting the prediction signal from original one, residue signal is obtained and then transformed using integer DCT. Finally, the transform coefficients are quantized and encoded using entropy coding methods.

When doing inter prediction for a macroblock, H.264/AVC supports a variety of partition sizes from 16 × 16 to 4 × 4. By performing full search in references frames, the prediction signal for each $M \times N$ luma block is derived and then specified by a translational motion vector and reference picture index. In order to reduce the motion-compensated residue, resolution of the motion vector is increased to 1/4 pel. In addition to the inter macroblock coding, intra macroblock coding is also supported by H.264/AVC. When using one of the nine $Intra\_4 \times 4$ mode, each 4×4 block is predicted from spatially neighboring samples. This is well suited for coding of image signals with significant detail. Another intra prediction mode $Intra\_16 \times 16$ performs prediction on the whole 16×16 luma block and is suited for coding smooth regions.

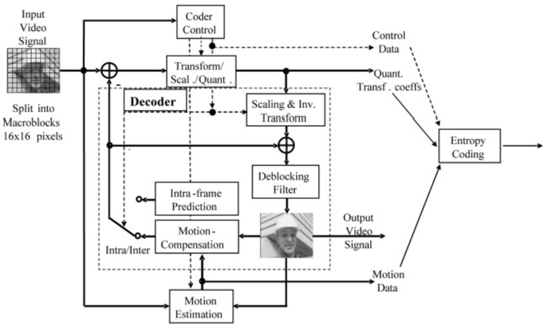


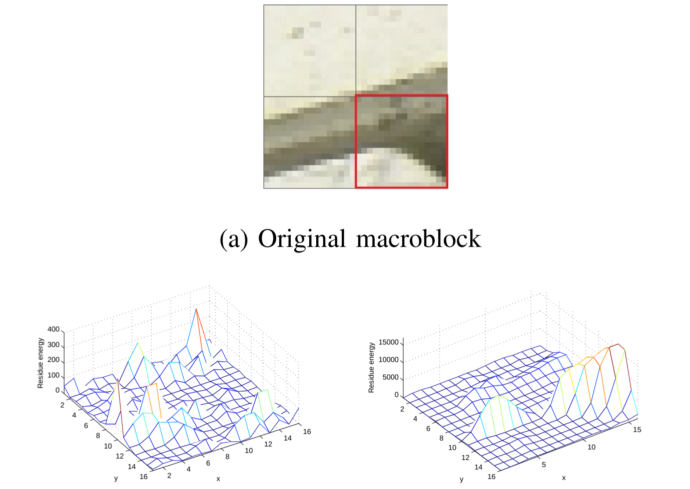
Fig. 1. Block diagram of H.264/AVC encoder.

Fig.2(a) shows a macroblock(with red border) selected from the second frame in foreman.yuv. Fig.2(b) and (c) shows corresponding residue surface using $Inter\_16 \times 16$ and $Intra\_16 \times 16\_Hor$ mode respectively. It can be seen that inter prediction produces smaller residue than intra prediction on the whole except edge regions while intra prediction works better for the pixels located near to neighboring samples. From this example, we know that inter prediction and intra prediction can be complementary in some cases. Combined inter and intra prediction was first proposed in [2] which added all possible mode combinations as candidate modes and select the best mode pair and weighted coefficient based on rate-distortion (R-D) criterion[3]. Although this method has improved coding efficiency, the encoder has to explicitly signal its selection in the bitstream, which limits the potential gain. Furthermore, this method requires much higher computational complexity. Researchers later improved this work and developed a more practical scheme $CII\_16 \times 16$ for encoders [4], which only adopted one additional combination mode and the weighted coefficient was trained to be fixed.

In order to further improve the coding efficiency of combined prediction and keep the complexity low, in this paper we present an modified combined prediction scheme including adaptively intra prediction mode selection with implicit signalling and fixed spatial-variant weighted coefficient. In Section 2, we first introduce the previous combined prediction methods. In Section 3, detailed description of the proposed scheme is presented. Experimental results and complexity analysis are shown in Section 4. And Section 5 concludes this paper.

