Low Complexity H.264/AVC Spatial Resolution Transcoding in the Transform Domain

A-S. Bacquet, C. Deknudt, P. Corlay, F-X. Coudoux, M. Gazalet

Université Lille Nord de France, IEMN UMR 8520, DOAE UVHC, France
• TOSCANE – A French ANR research project
• TOSCANE consortium:
• Aim of French ANR TOSCANE project: optimized H.264 video adaptation to DSL channel characteristics, and/or receiver capabilities thanks to
  – Video transcoding
  – Scalable video coding
• In this work:
  – Focus on H.264 spatial resolution transcoding
  – Targeted applications: low bit rate video, medium quality, over display devices with limited computational capacities

• In a first approach, we consider Intra-only coded frames, with 4x4 luminance blocks
Outline

• **Context**
• Description of the proposed H.264 spatial resolution transcoding architecture
• Simulation results
• Conclusion
Outline

• Context
• Description of the proposed spatial resolution transcoding architecture
• Simulation results
• Conclusion
The proposed architecture

• H.264/MPEG-4 AVC standard
  – Hybrid motion compensated transform compression scheme
  – New efficient coding tools
    • Intra prediction
    • Entropy coding: CAVLC, CABAC
  – Better performances compared to previous standards like MPEG-2
The proposed architecture

• Different types of spatial resolution transcoding architectures
  – Full Decode Full Recode (FDFR)
    • Optimal solution 😊
    • Computationally expensive 😞
  – Full Decode Enhanced Recode (FDER)
  – Pixel/ **Transform domain**
    • Avoid to go back to the pixel domain
    • Reduce computational cost
The proposed architecture

- **Transform Domain Manipulation (TDM):**
  - Applied on a group of 2x2 adjacent blocks of transform coefficients
  - Equivalent to geometrical zonal filtering
The proposed architecture

• 1D illustration:

Based on matrix operations...
The proposed architecture

- Let us note \((X_0, X_1, X_2, X_3)\) and \((Y_0, Y_1, Y_2, Y_3)\) the transform coefficients of two adjacent 1D-blocks or vectors of length 4.
- After frequency masking: \((X_0, X_1), (Y_0, Y_1)\).

\[
\begin{align*}
x_0 &= X_0 + X_1 \\
x_1 &= X_0 - X_1
\end{align*}
\]

\[
\begin{bmatrix}
Z_0 \\
Z_1 \\
Z_2 \\
Z_3
\end{bmatrix} =
\begin{bmatrix}
1 & 1 & 1 & 1 \\
2 & 1 & -1 & -2 \\
1 & -1 & -1 & 1 \\
1 & -2 & 2 & -1
\end{bmatrix}
\begin{bmatrix}
x_0 \\
x_1 \\
y_0 \\
y_1
\end{bmatrix}
\]

\[
\begin{bmatrix}
Z_0 \\
Z_1 \\
Z_2 \\
Z_3
\end{bmatrix} =
\begin{bmatrix}
2 & 0 & 2 & 0 \\
3 & 1 & -3 & 1 \\
0 & 2 & 0 & -2 \\
-1 & 3 & 1 & 3
\end{bmatrix}
\begin{bmatrix}
X_0 \\
X_1 \\
Y_0 \\
Y_1
\end{bmatrix}
\]

- 2D extension to convert a group of four adjacent 4x4 blocks into only one 4x4 block directly in the transform domain.
The proposed architecture

- **H264 Intra prediction**
  - In the spatial domain, typically
  - Use of neighboring decoded pixels
  - Residual error is encoded
  - Geometrical zonal filtering changes in INTRA prediction 😞
The proposed architecture

• Three consecutive steps in the algorithm

Input bitstream → Entropy Decoder → Inverse Quantization → TDM

Partial decoding in the transform domain

Buffer → Intra Prediction in transform domain → TDM

Output bitstream → Entropy Encoder → Quantization

Re-encoding

Buffer → Intra Prediction (DC mode) in transform domain
Outline

• *Context*

• *Description of the proposed spatial resolution transcoding architecture*

• *Simulation results*

• *Conclusion*
Simulation results

• Simulation setup
  – Test sequences: Crew, Soccer, Harbour, City
  – 4CIF@30Hz and CIF@30Hz formats available
  – Intra-only H.264 encoder, output bit rates range: 750 kb/s–2 Mb/s
  – Fixed QP, 4x4 block size, no deblocking
  – Luminance only

• PSNR and SSIM image quality metrics
Simulation results

The graphs depict the simulation results for Mean SSIM and Mean PSNR values for different QP parameter values. The FDFR algorithm and TDM algorithm are compared.

- **Mean SSIM value**: The graphs show a decreasing trend for both algorithms as the QP parameter value increases from 30 to 38. The FDFR algorithm has a consistently higher SSIM value compared to the TDM algorithm.

- **Mean PSNR value (dB)**: Similarly, the graphs indicate a decreasing trend for both algorithms as the QP parameter value increases. The FDFR algorithm again shows a better performance with higher PSNR values than the TDM algorithm.

These results suggest that the FDFR algorithm performs better than the TDM algorithm in terms of SSIM and PSNR under the given conditions.
Simulation results

Crew sequence, QP = 30

SSIM difference = 0.004 → visually negligible
Simulation results

Soccer sequence, QP = 24

SSIM difference = 0.089 → still visually negligible
Simulation results

• Reduced complexity
  – MPEG-4 downsampling filter:
    \[ [2 \ 0 \ -4 \ -3 \ 5 \ 19 \ 26 \ 19 \ 5 \ -3 \ -4 \ 0 \ 2]//64 \]
    \[ \rightarrow \ 6 \text{ multipliers} - 10 \text{ adders per sample} \]
    \[ \text{computartional savings} \]
  – Average processing time (4CIF@30Hz \[\rightarrow\] CIF@30Hz, Matlab, Intel Core 2 Duo CPU @2.40GHz, 3Go RAM) \textbf{reduced of about 20%} compared to FDFR
Outline

• Context
• Description of the proposed spatial resolution transcoding architecture
• Simulation results
• Conclusion
Conclusion & Perspectives

• Original spatial transcoding algorithm in the transform domain
  – Based on Transform Domain Manipulation
  – Reduced complexity compared to FDFR solution
  – Good visual quality

• Further work: extension of the proposed solution to inter-coded frames
Thank you for your attention. Any question?