Inter-Coding: multi-frame prediction

Motion Estimation in H.264

Quick Motion Estimation (as an example)
Inter-Coding: multi-frame prediction

- **Mode Decision in H.264**
  - **MV selection:** For each mode, perform multi-frame motion estimation to find the best motion vector.
    - \[ J = D + \lambda R(mvd, \text{ref}) \]
      - **Distortion estimation:** SAD of current MB and predicted MB
      - **Bit-Rate estimation:** entropy coding on mvd
  - **Mode decision:** Use rate-distortion optimization (RDO) algorithm to determine the best one among the seven modes (with selected best motion vector)
    - \[ J = D + \lambda R(mvd, \text{ref}, \text{mode}, \text{qDCTcoeff}) \]
  - The \( \lambda \) used for mv selection and mode decision are different.

- **SAD:** Sum of Absolute Difference
- **SSD:** Sum of Squared Difference
What is Entropy?

- **Statistical coding:**
  - Use less bits to common symbols and more bits to uncommon symbols.
  - Both *Arithmetic coding* and *Huffman coding* use statistical coding.

- **Entropy**
  - It is equal to the optimal number of bits to encode a symbol.
  - Entropy = $-\log_2(\text{Probability})$ or $\log_2(1/\text{Probability})$
  - Example: to compress “SQUEEZE”, with 5 symbols S, Q, U, E, Z.
    - **Without statistical coding**
      - Each symbol has probability: $1/5$  =>  Entropy = $\log_2(5)$ = 2.32bits
      - Total number of bits: $2.32 \times 7 = 16.25$bits
    - **With statistical coding**
      - Symbol E: probability = $3/7$  =>  entropy = ~1.22
      - other symbols: probability=$1/7$  =>  entropy= ~ 2.81
      - Optimal number of bits is $1.22 \times 3 + 2.81 \times 4 = 14.9$ bits
Bit-plane coding
- Bin n of a syntax of all mb in a frame form a bit-plane. A bit-plane are coded together with specific context models.

Context model: Conditional Probability
- depends on left and up mb,
- depends on previous bit of current mb.
- ... etc.

Different bit-planes may use different context model.
- Based on the binarization scheme, some bit-planes maybe better for context model using left-up mb, some are better for the model using previous bit of current mb.

<table>
<thead>
<tr>
<th>Syntax element</th>
<th>Slice type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SI/I</td>
</tr>
<tr>
<td>mb_type</td>
<td>0/3-10</td>
</tr>
<tr>
<td>mb_skip_flag</td>
<td>11-13</td>
</tr>
<tr>
<td>sub_mb_type</td>
<td>21-23</td>
</tr>
<tr>
<td>mvd (horizontal)</td>
<td>40-46</td>
</tr>
<tr>
<td>mvd (vertical)</td>
<td>47-53</td>
</tr>
<tr>
<td>ref_idx</td>
<td>54-59</td>
</tr>
<tr>
<td>mb_qp_delta</td>
<td>60-63</td>
</tr>
<tr>
<td>intra_chroma_pred_mode</td>
<td>64-67</td>
</tr>
<tr>
<td>prev_intra4x4_pred_mode_flag</td>
<td>68</td>
</tr>
<tr>
<td>rem_intra4x4_pred_mode</td>
<td>69</td>
</tr>
<tr>
<td>mb_field_decoding_flag</td>
<td>70-72</td>
</tr>
<tr>
<td>coded_block_pattern</td>
<td>73-84</td>
</tr>
<tr>
<td>coded_block_flag</td>
<td>85-104</td>
</tr>
<tr>
<td>significant_coeff_flag</td>
<td>105-165</td>
</tr>
<tr>
<td>last_significant_coeff_flag</td>
<td>166-226</td>
</tr>
<tr>
<td>coeff_abs_level_minus1</td>
<td>227-275</td>
</tr>
<tr>
<td>end_of_slice_flag</td>
<td>276</td>
</tr>
</tbody>
</table>
Entropy Coding - CABAC

**CABAC Context Model**
- Given a bit-plane of mb_type of a frame with 16 mb.
- **Assume without context model,**
  - For symbol 0 and 1, P(0)=12/16, P(1)=4/16.
  - Entropy: \( \frac{12}{16} \log_2(\frac{16}{12}) + \frac{4}{16} \log_2(\frac{16}{4}) = 0.81 \).
- **Assume context model which using left-up mb**

\[
\begin{align*}
&00: 13 \text{ times} \quad P(0|00) = \frac{12}{13} \\
&01: 1 \text{ times} \quad P(0|01) = 0 \\
&10: 1 \text{ times} \quad P(0|10) = 0 \\
&11: 1 \text{ times} \quad P(0|11) = 0
\end{align*}
\]

Entropy: \( \frac{12}{13} \log_2(\frac{13}{12}) + 1/13 \log_2(\frac{13}{1}) = 0.38 \).

\[
0.38 \times P(00) = 0.38 \times \frac{13}{16} = 0.3875 \quad \text{--> less than 0.81}
\]

(because context model with left-up mb is used)