1. Introduction

- **Task:** Professional applications often require lossless compression
- **Challenge:** Lossless compression leads to high bit rates
- **Solution:** Scalable lossless video coding based on transmitting a base layer (BL) with coarser quality and one or more enhancement layers (ELs), comprising the residual video data
- **Approach:** 3-D subband coding based on Wavelet Transform (WT) [1]

![Temporal scalability](Image)

Input video sequence

Temporal scalability $I_{t-1} \rightarrow I_t$

LP JPEG2000 BL

Spatial WT

HP JPEG2000 EL

1. By realizing $P$ as the warping operator $Y$, Motion Compensated Temporal Filtering (MCTF) is achieved [2]:

\[ h_{2t} = I_{2t} - I_{2t-1} \]

\[ I_{2t-1} = I_{2t-1} + \frac{1}{2} W_{2t-2} - t(h_{2t}) \]

2. Content Adaptive Wavelet Lifting (CA-WL)

- **Idea:** Adaptive temporal scaling based on significant changes among subsequent frames
- **Stopping Criterion:**
  - Haar WT can be represented with tree structures
  - With each node a basis vector $b_i$ and a wavelet coefficient vector $c_i$ is associated, which is the inner product of the signal $s$ with the basis $b_i$
  - If combined costs of child nodes exceed costs of parent node, i.e.
    \[ C(s, b_0, b_1, \ldots, b_i) \leq C(s, b_0, b_1, \ldots, b_{i-1}) \]
    the child nodes shall be pruned from the tree
  - $C()$ describes a Lagrangian cost functional, which represents the coding costs:
    \[ C(s, b) = D(s, b) + \lambda R(s, b) \]
  - Rate $R(s, b)$ is composed of the required rate for lossless coding of the LP and HP frames and, in case of MC, the file size of the motion vectors
  - Distortion $D(s, b)$ is calculated by the MSE of the corresponding wavelet coefficients compared to the original signal according to [3]

3. Experimental Results

- **Simulation Setup (8 bpp):**
  - Spatial resolution: Number of frames
    - AirportNight1: 640 x 480, 300
    - AirportNight2: 640 x 480, 300
    - AirportDay1: 640 x 480, 500
    - AirportDay2: 640 x 480, 500
  - Mod: 122 x 122, 30
  - HEVC: 144 x 240, 300

- **Coding parameters:**
  - LP and HP frames are encoded by JPEG2000 [5]
  - Block-based MC with block size equals $8$
  - Search range equals $8$ and is doubled for every decomposition level until a maximum size of $8$
  - Motion vectors are encoded using the QccPack library [6]

<table>
<thead>
<tr>
<th>Coding levels</th>
<th>Surv</th>
<th>Mod</th>
<th>HEVC</th>
<th>Total average</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta$ PSNR ($\Delta$s)</td>
<td>4.77</td>
<td>4.11</td>
<td>3.61</td>
<td>3.05</td>
</tr>
<tr>
<td>$\Delta$ File size [%]</td>
<td>5.99</td>
<td>4.99</td>
<td>3.48</td>
<td>2.18</td>
</tr>
<tr>
<td>$\Delta$ PSNR ($\Delta$s)</td>
<td>8.17</td>
<td>4.17</td>
<td>3.74</td>
<td>2.10</td>
</tr>
<tr>
<td>$\Delta$ File size [%]</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

4. Conclusion

- **Temporal resolution controlled by recursive application of WT**
- **Visual quality of BL is degraded by strong motion of underlying video**
- **CA-WL locally adapts temporal scaling by evaluating a Lagrangian cost functional**
- Realized by transmitting a vector $v$, whose length equals the number of input frames:
  - Initialize $v$: $(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$
  - $v$ after level $i=1$: $(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$
  - $v$ after level $i=2$: $(2, 0, 0, 2, 0, 0, 1, 0, 1, 0, 2, 0, 0, 1, 0, 1)$
  - $v$ after level $i=3$: $(3, 0, 0, 0, 0, 0, 1, 0, 1, 0, 2, 0, 0, 1, 0, 1)$
  - Non-zero entries correspond to the number of applied decomposition levels $i$
  - Distance $d$ to the corresponding HP frame is given by $d=2^{i-1}$
  - Encoded using mode-based arithmetic coding [4]

References: