ICME2012 Time Machine Session

High Order Entropy Coding – From Conventional Video Coding to Distributed Video Coding

Wenjun (Kevin) Zeng

University of Missouri

http://people.cs.missouri.edu/~zengw/

Acknowledgements: Wei Liu, Lina Dong, Yixuan Zhang, Ce Zhu
An Old Topic!  

But a new life!

- Exploiting high order statistics
  - Context modeling and conditional entropy coding
- Coding performance (PSNR) for 512x512 “lena”

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate (bpp)</td>
<td>JPEG</td>
<td><strong>Antonini ’92 ( * )</strong></td>
<td>ECSBC (ECVQ)</td>
<td><strong>EZW</strong></td>
<td>SPIHT</td>
<td>SFQ</td>
<td>EQ</td>
<td>Wu</td>
<td>JPEG 2000</td>
</tr>
<tr>
<td>0.21</td>
<td>29.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>31.6</td>
<td></td>
<td>33.98 (0.32 )</td>
<td><strong>33.17</strong></td>
<td>34.13</td>
<td>34.33</td>
<td>34.57</td>
<td><strong>34.81</strong></td>
<td>~34</td>
</tr>
<tr>
<td>0.37</td>
<td>30.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>34.9</td>
<td></td>
<td>35.96 (0.52)</td>
<td>36.28</td>
<td>37.24</td>
<td>37.36</td>
<td>37.68</td>
<td>37.92</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>37.9</td>
<td></td>
<td></td>
<td>39.55</td>
<td>40.45</td>
<td>40.52</td>
<td>40.88</td>
<td>40.85</td>
<td></td>
</tr>
</tbody>
</table>

*: Close to Woods & O’Neil’86 (SB+DPCM), Westerink’89 (SB+SQ)
Video Coding: from Hybrid to Distributed

- New applications demand paradigm-shifting approaches
  - Low-complexity wireless video, video surveillance and camera arrays, sensor networks, compression of encrypted data, etc.
  - New structure:
    - Simple encoding, but can afford complex decoding
    - Distributed encoders/processing
- Distributed video coding
  - Theoretical foundations: DSC, laid in 1970s
  - Very good codes for ideal i.i.d. sources found (since 1999)
  - How about real-world sources such as image/video? (since 2001)

<table>
<thead>
<tr>
<th></th>
<th>Coding Efficiency</th>
<th>Encoder Complexity</th>
<th>Decoder Complexity</th>
<th>Error Resilience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Hybrid</td>
<td>Highest</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td>Distributed</td>
<td>? 😊</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>
DVC Performance (2002)

Aaron et al’ 2002, r-d for WZ frames only (direct quote)

5 to 7 dB gap from H.263+ inter coding
DVC Performance (2007)

Ed. Delp, Nov. 2007, Discover Workshop (direct quote)
(odd frame: Decoder ME, MV sent back to encoder through backward channel
even frame: WZ coded)

Real gap could be 6 - 8 dB

4 to 6 dB gap (diluted!) from H.264 inter coding
DVC Performance (2009)

F. Pereira, ICME’09 (direct quote)

“it can be concluded that the SIR based IST-TDWZ codec performs better than H.264/AVC Intra for the sequences with low motion content, notably for the lower GOP sizes.”

What is wrong with the field???

RD performance, CIF, 30 Hz, GOP size 2
Conventional Video Coding

- **Major tools**
  - Motion compensated prediction
  - Spatial transform
  - High order entropy coding
    - run-length coding, end-of-block symbol
    - zero-tree coding,
    - context based coding
**Distributed Source Coding (Lossless)**

- Slepian and Wolf, 1973 (SWC)
  - No rate loss even if encode separately
- A practical approach: using **Forward Error Correcting Code** (Wyner, 1974)
  - \( R_Y = H(Y) \), \( R_X = H(X|Y) \) (treat \( Y \) as noisy version of \( X \))
- Recent advances in channel codes produces practical SWC with capacity-approaching performance
  - Turbo codes, LDPC codes

\( X: \text{i.i.d.; } Y: \text{i.i.d.} \)
A Generalized WZVC Architecture

For a single video

Video Signal

Request bits

Side info Estimation

Side Info.