II. Related Work

Combined prediction scheme $CIIP$ proposed in [2] use all possible combinations of existing inter-modes and intra-



(a) Original macroblock

(b) Inter residual energy surface(c) Intra residual energy surface

Fig. 2. Residual energy surface of inter and intra prediction

modes as candidate modes. For a given position (i, j), $i \in \{1, ..., 16\}, j \in \{1, ..., 16\}$ in a macroblock, predicted pixel value of new modes are defined as:

$$CIIP(i,j) = w_1 \cdot P\_Inter(i,j) + w_2 \cdot P\_Intra(i,j) \quad (1)$$

where w_1 is weighted coefficient for inter prediction block P_Inter and w_2 for intra prediction block P_Intra. Rate distortion search[3] is performed to select the best weights and best intra-modes corresponding to the best inter-modes. If the RD cost of one of the combination modes is the minimum among all the candidates modes, the residue coefficients will be coded along with weighted coefficients, motion vectors, reference frames, DQuant, sub macroblock types, Intra prediction modes and intra chroma prediction modes.

Since the method above will conduct rate distortion search for all the mode combinations and all the possible weighted coefficients, computational complexity becomes quite high. Scheme $CII\_16 \times 16$ in [4] pays more attention on the complexity reduction in which only one new mode which combines $PInter\_16 \times 16$ and $Intra\_16 \times 16\_Hor$ is added:

$$\begin{aligned} CII\_16 \times 16(i,j) &= w_1 \cdot P\_Inter\_16 \times 16(i,j) \\ &+ w_2 \cdot P\_Intra\_16 \times 16\_Hor(i,j) \end{aligned} \quad (2)$$

Furthermore, in order to avoid additional overhead bits for transmitting weighted coefficient, w_1 and w_2 are designed to be 13/16 and 3/16 through training which minimizes the SAD between the original MBs and the predicted MBs generated by using different values of weighted coefficient w_1.

III. Efficient Combined Inter-Intra Prediction Scheme

Although algorithm in [4] solved the complexity problem, there are still two problems of this method. First, only horizontal direction for intra prediction is not enough since texture in a macroblock can be in various direction. Second, in both [2] and [4] , weighted coefficient for different position in a macroblock is same, which violates the fact that pixels near to prediction samples are more related to prediction samples so intra prediction should be given higher weight for these pixels. Here, adaptive intra prediction mode selection and fixed spatial-variant weighted coefficient are proposed in order to tackle these problems and keep the complexity low in the meantime.

A. Adaptive intra prediction mode selection

As we know, direction information of image textures can be various while horizontal and vertical textures exist most. In this paper, only one additional mode $Inter\_16 \times 16$ combining with $Intra\_8 \times 8$ is added in which the direction of intra prediction is adaptively decided to be horizontal or vertical. Complexity is kept as low as [4] since only one more RDO loop is implemented in mode decision process.

To decide whether to use upper or left prediction samples, edge detection is first performed on the prediction block indicated by motion vector of $Inter\_16 \times 16$ mode. And due to the fact that one macroblock may contain more than one object or has unsmooth textures, block size for edge detection is set to be 8×8. The reason why we use $Inter\_16 \times 16$ rather than $Inter\_8 \times 8$ is that $Inter\_8 \times 8$ requires more overhead to transmit motion vectors, which may exceed the bits we save especially for some small sequences. The edge detection algorithm in [5] we adopt is based on a spatial-domain synthetic edge model, which is defined using interrelationship of two DCT edge features: horizontal and vertical:

$$\begin{cases} Horizontal\_feature : \{F_{u,0} : u = 1, 2, ...7\}. \\ Vertical\_feature : \{F_{0,v} : v = 1, 2, ...7\}. \end{cases} \quad (3)$$

