


# Decoder-derived Adaptive Intra Prediction (JCTVC-B080)

Yu Chen, Lu Yu

Zhejiang Univ.  
2010-07



# Outline

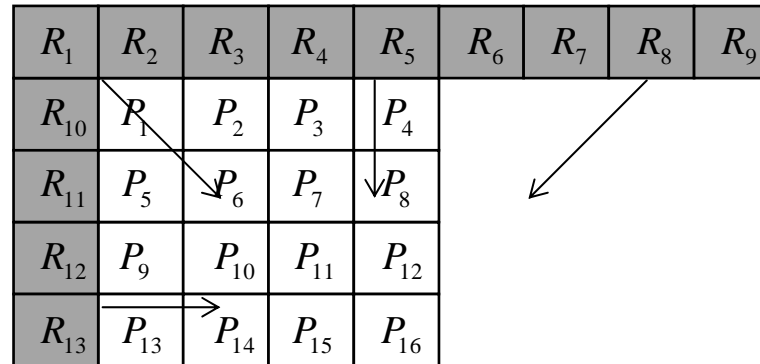
- **Motivation**
  - **Proposed Algorithm**
    - Spatial Extrapolation Prediction Filters
    - Adaptive Intra Prediction Mode Generation
    - Decoder-derived Adaptive Intra Prediction
  - **Experiments**
  - **Conclusions**
- 

# Motivation

- H.264/AVC uses a predetermined intra prediction method — 9 predetermined and fixed prediction modes
- The texture feature of video contents vary significantly by time and space which may possibly cause fixed prediction modes fail
- To capture the dynamics of different video contents, an adaptive intra prediction scheme is proposed where each intra prediction mode is regarded as a set of extrapolation prediction filters which is adaptively generated for each piece of video



# Spatial Extrapolation Prediction Filters



$\vec{p} = [P_1, P_2, \dots, P_{16}]$ , predictive pixel values of the current block

$\vec{r} = [R_1, R_2, \dots, R_{13}]$ , reference pixel values

Each intra prediction mode can be expressed by a extrapolation prediction filter matrix  $W^m$

$$\vec{p} = W^m \cdot \vec{r}$$

$$\begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{16} \end{bmatrix} = \begin{bmatrix} w_{1,1}^m & w_{1,2}^m & \dots & w_{1,13}^m \\ w_{2,1}^m & & & \\ \vdots & & \ddots & \\ w_{16,1}^m & & & w_{16,13}^m \end{bmatrix} \cdot \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_{13} \end{bmatrix}$$

# Adaptive Intra Prediction Mode Generation (1/3)

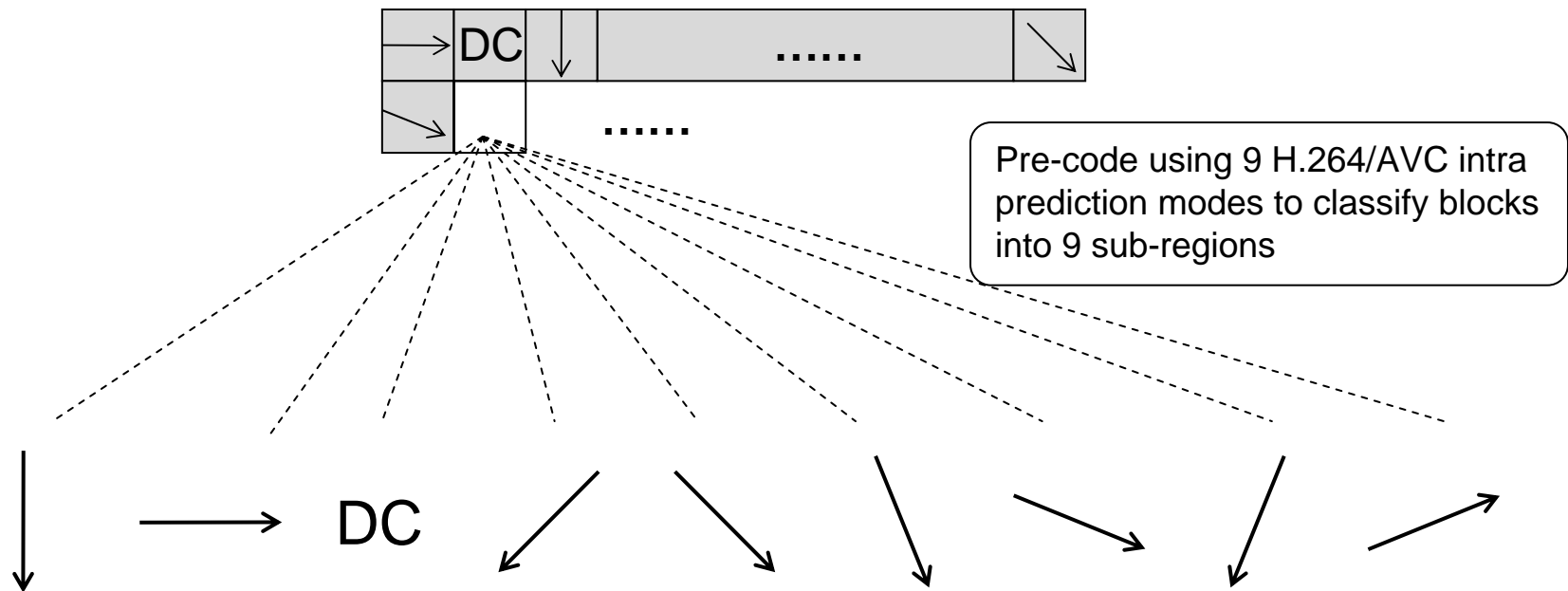
- 2-step algorithm to generate a set of 9 adaptive intra prediction modes ( $W^m$ )



