Efficient Wavelet-Based Video Compression wg1n3954

N. Adami*, E. Izquierdo°, R. Leonardi*, M. Mrak°, A. Signoroni*, T. Zgaljic°

° Queen Mary Univ. London, UK
(on behalf of ACEmedia Consortium)

* University of Brescia, IT

Summary

- Where from?
- Scalability requirements in video
- Making use of time redundancy
  - Decorrelating in time
  - MCTF
  - Architectural issues (decomposition order)
- Motion Adaptive WT versus
  - JPEG 2000
  - JSVM
- Recommendations

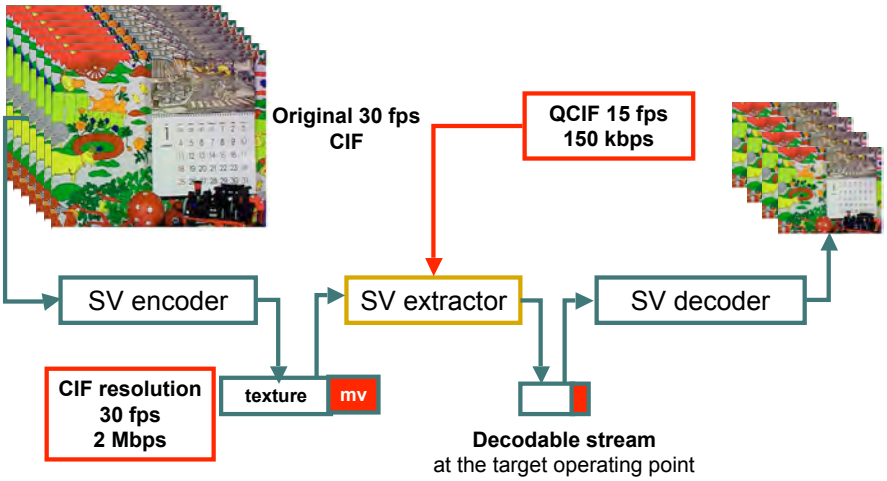
Where from?

- Research
  - Transform coding, 1970's (Jain)
  - Motion-Compensated Transform coding, 1980's (Mussman)
  - WT based compression, late 1980's (Woods, Shapiro, …)
  - Motion Compensated Temporal Filtering (Ohm 1994)
  - X-lets (late 1990's)
  - Advanced space-time transforms (late 1990's, Taubman, Pesquet-Popescu, …)

Where from?

- Standardization
  - JPEG (Transform coding, late 1980's)
  - H.261 (MC Transform coding, late 1980's)
  - MPEG1, MPEG2 (MC Transform coding with B pictures, early 1990's)
  - JPEG 2000 (WT compression, early 2000)
  - MPEG4 AVC, H.264 (MC Transform with improved entropy coding, early 2000)
  - MPEG21 SVC (Interlayer MC predictive Transform coding, currently)
  - In addition, MPEG exploration on Wavelet Video Coding
  - (stopped at the last MPEG meeting)

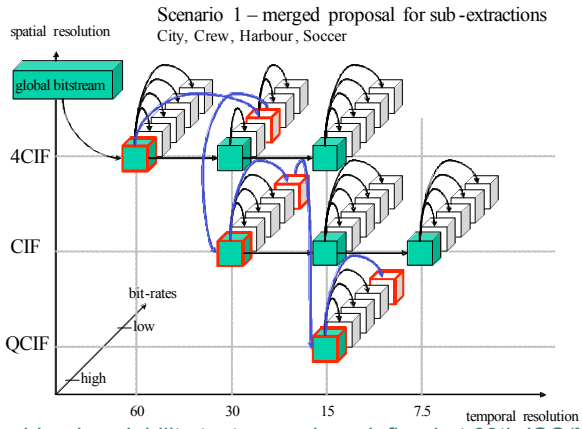
Scalable Video Coding



Scalability requirements in video

- Typical (image) scalability
  - Spatial resolution (dyadic, non dyadic)
  - Quality (SNR)
- Add temporal scalability, i.e.
  - reduced frame rate (dyadic, non dyadic)
- Encapsulation of decoded bit-stream may turn out more complex for any desirable S-T-Q decoding path.

