INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG2005/M12643
October 2005, Nice, F

Source	University of Brescia (Italy) – Signals & Communications Laboratory
Title	**Performance evaluation of the current Wavelet Video Coding Reference Software**
Status	Contribution
Sub group	Video Group
Authors	Nicola Adami, Michele Brescianini, Riccardo Leonardi, Alberto Signoroni
	Contacts: nicola.adami@ing.unibs.it, michele.brescianini@ing.unibs.it riccardo.leonardi@ing.unibs.it, alberto.signoroni@ing.unibs.it

1 2D+t+2D architecture in the VidWav reference software

The 2D+t+2D architecture is based upon STool interband prediction scheme: see [1, 2, 3].

1.1 STool interband prediction scheme

For a detailed presentation of the STool based architecture, please see [1, 2, 3].

1.2 Modifications of Poznan VidWav reference software

Modifications to the Wavelet Video Coding Reference Software [4] to accommodate the 2D+t+2D Stool interband prediction mechanism have been described in Document [5].

1.3 Configuration parameters

Configuration parameters which enable and regulate the Wavelet Video Reference Software working in 2D+t+2D mode are described in Document [5].

1.4 Extractor and decoder

An example is shown hereafter to explain how extractor and decoder work, according to one of the multiple adaptations extraction paths defined in [6].

Considering the City test video and the three level of spatial scalability scenario, after the encoding process, three bistreams are created, one for each spatial layer (compress4CIF, compressCIF and compressQCIF): the lower ones are the base-layers for the 4CIF sequences and, in this example, the sum of the three bitstreams is equal to the highest bit-rate point of the CE1 in [6]. The bit-streams have been generated by the encoder following the multiple extraction paths constraint.

For example the sequence city_QCIF_15_96 has been extracted with these command lines:
..\bin\Extractor.Exe 1 city4CIF.cfg compress4CIF 0 0 2048 1 city4CIF60_2048.svc
..\bin\Extractor.Exe 1 city4CIF.cfg city4CIF60_2048.svc 0 1 1024 1 city4CIF30_1024.svc
..\bin\Extractor.Exe 1 cityCIF.cfg compressCIF 0 0 768 1 cityCIF30_768.svc
..\bin\Extractor.Exe 1 cityCIF.cfg cityCIF30_768.svc 0 0 384 1 cityCIF30_384.svc
..\bin\Extractor.Exe 1 cityQCIF.cfg compressQCIF 0 1 192 1 cityQCIF15_192.svc
..\bin\Extractor.Exe 1 cityQCIF.cfg cityCIF30_192.svc 0 1 96 0 cityQCIF15_96.svc
..\bin\MCWDecD.exe cityQCIF.cfg cityQCIF15_96.svc 0
rename DecSeq.yuv city_QCIF_15_96.yuv

For all other points similar scripts can be used. Multiple extraction paths as defined in [6] were always guaranteed in all experiments using the 2D+t+2D “Stool” like architecture in Vidwav Reference Software.

2 Simulation results

A comparison among the decoded sequences by JSVM3.0, VidWav reference software in “t+2D” working condition and AVC base-layer (with optimal configuration files, provided by MSRA) and Vidwav reference software in “2D+t+2D” working condition (with configuration as described in the m12642 document [3]) is reported. All the points have been extracted following the Palma extraction path and the bitstream size have been verified: VidWav reference SW in “t+2D” configuration and JSVM3.0 do not respect the bit-rate constraint in all the sequences.

All PSNR results are reported in the excel file attached to this document.

C:\HOME\alby\
RICERCA\MPEG-SVC\

NOTE: we have been unable to correctly extract some JSVM3 working points with the available configuration files.