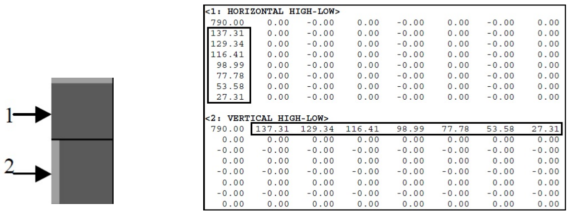
where $F_{u,v}$ represents DCT coefficient at location (u, v). To illustrate these two features, two synthetic edge models and their corresponding DCT coefficients are given in Fig.3. It is clearly seen that horizontal and vertical edges correspond to only the horizontal and vertical features. Here by performing DCT on the prediction block, horizontal and vertical features are derived and we further decide the edge information by comparing sum of these features as follows:

$$\begin{cases} Horizontal\_edge, & \text{if } \sum\limits_{u=1}^{7} |F_{u,0}| \geq \sum\limits_{v=1}^{7} |F_{0,v}|. \\ Vertical\_edge, & \text{Otherwise.} \end{cases} \quad (4)$$

In JM software[6], when encoding macroblock using mode $Inter\_16 \times 16$, current macroblock will be divided into 4 8×8 sub-blocks and processed one 8×8 block after another. So after prediction direction has been derived for each 8×8 block, $Intra\_8 \times 8$ will be implemented to get the intra prediction block since we have already obtained the upper and left prediction samples for each current 8×8 block.

B. Fixed Spatial-Variant Weighted Coefficient Derivation

After obtaining the intra prediction block and inter prediction block, weighted combination needs to be done to produce the final prediction block. In previous methods[2],



```
<1: HORIZONTAL HIGH-LOW>
 790.00     0.00    -0.00     0.00    -0.00     0.00    -0.00     0.00
 137.31     0.00    -0.00     0.00    -0.00     0.00    -0.00     0.00
 129.34     0.00    -0.00     0.00    -0.00     0.00    -0.00     0.00
 116.41     0.00    -0.00     0.00    -0.00     0.00    -0.00     0.00
  98.99     0.00    -0.00     0.00    -0.00     0.00    -0.00     0.00
  77.78     0.00    -0.00     0.00    -0.00     0.00    -0.00     0.00
  53.58     0.00    -0.00     0.00    -0.00     0.00    -0.00     0.00
  27.31     0.00    -0.00     0.00    -0.00     0.00    -0.00     0.00

<2: VERTICAL HIGH-LOW>
 790.00   137.31   129.34   116.41    98.99    77.78    53.58    27.31
   0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00
  -0.00    -0.00    -0.00    -0.00    -0.00    -0.00    -0.00    -0.00
   0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00
  -0.00    -0.00    -0.00    -0.00    -0.00    -0.00    -0.00    -0.00
   0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00
  -0.00    -0.00    -0.00    -0.00    -0.00    -0.00    -0.00    -0.00
   0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00
```

(a) Synthetic edge models (b) Corresponding DCT coefficients

Fig. 3. Illustration of horizontal and vertical features

[4], combination is performed based on macroblock size. And weighted coefficient for different location in a macroblock is same. Differently, we implement combination on 8×8 block size. And as mentioned above, prediction block should be weighted using different coefficient for different pixel locations. To verify this, optimal weighted coefficient of inter prediction and horizontal intra prediction for each location in a 8×8 block obtained by minimizing residue energy is provided as follows:

Unlike previous methods using same coefficient for all pixel locations, spatial-variant coefficient $w_1(i,j)$ for inter-predicted block P_Inter and $w_2(i,j)$ for intra-predicted block P_Intra are employed here to get prediction block $CIIS$.

$$CIIS(i,j) = w_1(i,j) \cdot P\_Inter(i,j) + w_2(i,j) \cdot P\_Intra(i,j) \tag{5}$$

The weighted coefficients satisfy $w_1(i,j) + w_2(i,j) = 1$. Then by subtracting prediction signal from original signal c, we derive the residue signal for each location (i,j), and residue energy based on statistics over one frame I can be further written as:

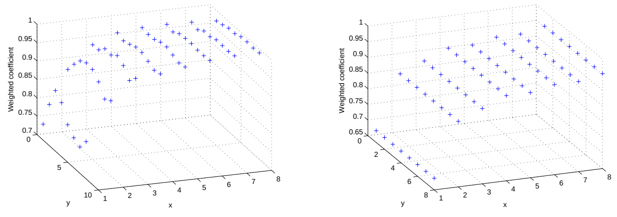
$$\begin{aligned} & E_I[R^2(i,j)] \\ & = E_I[(c(i,j) - CIIS(i,j))^2] \\ & = E_I[(w_1(i,j) \cdot R\_Inter(i,j) + w_2(i,j) \cdot R\_Intra(i,j))^2] \\ & = w_1^2(i,j) \cdot \sigma^2_{R\_Inter}(i,j) + w_2^2(i,j) \cdot \sigma^2_{R\_Intra}(i,j) \end{aligned} \tag{6}$$

where R_Inter and R_Intra denote the residue block of inter and intra prediction respectively, σ_{R_Inter} and σ_{R_Intra} denote variance of residue signal of inter and intra prediction. Here, it assumes that residue signals of both inter and intra prediction are zero-mean. To minimize residue energy of each location (i,j) in a macroblock, the derivative of $E_I[R^2(i,j)]$ with respect to the weighted coefficient w_1 is taken. This way optimal weighted coefficients w_1 and w_2 are able to be expressed as formula 8 shows.

$$0 = \frac{\partial(E_I[R^2(i,j)])}{\partial(w_1(i,j))} \tag{7}$$

$$\begin{cases} w_1(i,j) = \frac{\sigma^2_{R\_Intra}(i,j)}{\sigma^2_{R\_Inter}(i,j) + \sigma^2_{R\_Intra}(i,j)} \\ w_2(i,j) = \frac{\sigma^2_{R\_Inter}(i,j)}{\sigma^2_{R\_Inter}(i,j) + \sigma^2_{R\_Intra}(i,j)}. \end{cases} \tag{8}$$

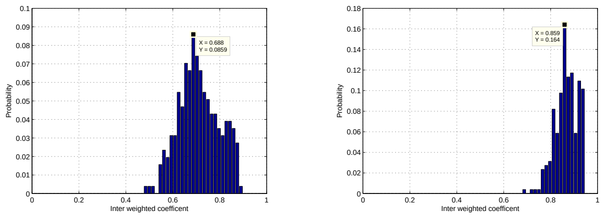
So the optimal weighted coefficient of location (i,j) can be solved by computing the variance of residue signal $R\_Inter(i,j)$ and $R\_Intra(i,j)$. Fig.4(a) shows the optimal weighted coefficients w_1 of one P frame in sequence foreman_cif.yuv. It indicates that the weighted coefficient varies for different locations and weight for inter-predicted signal becomes smaller when pixel locates closer to prediction sample.



(a) Optimal weighted coefficient (b) Fixed weighted coefficient

Fig. 4. Weighted coefficient w_1 for inter prediction

To enable implicit signalling of weighted coefficients to the decoder side, we design fixed weighted coefficient for each location in 8×8 block. Six sequences with different content features as random or fast motion, detailed texture, zoom up, panning, etc. are tested by increasing the QP by five at a time from 23 to 38. By minimizing mean square error as above, optimal weighted coefficient for each location of each row in 8×8 block is recorded for each frame. And to enable implementation using bit-shift instead of multiplication, calculated weighted coefficients are further quantized to be from 1/64 to 1 with 1/64 as step size. Fig.5 shows the probability of weighted coefficient at the first position and second position respectively. By doing so, we design the weighted coefficients for each row to be [44/64,55/64,57/64,59/64,59/64,60/64,60/64,61/64] when prediction direction is selected as horizontal. For vertical intra prediction, weighted coefficients of each column are set to be the transpose of coefficients of horizontal case. If it is horizontal intra prediction, same weighted coefficients for different rows in 8×8 block are adopted, while for vertical intra prediction, same weighted coefficients for different columns are utilized. Fig.4(b) illustrates the designed weighted coefficient for a 8×8 block in case of horizontal intra prediction.