Scalability requirements in video

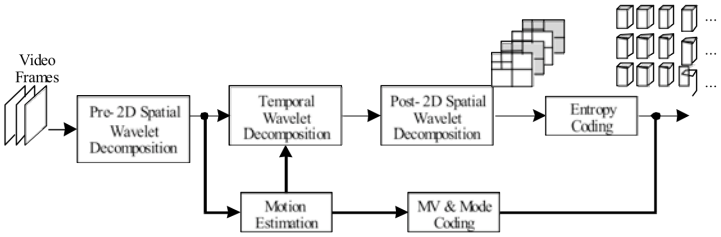


Combined scalability test procedure defined at 69th ISO/MPEG Redmond, WA, USA (w6521)

Making use of time redundancy

- Decorrelating in time
  - 3D wavelet transform (space+time)
    - > Poorly adapts to low temporal rate…
    - > strong discontinuities in time at moving object boundaries
  - As in MC Transform Coding, need to take into account motion information
- Motion Compensated Temporal Filtering
  - Effective lifting implementation for Haar Transform, 5-3 filters, along motion trajectories.

General framework for WT based video compression



Issues in WT based video compression

- Time then Space decomposition
  - Block-based motion estimation
    - > blocky displaced frame difference (DFD)
      - OBMC can reduce problem
  - DFD is not a piecewise smooth function
    - > WT transform is not optimal
    - > Less spatial Interband dependencies + difficult handling of subband coefficients due to poor energy compaction (sophistication of entropy coding)
  - Shift variant nature of the WT
    - > Non invertibility of the transform if lower spatial resolution needs to be reconstructed (unknown motion information)

Issues in WT based video compression

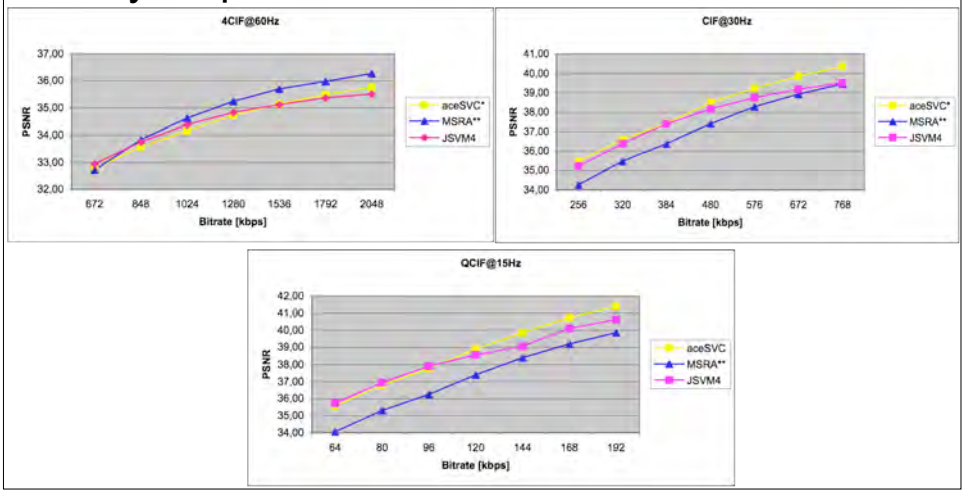
- Space then Time decomposition
  - Shift variant nature of the WT
    - > Different structure of spatial subband in presence of translation makes Motion Based Model inappropriate
      - Overcomplete spatial transform can reduce the problem
  - Motion fields can be quite different at various resolutions (loss of physical like motion)
    - > small coding efficiency in motion representation
- Layered representation (Laplacian pyramid)
  - Increased number of coefficients in transformed domain
  - May be adequate to preserve motion field consistency across layers

JSVM comparison (see W8043)

- Objective comparison (Y PSNR)
  - SNR scalability: WT
  - Spatial / Combined scalability: JSVM
- Subjective comparison
  - on average JSVM is slightly superior.

