ISO/IEC JTC 1/SC 29/WG1N4049
July 7, 2006

ISO/IEC JTC 1/SC 29/WG 1
(ITU-T SG16)

Coding of Still Pictures

JBIG
Joint Bi-level Image Experts Group

JPEG
Joint Photographic Experts Group

TITLE: Performance evaluation of wavelet-based HD video coding

SOURCE: Nicola Adami, Livio Lima, Francesca Manerba, Riccardo Leonardi, Alberto Signoroni - University of Brescia, Italy

PROJECT: Advanced Image Coding

STATUS: Draft

REQUESTED ACTION: For discussion and feed-back

DISTRIBUTION: WG 1

INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG1
CODING OF STILL PICTURES

ISO/IEC JTC 1/SC 29/WG 1 N4049

Date: 2006-06-11

Title: Performance evaluation of wavelet-based HD video coding

SOURCE: Nicola Adami, Livio Lima, Francesca Manerba, Riccardo Leonardi, Alberto Signoroni
University of Brescia, Italy

---

1. Introduction

This paper is intended as a complement of document n3954 and presents some preliminary evaluation of coding efficiency of a scalable wavelet-based encoder.
Two HD video sequences have been encoded according to test conditions derived from those used to evaluate the coding efficiency of JSVM and VidWav[4].
Assuming that exploiting temporal redundancy in video coding improves compression efficiency, the aim of this work is to further investigate advantage and disadvantage of applying a motion compensated temporal filtering, in term of compression gain, with respect to pure intra coding.

2. The STP-tool codec

The codec used during the tests is an implementation of the STP-tool scheme, which has been also implemented in MPEG VidWav reference system. The STP-tool is a 2D+t+2D scalable architecture with an original inter-scale prediction mechanism. As it can be seen in Fig 1, contrarily to what happen in classical 2D+t+2D, the decoded lower spatial resolution is not interpolated before the inter-scale prediction stage. In STP-tool scheme only the LL spatial sub-band, generated by applying a one level 2DWT to the signal after the MCTF module, is predicted using the decoded lower spatial resolutions temporal sub-bands. In this way the interpolation can be avoided and the inter-scale error and details signals can be separated allowing a more flexible controls.

The STP-tool software implementation tested in the present work employ a temporal module wich a Hierarchical B Picture temporal decomposition and Unconstrained MCTF. The entropy encoder used to compress the texture is an implementation of the

wavelet-based EMDC algorithm (2D) which produce a progressive bit-stream with embedded rate distortion optimization. The coding efficiency of this codec is comparable with the one offered by JPEG2000 so the results obtained in this work can be expected also in the case the EMDC module is replaced with JPEG2000.

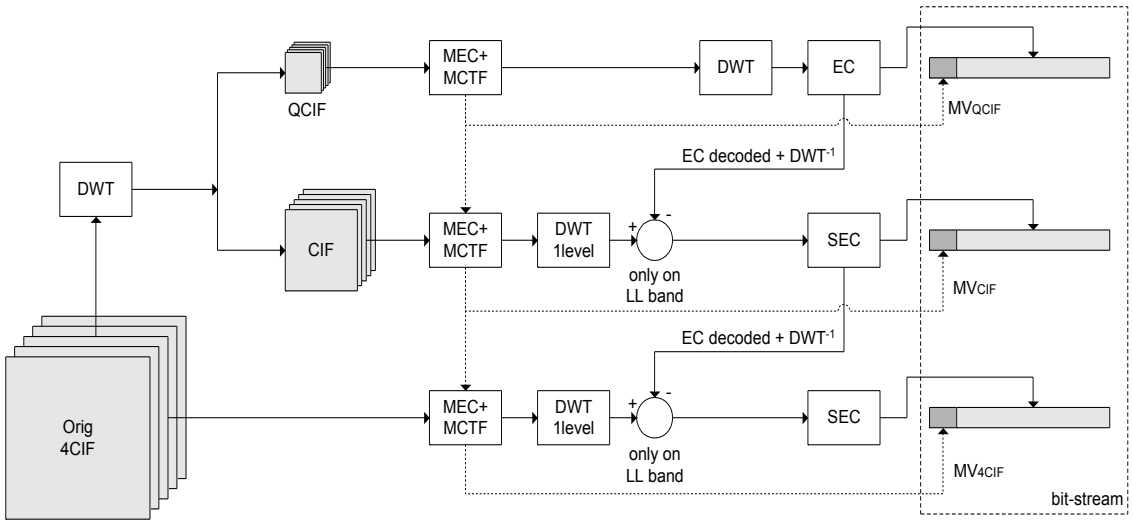

Fig 1. 2D+t+2D STP-tool scheme: ISP without interpolation

3. Tests conditions

The conditions used in the tests have been derived from document [4]. With respect to the original conditions here we changed the bit rate specifications selecting the operating point in order to decode the original video in the range of 35–40 dB. Additionally in order to conduct a more consistent evaluation, the decoded bit-streams have been extracted trying to minimize the PSNR fluctuation between adjacent frames. When temporal filtering is applied, this constraint penalizes the coding efficiency, at least in term of average PSNR.