(a) w_1 at first location (b) w_1 at second location

Fig. 5. Probability of w_1 at different locations

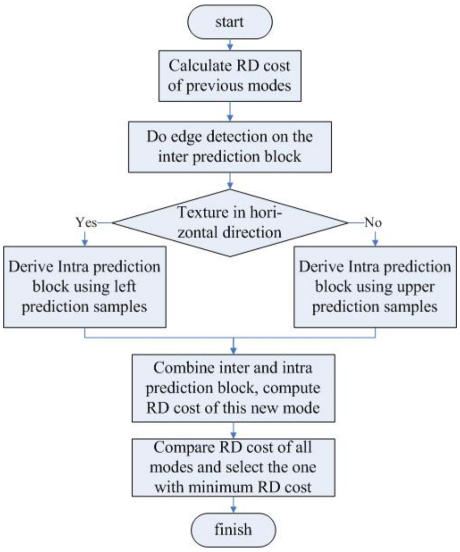

start
Calculate RD cost of previous modes
Do edge detection on the inter prediction block
Texture in horizontal direction
Yes
No
Derive Intra prediction block using left prediction samples
Derive Intra prediction block using upper prediction samples
Combine inter and intra prediction block, compute RD cost of this new mode
Compare RD cost of all modes and select the one with minimum RD cost
finish

Fig. 6. Flow chart of the proposed scheme

C. Algorithm description

For each macroblock in P frame, the proposed scheme decides whether to use the new combination mode or to use an existing mode as Fig.6 shows:

IV. EXPERIMENTAL RESULTS

In order to illustrate the performance of different algorithms, we implement the proposed scheme and the $CII\_16 \times 16$ scheme in [4] in the H.264/AVC reference software JM15.1[6] for a low delay IPPP coding using one reference frame. The reason why we employ JM15.1 rather than JSVM[7] is that JM15.1 is recent-revised version and JSVM is mainly applied in scalable video coding. Table 1 lists detailed coding conditions of implementation and five cif sequences including various texture and motion characteristics are tested here.

Experimental results of the proposed scheme and $CII\_16 \times 16$ are listed in Table 2 and they are both measured against H.264/AVC according to coding configurations in Table 1. Here, average Bjontegaard Delta Rate (BD-rate)[8] is used to measure coding gains. A negative value in the table indicates an average percentage of bit rate reduction between two rate-distortion curves. It can be seen that the proposed scheme achieves higher BD PSNR for all test cases. Compared to H.264/AVC bit rate reduction is up to 0.68%. Our method also achieves up to 0.35% additional bit rate reduction compared to the most recent combined prediction scheme $CII\_16 \times 16$. Coding gain for sequences ”foreman” and ”bus” is relatively bigger than other test sequences because the motion between neighboring frames is translational and slow, and image texture is more smooth. In this case, inter prediction using $Inter\_16 \times 16$ is able to perform better and intra prediction can further reduce prediction error for pixels near to prediction samples. While coding gain for sequence like ”stefan” and ”football” is relatively smaller because of complex textures and un-translational motion, for which smaller block size $16 \times 8, 8 \times 16, 8 \times 8..$ in the combination mode is more suitable. Table 3 shows the average usage of the new combination mode of different sequences and it is clear that encoder chooses the new mode more in the proposed algorithm.

TABLE I
CODING CONDITIONS

Parameter	**Settings**
Version of the reference software	JM15.1
Profile	Baseline(66)
GOP structure	IPPP
Intra period	0(first frame only)
Number of reference images	1
Search range	64pel
Block sizes for MCP	all 16 × 16 to 4 × 4
Rate-distortion optimization	on
Quantization parameter(I/P)	22/23,27/28,32/33,37/38
Adaptive rounding	off
CAVLC	on