Decoded Video

WZVC Encoder

WZVC Decoder

- I frames
  - Conventional Intra Encoding
  - Conventional Intra Decoding

- WZ frames
  - Transform (optional)
  - Syndrome Buffering
  - SWC Encoding
  - SWC Decoding
  - Dequantizing
  - Inverse Transform

- Side info
  - Estimation

- Request bits
Wyner-Ziv Video Coding

- Wyner-Ziv Video Coding (WZVC)
  - Shift the computational burden (exploiting redundancy) to the decoder side
  
  - The decoder usually does
    - Generate side information $Y$
    - Estimate the (noisy) channel statistics between $Y$ and the frame to be decoded ($X$)
    - Wyner-Ziv (channel) decoding

- Better performance of WZVC can be achieved if any of the above aspects is improved
Common Assumptions and Models

- Stationary motion field (?)
  - interpolation/extrapolation of motion vectors from neighboring decoded frames
Common Assumptions and Models

- Correlation between source and side information (i.e., virtual channel model)
  - Virtual channel noise: independently distributed (?)
  - Laplacian distribution
    - Variance estimated from previous decoded frames

\[ \frac{1}{2b} \exp \left( -\frac{|x - \mu|}{b} \right) \]
Progressive-learning-based DVC

- Allow decoder learning
  - Learn from lower/earlier layers and use for higher/later layers
- Allow better side info estimation
- Better exploitation of local redundancy at the decoder
- Allow employment of symmetric DSC
- Scalability
WZVC with Multi-Resolution Motion Refinement (MRMR) (Liu et al’09,10)

- **Basic idea**
  - Multi-resolution:
    - Resolution-progressive packetization and decoding
    - From low-frequency to high-frequency components.
  - **Motion Refinement:**
    - Decode a low-resolution frame using current motion field
    - Refine the motion field based on partially decoded frame
    - Go to the next resolution level, repeat the process
- **Comprehensive R-D analysis**
- **A wavelet-domain codec**
Performance Comparison

- Recall
  \[ \Delta R = R_{MCP} - R_{\text{intra}} \approx \frac{1}{8\pi^2} \int \log_2 \left[ 1 - \exp\left( -\omega^T \omega \sigma_{\Delta d}^2 \right) \right] d\omega \]
  - An equivalent transfer function \( \sqrt{1 - \exp\left( \omega^T \omega \sigma_{\Delta d}^2 (\omega) \right)} \)
  - **Smaller value \( \rightarrow \) better performance**

- Comparison
  - Encoder inter-frame coding is always the best.
  - MRMR is the worst in very low frequency, but overall much better than motion extrapolation (even for a large \( \rho \)).
Result: Sample SI Estimation (Residue)

Motion Extrapolation

MRMR
Wavelet Domain WZVC with MRMR

Decoder side diagram

- Redundant $LL_n(t-1)$
- Redundant $HL_n(t-1)$
- SI for $HL_n(t)$
- ME
- Refined Motion
- WZC$^{-1}$
- Syndromes of $HL_n(t)$
- $LL_n(t)$
- DWT$^{-1}$
- $HL_n(t)$

Bit-plane coder
Virtual channel model: Laplacian
**Result: MRMR vs. H.264/AVC Predictor**

**GOP pattern:** first frame I, all rest P/WZ; only one reference frame, QCIF

For MRMR, $B = 2$, $q = \frac{1}{4}$, $M = 8$ (*extensive motion exploration*).

For AVC baseline, $B = 4$, $q = \frac{1}{4}$, $M = 1$.