In this work two set of coding conditions are considered for testing respectively SNR and Combined Scalability features.

SNR scalability
For the evaluation of quality scalability each spatial resolution is encoded independently. This means that no spatial scalability is embedded in the generated bit-stream. In this case the each working point is successively extracted from the highest to the lowest rate.

Combined scalability

The evaluation of combined scalability features of a given codec can be determined by building a plot of quality values of decoded bit-streams obtained by extracting a set of operating point. In this case the extracted point are characterized by different values of spatial, temporal and quality resolutions. In addition, a predefined extraction can be imposed which in principle defines the embedding strategy.

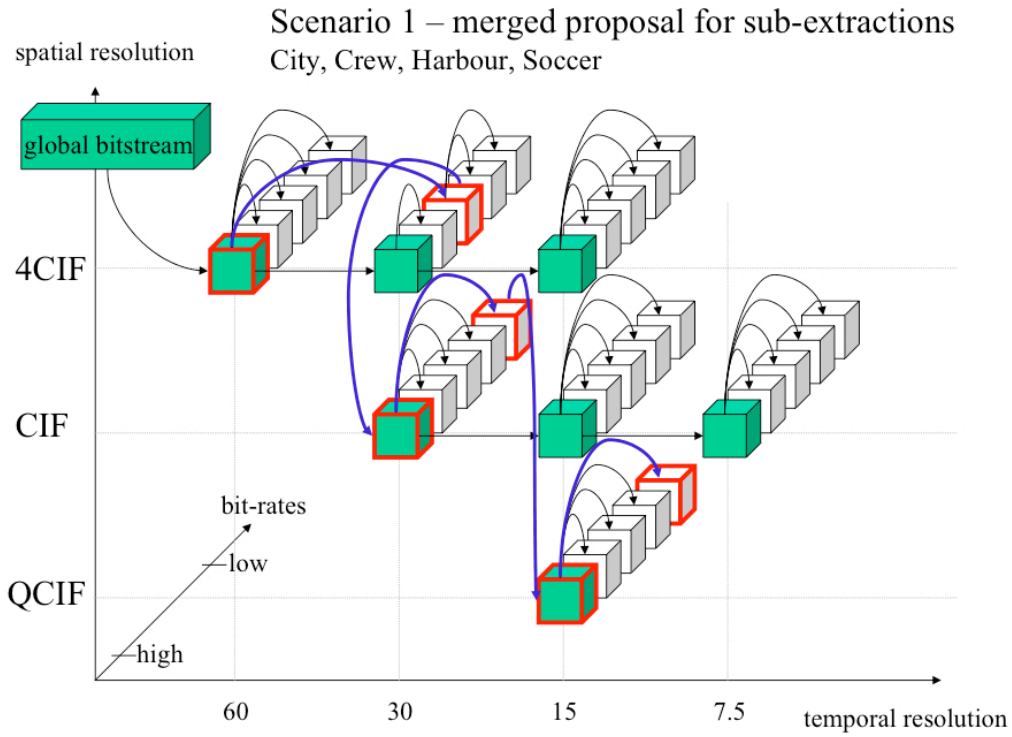

Fig 2. Example of Combined Scalability test. Procedure defined at 69th ISO/MPEGRedmond, WA, USA (w6521)

4. Test material

The High Definition video sequences have been obtained from the *European Broadcasting Union* website [2]. Among the five available sequences the two classified as "Difficult" to compress have been chosen: *CrowdRun* and *DucksTakeOff*.
As full resolution, we selected the closest to the 2K format i.e. the 1920x1080@50Hz. Due to some limitation of the tested codec, originals have been cropped to 1920x1024@50Hz. For the Combined Scalability test, two lower resolution version of the original videos have been generated by applying the 3LS and 9x7 low pass filters prior to the decimation stage. These two filters have been selected in order to have a good trade of between the inter-layer prediction effectiveness and coding efficiency. In fact, as it can be noticed in Fig 1 the residual generated by the inter-layer prediction is minimal when the low pass filters used in the image pyramid generation and the one used just before the differential node are the same.

a) b) c)

Fig 3. Sequence *City* obtained by using: a) MPEG down-sampling filter, b) 9x7 wavelet low-pass filter and c) 3LS wavelet low-pass filter.

In Fig 3 we can compare the effect of the used filters with the MPEG down-sampling filter which, could be considered not adequate for HD material due to its signal over smoothing.