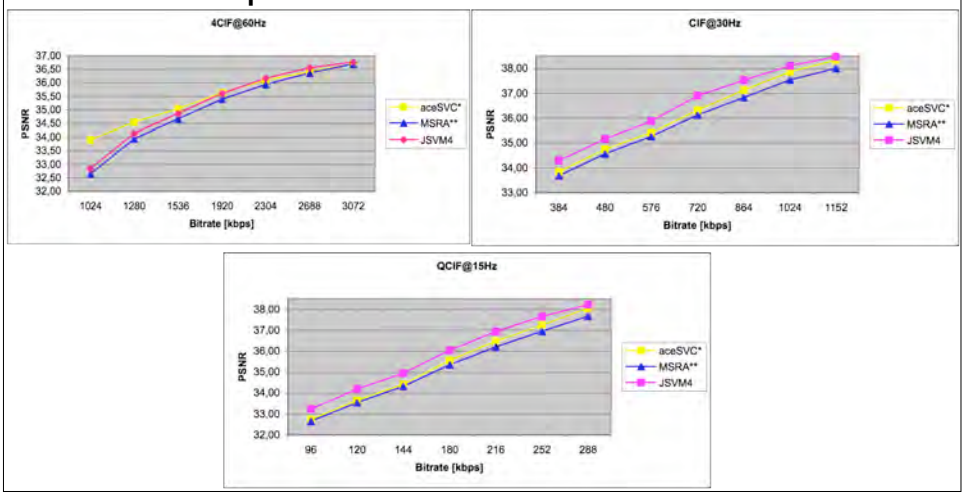
Objective performance (Y PSNR only)
SNR scalability only

City sequence



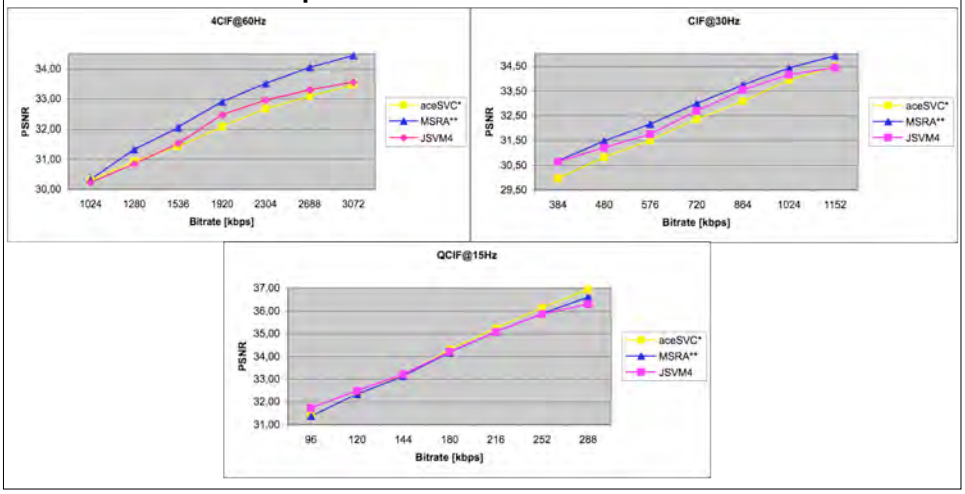
Objective performance (Y PSNR only)
SNR scalability only