# Adaptive Intra Prediction Mode Generation (2/3)

**Step1**  
Classify the coded frame into 9 sub-regions

- Each block of the frame is classified into one of the 9 sub-regions



# Adaptive Intra Prediction Mode Generation (3/3)

## Step2

Analytically calculate the optimal prediction filter matrix for each sub-region

Take mode 0 as an example

Let  $T_i^k$ ,  $P_i^k$  and  $R_j^k$  denote the original pixel values, predictive pixel values and reference pixel values of the  $k^{th}$  4x4 block in region  $G^0$  ( $i = 1, 2, \dots, 16, j = 1, 2, \dots, 13, k = 1, 2, \dots, K^0$ )

$$P_i^k = \sum_{j=1}^{13} w_{i,j}^0 \times R_j^k, \quad i=1,2,\dots,16, k=1,2,\dots,K^0;$$

Calculate  $w_{i,j}^0$  by minimizing the sum of prediction residue energy (SPRE)

$$SPRE = \sum_{k=1}^{K^0} \left\{ \sum_{i=1}^{16} [T_i^k - P_i^k]^2 \right\}$$

$$= \sum_{k=1}^{K^0} \left\{ \sum_{i=1}^{16} \left[ T_i^k - \sum_{j=1}^{13} R_j^k \times w_{i,j}^0 \right]^2 \right\}$$

$$\frac{\partial SPRE}{\partial w_{I,J}^0} = 0 \Rightarrow$$

$$\sum_{k=1}^{K^0} \left\{ \sum_{i=1}^{16} 2 \times \left( T_i^k - \sum_{j=1}^{13} R_j^k \times w_{i,j}^0 \right) \times (-w_{I,J}^0) \right\} = 0$$

$$\sum_{k=1}^{K^0} \sum_{i=1}^{16} \sum_{j=1}^{13} R_j^k \times w_{i,j}^0 \times w_{I,J}^0 = \sum_{k=1}^{K^0} \sum_{i=1}^{16} T_i^k \times w_{I,J}^0$$

$$I = 1, 2, \dots, 16, J = 1, 2, \dots, 13$$

Solve the linear equations to derive  $w_{i,j}^0$  and the prediction filter matrix

$$\begin{bmatrix} w_{1,1}^0 & w_{1,2}^0 & \dots & w_{1,13}^0 \\ w_{2,1}^0 & & & \\ \vdots & & \ddots & \\ w_{16,1}^0 & & & w_{16,13}^0 \end{bmatrix}$$