CrowdRun	1920x1024@50Hz	960x512_3LS@50Hz	960x512_9x7@50Hz
DucksTakeOff	1920x1024@50Hz	960x512_3LS@50Hz	960x512_9x7@50Hz

5. Results evaluation

Hereafter are reported the results for concerning the SNR scalability test for the lowest resolution while those obtained for the highest one are only partial and could be completed during the meeting. The Combined Scalability test will require more time and a better definition of testing conditions and scenarios.

Results description

The PSNR results obtained by using the STP-tool encoder has been compared with those provided by MPEG-4 AVC high profile and MJPEG2000. In Fig 4 and Fig 5 two type of curves can be identified: those labelled with “codecName_gop4” indicating that two levels of temporal decomposition have been used and those reporting only by the name of the corresponding codec indicating the intra coding mode.

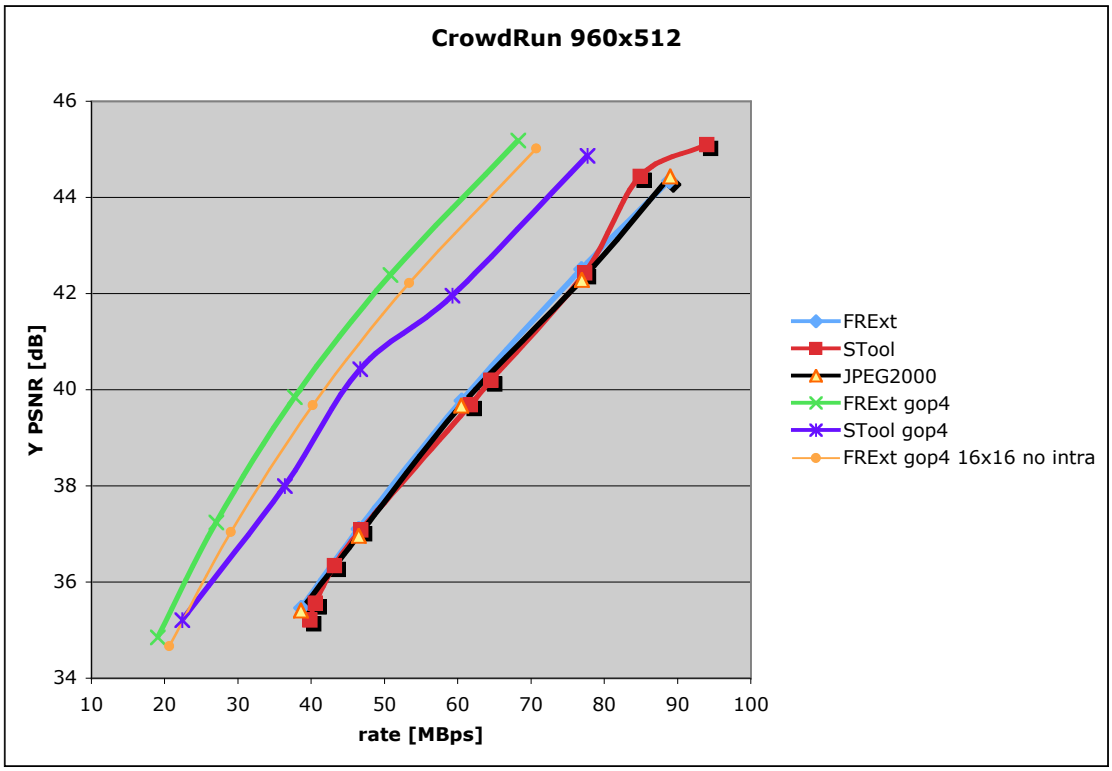

Fig 4. PSNR curves for the CrowdRun sequence

Consideration: CrowdRun

As it can be all the tested codec exhibit similar performances if intra coding is considered. It has to be mentioned that FRExt provide a single point encoding so we are comparing STP-tool and JPEG2000 with a non scalable codec. The second consideration that can be made concern the consistent coding efficiency gain obtained

by exploiting temporal redundancy.