Estimate rate saving over intra-coding

- Calculate $\Delta R = \frac{1}{2} \log_2 \frac{MSE_1}{MSE_0}$, for each subband, then do a weighted average.

Very small gap is observed.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MRMR</td>
<td>-2.89</td>
<td>-1.70</td>
<td>-3.48</td>
<td>-0.42</td>
<td>-1.88</td>
<td>-1.46</td>
<td>-2.09</td>
<td>-2.80</td>
<td>-1.10</td>
<td>-2.44</td>
<td>-2.03</td>
</tr>
<tr>
<td>AVC</td>
<td>-2.58</td>
<td>-1.88</td>
<td>-3.04</td>
<td>-0.75</td>
<td>-2.03</td>
<td>-1.53</td>
<td>-2.12</td>
<td>-2.82</td>
<td>-1.31</td>
<td>-2.54</td>
<td>-2.06</td>
</tr>
</tbody>
</table>
High Rate Coding Comparison

Y components of the **P/WZ-frames only**
I followed by all P @30fps
Distributed Video Coding (DVC)

- Simple encoder, complex decoder
  - burden of removing redundancy shifted to decoder
    - Hard!
- Major steps
  - Side information estimation (at decoder)
    - ill posed problem when the source is non-stationary
    - Major obstacle
    - NOT solved until recently
  - Entropy coding
    - Channel coding approach
      - Much less flexible
    - Little effort so far for high order entropy
High Order Entropy Coding for DVC

- Now is the time!
  - Side info estimation well understood/tackled

- How?
  - Design specific channel code for virtual channel
    - Complicated (Hard!)
  - Context modeling and **bit-level** conditional probability
    - Fit channel coding paradigm well
    - Borrow ideas from conventional image/video coding!!!
Embedded Zerotree Wavelet Coding (Shapiro'93)

Fig. 6. Flow chart for encoding a coefficient of the significance map.
Can we use zero-tree in DVC???

- It relies on **direct** coding of the tree.
- Not straightforwardly!
- But concept could be borrowed.
Embedded Conditional Entropy Coding (Wu’97)

- Modeling contexts in different sub-bands

- High order statistical model at bit-level:

\[
\hat{P}(C_b \mid N_{m..b}, W_{m..b}, S_{m..b+1}, E_{m..b+1}, NW_{m..b}, NE_{m..b}, P_{m..b} \cdots)
\]

- Learn on the fly
Each coefficient in a code-block has an associated binary state variable called its *significance state.*

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_0$</td>
<td>$V_0$</td>
<td>$D_1$</td>
</tr>
<tr>
<td>$H_0$</td>
<td>$X$</td>
<td>$H_1$</td>
</tr>
<tr>
<td>$D_2$</td>
<td>$V_1$</td>
<td>$D_3$</td>
</tr>
</tbody>
</table>
Exploiting High Order Correlation for DVC (Yixuan Zhang, Ce Zhu, Wenjun Zeng’ 2012)

- Exploit the inter-coefficient correlation across scales in the wavelet domain
- Modeling of the distribution of $p(x_n|x_{n+1})$ for current symbol $x_n$ given its parent $x_{n+1}$

1 dB gain, still 4 - 5 dB from H.264 Inter.

I followed by all P/WZ.
An Old Topic!

But a new life!

~ 3 dB

Coding performance (PSNR) for 512x512 “lena”

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate (bpp)</td>
<td>JPEG</td>
<td>Antonini '92 ( *)</td>
<td>ECSBC (ECVQ)</td>
<td>EZW</td>
<td>SPIHT</td>
<td>SFQ</td>
<td>EQ</td>
<td>Wu</td>
<td>JPEG 2000</td>
</tr>
<tr>
<td>0.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*: Close to Woods & O’Neil’86 (SB+DPCM), Westerink’89 (SB+SQ)
Now is the time!
  - Side info estimation well understood/tackled

From Conventional to Distributed Video Coding
  - Leverage the past for the future!
Thanks

Q & A
References