Crew sequence



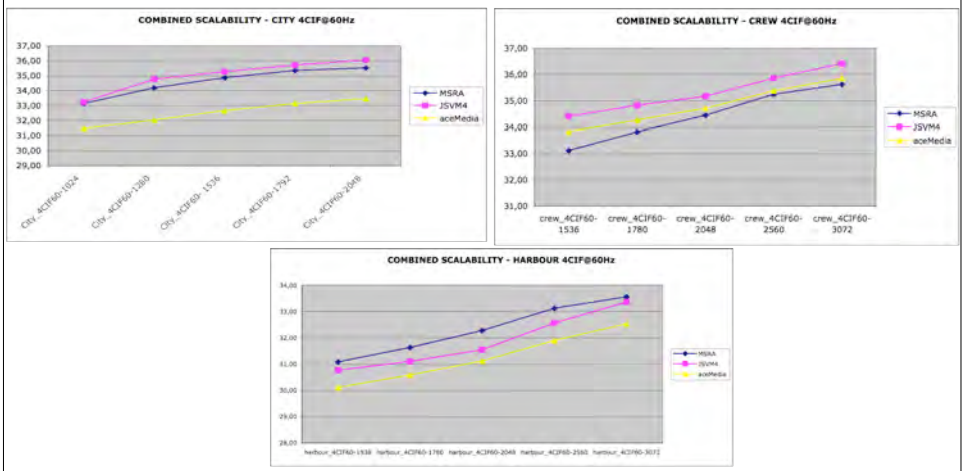
Objective performance (Y PSNR only)
SNR scalability only

Harbour sequence



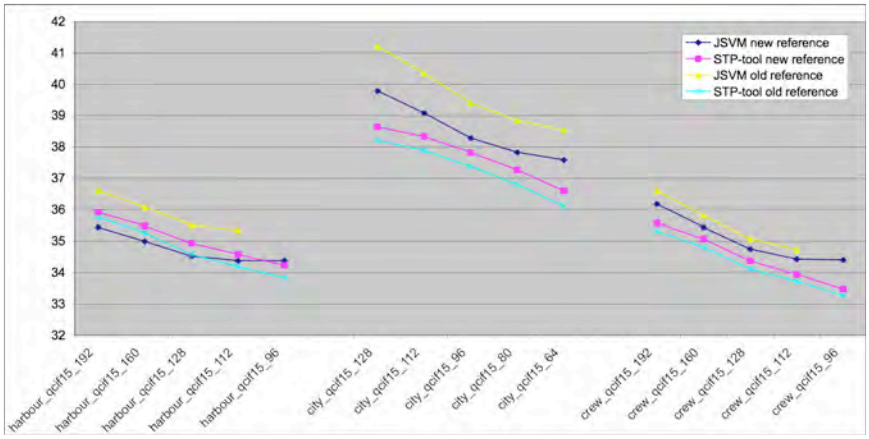
Objective performance (Y PSNR only)
Combined scalability

4CIF resolution



Objective performance (average PSNR)
Combined scalability (QCIF resol.)

Multiple reference adjustments



Visual comparison
City CIF 15fps 192kbps

STP-tool

JSVM3 (JVT)

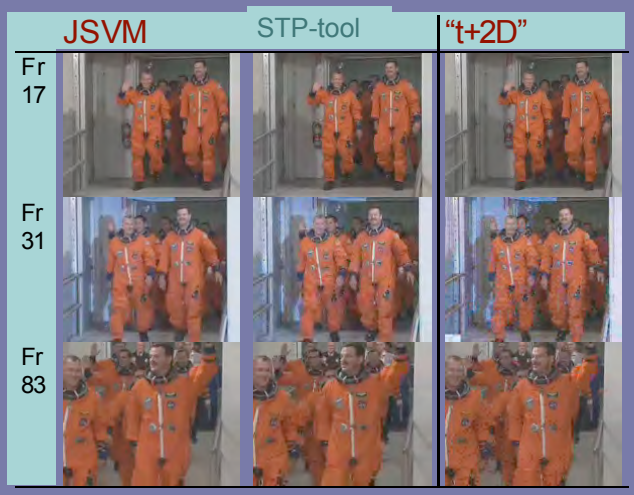
Visual Comparison
Harbour 4CIF
30fps 1024kbps

JSVM3
JVT

Visual Comparison:
Harbour 4CIF
30fps 1024kbps

STP-tool

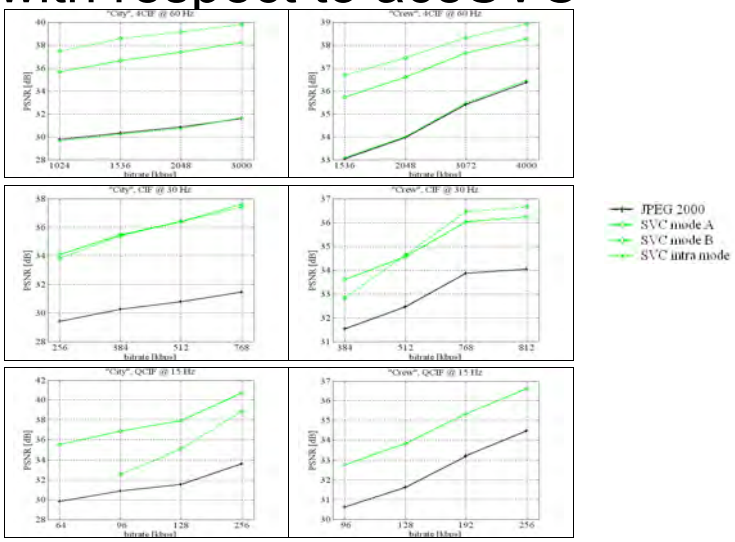
Visual Comparison:
Crew QCIF 15fps 128kbps



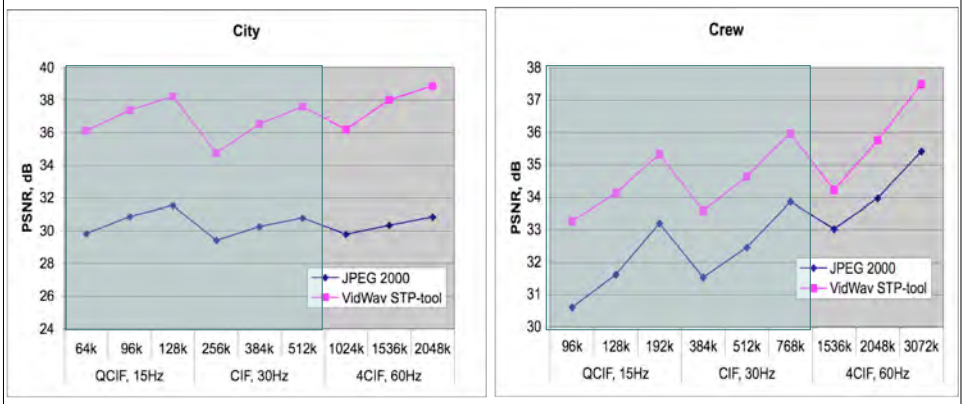
JPEG 2000 comparison

- Objective comparison (combined scalability experiment)
  - Average PSNR: (4.Y+Cr+Cb)/6
  - Motion Adaptive WT significantly superior (both aceSVC and STP-tool)
    - > performance gain for City 4-6 dB
    - > Performance gain for Crew 2-3 dB
- Subjective comparison
  - Unquestionable for City
  - Superior for Crew

JPEG 2000 comparison with respect to aceSVC



JPEG 2000 comparison with respect to STP-tool



Low spatial resolution points use different references

Concluding remarks

- For a variety of applications that handle Moving Image Sequences
  - needs to have a more effective bandwidth representation
  - ensures a baseline JPEG 2000 compatibility
- Preliminary simulations indicate
  - a substantial performance gain
  - at a slightly higher complexity
- Many additional interesting features, e.g.
  - Compressed domain processing for content description and information retrieval

THUS

- ISO/JPEG should open within AIC a study of technologies for better handling of temporal correlation