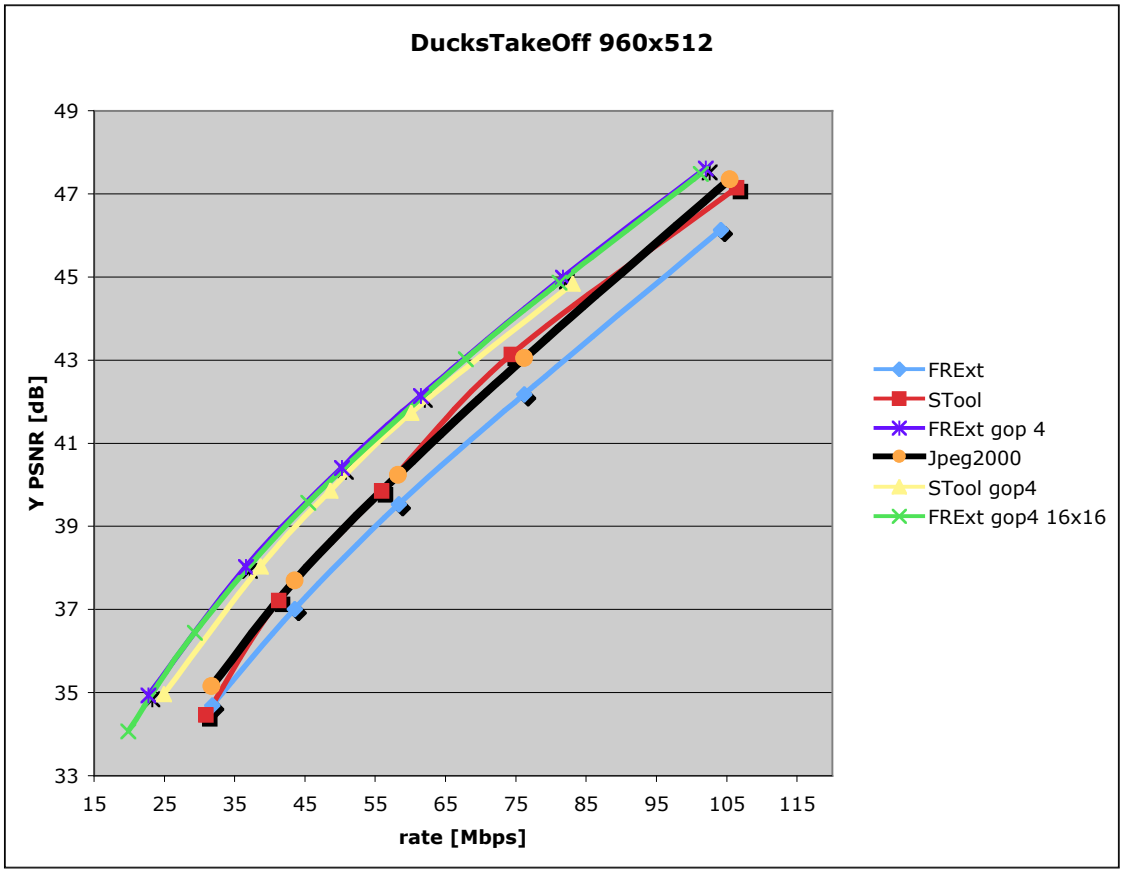

Fig 5. PSNR curves for the DuksTakeOff sequence

Comments: DuksTakeOff

In this case it can be noticed that intra coding performances of wavelet-based codecs are superior of those provided by FRExt while the gain, obtained by adding the temporal filtering, is reduced. This is also an indicator of the complexity of the considered video sequences.

6. Conclusion

This work is a preliminary attempt to test scalable video coding of new HD video material. The obtained results have been produced under redefined test conditions in order to better fit HD coding needs.
Concluding, it has been shown how the exploitation of temporal redundancy can significantly decrease the compressed bit-rate even under unfavourable coding constraints such as the near constant quality. Although, results have been obtained with experimental software, they can surely be expected for a similar architecture implementation, which employ JPEG2000 compliant technology for texture compression. It is our intention to complete tests for SNR and Combined Scalability by

also considering the adoption of fast motion estimation techniques, which remain the bottleneck in term of time needed to compress the original sequence.

7. References

[1] N. Adami, E. Izquierdo, R. Leonardi, M. Mrak, A Signoroni and T. Zgaljic, "Efficeient wavelet-based video compression", ISO/IEC JTC1/SC29/WG1, N3954, 39th JPEG Meeting, Assisi, Italy, July 2006.

[2] *European Broadcasting Union*, http://www.ebu.ch/en/technical/hdtv/test_sequences.php

[3] V. Bottreau, C. Guillemot, R. Ansari and E. Francois, "SVC CE5: spatial transform using three lifting steps filters," ISO/IEC JTC1/SC29/WG11, M11328, 70th MPEG Meeting, Palma de Mallorca, Spain, Oct. 2004.

[4] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, "Testing Conditions for Coding Efficiency and JSVM Performance Evaluation", ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-P205, 16th Meeting: Poznan, PO, 23-29 July, 2005.

[5] N. Adami, M. Brescianini, M. Dalai, R. Leonardi, and A. Signoroni, "A fully scalable video coder with inter-scale wavelet prediction and morphological coding," in Visual Communications and Image Processing (VCIP'05, SPIE, Ed., vol. 5960, Beijing, China, July 2005, pp. 535–546.

[6] R. Leonardi, T. Oelbaum, and J.-R. Ohm, "Status report on wavelet video coding exploration," ISO/IEC JTC1/SC29/WG11, 76th MPEG Meeting, Montreux, Switzerland, Tech. Rep. N8043, Apr. 2006.