A Concealment based Approach to Distributed Video Coding

Nantheera Anantrasirichai,
Dimitris Agrafiotis,
Dave Bull
Contents of Presentation

• Introduction
  ➔ Distributed Video Coding (DVC)
• Proposed schemes
  ➔ Hybrid Key/WZ frames
  ➔ Block-based concealment
• Results
• Conclusions
Introduction

- Distributed Video Coding (DVC)
  - Very low complex video transmitters
  - Two information theory principles: Slepian-Wolf (SW) and Wyner-Ziv (WZ)
  - Separate encoding, but joint decoding
  - Single source video frames are split into two categories, key frames and WZ frames
Introduction

• **Distributed Video Coding (DVC)**
  
  • Key frames are intra coded with a conventional encoder and are made available at the decoder.
  
  • WZ frame coding, which may involve transformation, uses quantisation followed by channel coding applied.
  
  • At the decoder, the key frames are used for creating an estimate of the WZ frames (side information)
  
  • The side information is seen as a noisy version of the original coded WZ frame.
  
  • Finally, the received parity bits are used to correct the errors present in this noisy version of the WZ data.
Proposed System

This paper presents a concealment based approach to distributed video coding with:

- hybrid Key/WZ frames via an FMO type interleaving of macroblocks.
- spatio-temporal concealment for generating the side information on a MB basis
- mode selection for switching between the two concealment approaches and for deciding how the correlation noise is estimated
- local (MB wise) correlation noise estimation
- modified B frame quantisation
Proposed System

- FMO DVC codec
Proposed System

- **GOP Structure**

  **IPP (Low Delay)**
  
  ![IPP Diagram]

  **IBP**
  
  ![IBP Diagram]
Proposed System Encoder

- Interleaving

The first step involves splitting of the current input frame into KEY and WZ groups, in a similar fashion to the dispersed type of flexible macroblock ordering (FMO) specified in H.264.
Proposed System Encoder

- KEY group coding

Then, the KEY MBs are horizontally shifted to make a new frame of the same height but half the width of the original which is then encoded with H.264 in intra mode.
Proposed System Encoder

- **WZ group coding**

  *Transform* might be applied to remove spatial redundancy at the expense of a slight increase in complexity.

  In addition to the matrices defined in [1] and [2], two more matrices (shown below) are proposed for very low bitrates.

  \[
  \begin{array}{cccc}
  8 & 2 & 0 & 0 \\
  2 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 \\
  \end{array}
  \quad
  \begin{array}{cccc}
  12 & 4 & 0 & 0 \\
  4 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 \\
  \end{array}
  \]

  Note than the use of 0 quantisation levels implies that no bits are sent.


Proposed System Encoder

• B-frame quantisation

To further reduce this bitrate we increase the quantisation step size of these WZ MBs relative to the step size used for the WZ MBs of the P frames.

The relationship between the number of quantisation levels $Q_B^n$ and $Q_P^n$ for B and P frames respectively, for $n$ bit-planes can be written as

$$Q_B^n = \left| a \cdot Q_P^n \right|$$

Based on experimental results, we have selected $a$ so that $Q_B^n = 1/2(Q_P^n + Q_P^{n-1})$ i.e. $a = 0.75$. 
Proposed System Encoder

- **B-frame quantisation**

<table>
<thead>
<tr>
<th># of bit-planes ($n$)</th>
<th>$Q^n_P$</th>
<th>$Q^n_B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Bits spent for each decoded bit-plane of a B-Frame WZ group for different number of quantisation levels (BER $10^{-3}$)

<table>
<thead>
<tr>
<th>Bit-plane</th>
<th>$Q^3_B = Q^3_P$</th>
<th>$Q^3_B = 0.75Q^3_P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (MSB)</td>
<td>12672</td>
<td>6336</td>
</tr>
<tr>
<td>2</td>
<td>22176</td>
<td>12672</td>
</tr>
<tr>
<td>3 (LSB)</td>
<td>31680</td>
<td>31680</td>
</tr>
</tbody>
</table>
Proposed System Decoder

- Side Information Generation

Temporal Error Concealment

Each missing WZ block is divided into four sub-blocks. Each sub-block is concealed using two motion vectors of neighbouring KEY blocks and two motion vectors generated using external boundary matching error (EBME) of its own WZ block and adjacent WZ blocks. These four possible sub-blocks are blended using a cosine weighting matrix.
Proposed System Decoder

- Side Information Generation

**Spatial Error Concealment**

SEC uses bordering KEY pixels to conceal the missing WZ pixels of each WZ block through
(a) directional interpolation or
(b) bilinear interpolation along detected edges.

Proposed System Decoder

• Side Information Generation

Mode Selection

• Evaluating the levels of motion compensated activity (TA) and spatial activity (SA) in the neighbourhood of that block

\[ SA = E \left[ (x - \mu)^2 \right] \] and \[ TA = E \left[ (x - x^*)^2 \right] \]

If (SA<TA) and (TA>3), MB is processed by SEC.

TA = mean squared error between the key blocks surrounding the missing block in the current frame and those surrounding the replacement block in the reference frame

SA = variance of the surrounding key blocks in the current frame
Proposed System Decoder

- Correlation Noise Estimation

Using the 4-neighbouring Key macroblocks of each WZ MB.

Average bits spending on the first and second bitplanes when local or global noise distribution is used

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Global</th>
<th>Local</th>
<th>% improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st</td>
<td>2nd</td>
<td>1st</td>
</tr>
<tr>
<td>Akiyo</td>
<td>8673</td>
<td>15684</td>
<td>8413</td>
</tr>
<tr>
<td>Foreman</td>
<td>8933</td>
<td>15840</td>
<td>6855</td>
</tr>
<tr>
<td>Breakdancers</td>
<td>6024</td>
<td>18696</td>
<td>3428</td>
</tr>
</tbody>
</table>
Results

Pixel domain DVC performance

Transform domain DVC performance
Results

Rate-distortion performance of Akiyo

Rate-distortion performance of Breakdancers
Results

WZ performance

[Graphs showing PSNR vs. bitrate for Akiyo, Foreman, and Breakdancers with different QP settings.]
Conclusions

This paper presents

- Extension of the Key-Wyner/Ziv framework to the IBP and transform domain DVC scenario
- Use of spatio-temporal concealment for generating the side information on a MB basis
- Mode selection for switching between the two concealment approaches and for deciding how the correlation noise is estimated
- Local (MB wise) correlation noise estimation
- Modified B frame quantisation
Thank you