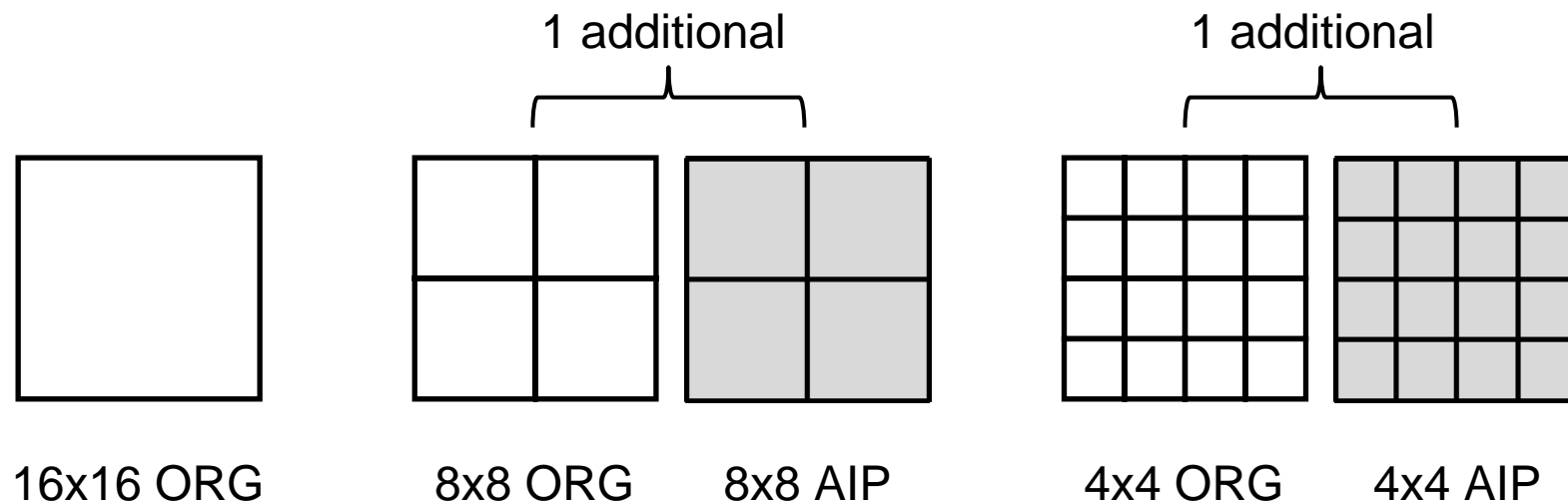
# Decoder-derived Adaptive Intra Prediction (1/2)

- To avoid transmitting prediction filter coefficients, DDAIP is proposed. Adaptive prediction modes are generated based on the previously coded and reconstructed image.
- The first part of each frame is coded using H.264 intra prediction scheme. The adaptive intra prediction modes (AIP) generated from the first part are used to code the remaining part.
- AIP modes will be updated as coding process goes on.



# Decoder-derived Adaptive Intra Prediction (2/2)

- DDAIP is implemented on 4x4 and 8x8 blocks.
- The adaptively generated intra prediction modes (AIP) are used as a complementary set for the original intra prediction modes of H.264/AVC.
- For I4x4 and I8x8 MB, 1 additional bit is transmitted to indicate whether this MB applies the original H.264/AVC prediction modes or the AIP modes.



# Experiments

The proposed intra coding scheme has been implemented in the JMKTA 2.1 software and experimental conditions are as follows:

- All CFP test sequences are tested (Class A to Class E)
- All I frame coding structure,  
frame interval according to GOP structure of Alpha anchor
- DDAIP is implemented on 4x4 and 8x8 blocks
- RDO and RDOQ enabled
- Anchor use 5 QPs {22, 26, 30, 34, 38}, DDAIP use 5 QPs {21, 25, 29, 33, 37} to match Anchor rate at each QP point

**Table 1 Bit Rate Reduction for low and high bit ranges**

| Sequence/Resolution    |                 | High Bitrate % | Low Bitrate % |
|------------------------|-----------------|----------------|---------------|
| Class A<br>2560x1600   | PeopleOnStreet  | -5.49          | -4.50         |
|                        | Traffic         | -4.40          | -4.07         |
| <b>Class A Average</b> |                 | <b>-4.94</b>   | <b>-4.29</b>  |
| Class B<br>1920x1080   | BasketballDrive | -4.36          | -4.75         |
|                        | BQTerrace       | -4.49          | -4.46         |
|                        | Cactus          | -2.81          | -2.19         |
|                        | Kimono          | -5.68          | -4.96         |
|                        | ParkScene       | -3.74          | -2.98         |
| <b>Class B Average</b> |                 | <b>-4.21</b>   | <b>-3.87</b>  |

# Experiments

**Table 1 Bit Rate Reduction for low and high bit ranges**

| Sequence/Resolution       |                 | High Bitrate % | Low Bitrate % |
|---------------------------|-----------------|----------------|---------------|
| Class C<br>832x480        | BasketballDrill | -2.99          | -2.43         |
|                           | BQMall          | -2.43          | -2.26         |
|                           | PartyScene      | -2.47          | -2.56         |
|                           | RaceHorses      | -2.15          | -2.04         |
| <b>Class C Average</b>    |                 | <b>-2.51</b>   | <b>-2.32</b>  |
| Class D<br>416x240        | BasketballPass  | -1.46          | -1.16         |
|                           | BlowingBubbles  | -0.98          | -1.46         |
|                           | BQSquare        | -1.25          | -0.92         |
|                           | RaceHorses      | -0.76          | -0.68         |
| <b>Class D Average</b>    |                 | <b>-1.11</b>   | <b>-1.05</b>  |
| Class E<br>1280x720       | vydio1          | -2.17          | -2.52         |
|                           | vydio3          | -1.87          | -2.68         |
|                           | vydio4          | -1.98          | -2.42         |
| <b>Class E Average</b>    |                 | <b>-2.01</b>   | <b>-2.54</b>  |
| <b>Class A, B Average</b> |                 | <b>-4.42</b>   | <b>-3.99</b>  |
| <b>Overall Average</b>    |                 | <b>-2.86</b>   | <b>-2.72</b>  |

# Experiments

**Table 2 Percentage of I4x4/I8x8 MBs that select DDAIP mode**

| Sequence/Resolution    |                 | I8x8 MB    |            | I4x4 MB    |            |
|------------------------|-----------------|------------|------------|------------|------------|
|                        |                 | DDAIP      | AVC        | DDAIP      | AVC        |
| Class A<br>2560x1600   | PeopleOnStreet  | 52%        | 48%        | 72%        | 29%        |
|                        | Traffic         | 73%        | 27%        | 72%        | 28%        |
| <b>Class A Average</b> |                 | <b>63%</b> | <b>37%</b> | <b>72%</b> | <b>28%</b> |
| Class B<br>1920x1080   | BasketballDrive | 48%        | 52%        | 64%        | 36%        |
|                        | BQTerrace       | 33%        | 67%        | 58%        | 42%        |
|                        | Cactus          | 39%        | 61%        | 62%        | 38%        |
|                        | Kimono          | 63%        | 37%        | 81%        | 19%        |
|                        | ParkScene       | 50%        | 50%        | 73%        | 27%        |
| <b>Class B Average</b> |                 | <b>47%</b> | <b>53%</b> | <b>67%</b> | <b>33%</b> |
| Class C<br>832x480     | BasketballDrill | 16%        | 84%        | 41%        | 59%        |
|                        | BQMall          | 24%        | 76%        | 41%        | 59%        |
|                        | PartyScene      | 25%        | 75%        | 48%        | 52%        |
|                        | RaceHorses      | 21%        | 79%        | 54%        | 46%        |
| <b>Class C Average</b> |                 | <b>22%</b> | <b>78%</b> | <b>46%</b> | <b>54%</b> |
| Class D<br>416x240     | BasketballPass  | 18%        | 82%        | 38%        | 62%        |
|                        | BlowingBubbles  | 5%         | 95%        | 34%        | 66%        |
|                        | BQSquare        | 8%         | 92%        | 42%        | 58%        |
|                        | RaceHorses      | 7%         | 93%        | 34%        | 66%        |
| <b>Class D Average</b> |                 | <b>9%</b>  | <b>91%</b> | <b>37%</b> | <b>63%</b> |
| Class E<br>1280x720    | vydio1          | 32%        | 68%        | 59%        | 41%        |
|                        | vydio3          | 30%        | 70%        | 49%        | 51%        |
|                        | vydio4          | 30%        | 70%        | 57%        | 43%        |
| <b>Class E Average</b> |                 | <b>31%</b> | <b>69%</b> | <b>55%</b> | <b>45%</b> |
| <b>Overall Average</b> |                 | <b>32%</b> | <b>68%</b> | <b>54%</b> | <b>46%</b> |

# Conclusions

- An adaptive intra prediction scheme (DDAIP) is proposed to improve prediction accuracy and better capture the dynamics of different video contents.
- In DDAIP, intra prediction modes are adaptively optimized and generated by analytical calculations according to the local feature of the coded video content.
- Experimental results show that DDAIP outperforms the H.264/AVC intra prediction scheme in all CFP test sequences (**2.86%** bit rate reduction) and is especially effective for high resolution sequences (**4.42%** bit rate reduction for Class A and B).