TABLE II
RD PERFORMANCE OF $CII\_16 \times 16$ AND PROPOSED METHOD COMPARED TO H.264/AVC

	$CII\_16 \times 16$ .vs. H.264		Proposed .vs. H.264	
Sequence	BD Bitrate (%)	BD PSNR (dB)	BD Bitrate (%)	BD PSNR (dB)
foreman	-0.43	0.018	-0.68	0.027
football	-0.11	0.008	-0.19	0.015
stefan	-0.003	0.072	-0.14	0.007
bus	-0.339	0.017	-0.41	0.021
mobile	-0.015	0.001	-0.36	0.019
Average	-0.16	0.008	-0.36	0.018

With regard to the computational complexity, since we already have inter prediction block before computing rate distortion cost of the new mode, no additional motion estimation search needs to be done in this evaluation. And the proposed method is comparable to $CII\_16 \times 16$ since only one mode is added as candidate mode. Although the proposed method needs to implement transform on the prediction block, DCT has fast algorithm which makes it possible for real applications. Furthermore, as mentioned above, the spatial-variant weighted coefficient is designed to be fixed and able to be implemented using bit-shift. On the whole, the proposed scheme achieves a better tradeoff in terms of coding efficiency and computational complexity than [2][4].

TABLE III
AVERAGE USAGE OF COMBINED MODE

Method	$CII\_16 \times 16$	Proposed
foreman	0.96%	2.80%
football	1.21%	2.56%
stefan	1.13%	2.73%
bus	0.89%	3.30%
mobile	1.60%	4.19%

V. Conclusions

In this paper we have presented a new combined inter and intra prediction algorithm for H.264/AVC. The proposed scheme adds an extra mode $Inter\_16 \times 16$ combining with $Intra\_8 \times 8$, which has better coding performance and lower computational requirement compared with previous methods. When doing intra prediction, prediction direction is decided by detecting edge information of the corresponding inter prediction block. After deriving both inter and intra prediction block, we combine them using spatial-variant weighted coefficient which is capable of bit-shift implementation. Simulation results dedicate that the proposed scheme achieves up to 0.68% compared to H.264/AVC and 0.35% additional coding gain when compared to previous scheme using spatial-invariant weighted coefficient. In the future, combination mode which is able to select both inter and intra mode and weighted coefficient adaptively to the current macroblock will be investigated.

VI. Acknowledgement

This work has been supported in part by the the Research Grants Council (RGC) of the Hong Kong Special Administrative Region, China. (GRF Project no. 610210)

References

[1] Thomas Wiegand and Gary J. Sullivan, *Overview of the H.264/AVC Video Coding Standard*, IEEE Trans. Circuits Syst. Video Technol., vol.13, no.7, pp.560-576, Jul.2003.

[2] Kenneth Andersson, *Combined Intra Inter Prediction Coding Mode*, Picture Coding Symposium, ITU-T SG16/Q6 (VCEG), Doc. VCEGAD11, Hangzhou, China, Oct. 2006.

[3] Sullivan, Gary J., Wiegand, Thomas, *Rate-distortion optimization for video compression*, IEEE Signal Processing Magazine, pp.74-90, Nov. 1998.

[4] Xin Jin, King Ngi Ngan and Guangxi Zhu, *Combined Inter-Intra Prediction for High Definition Video Coding*, Picture Coding Symposium, Nov. 2007.

[5] M. Lee, S. Nepal, U. Srinivasan, ”Role of Edge Detection in Video Semantics,” Pan Sydney Workshop on Visual Information Processing (VIP2002), Sydney, Australia, Conferences in Research and Practice in Information Technology, 22. Jin, J. S., Eades, P., Feng, D. D. and Yan, H., Eds., pp.59-68, 2003.

[6] H.264/AVC Reference Software[Online]. Available http://iphome.hhi.de/suehring/tml/download/.

[7] H.264/SVC Reference Software[Online]. Available http://ip.hhi.de/imagecom-G1/savce/downloads/SVC-Reference-Software.htm.

[8] Gisle Bjontegaard, *Calculation of average PSNR differences between RD-curves*, in Video Coding Expert Group(VCEG), VCEG-M33, Bangkok, Thailand, April, 2001.