2.1 Lowest spatial resolution results

2.1.1 PSNR comparison with original references

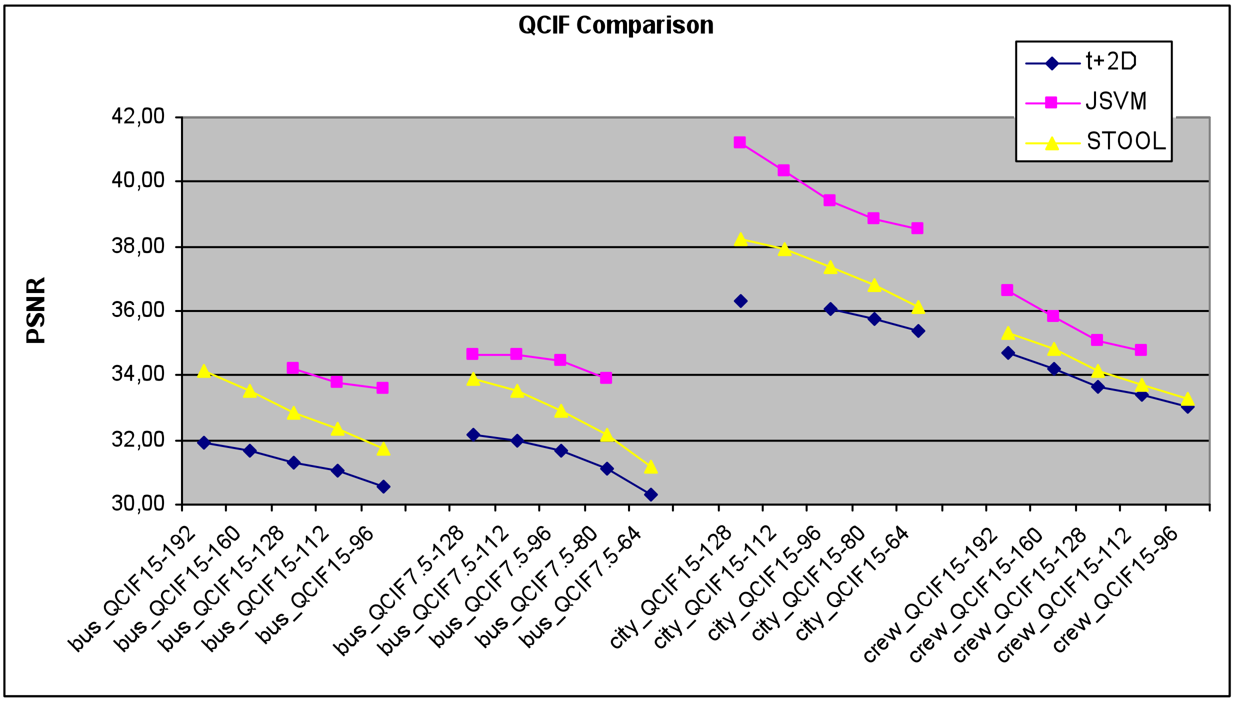
In Fig.1 we present a complete PSNR comparison at QCIF resolution. As known only trends for each system are meaningful since the different coding schemes use different reference sequences (MPEG downsampling filters for JSVM3.0, 9/7 wavelet filterbank for “t+2D” Vidwav Reference Software configuration, 3-LS filters for “2D+t+2D” configuration) relative difference in PSNR between the three coding schemes lose significance.

In the next section we propose a possible solution to this problem and we give some hints in order to “re-interpret” Fig.1 results.

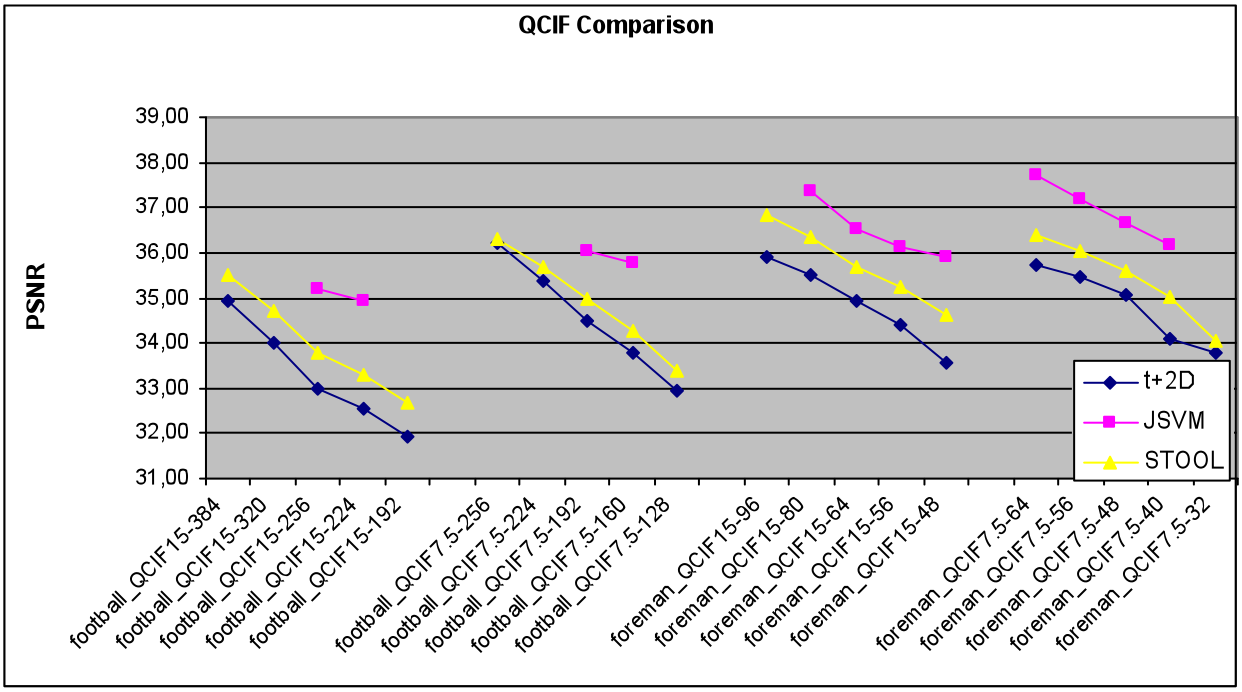
2.1.2 PSNR comparison with averaged references

A method to create a fair reference between two system with their own reference video V_1 and V_2 is to create a weighted reference $V=\alpha_1 V_1+\alpha_2 V_2$, and in particular with $\alpha_1=\alpha_2=1/2$ it can be easily verified that $PSNR(V,V_1)=PSNR(V,V_2)$. This means that V_1 and V_2 are equally disadvantaged by the creation of the common reference V, and then V is a fair common reference for both. In Figure 2 we compare the PSNR results obtained on two sequences using both system related references and a common reference for JSVM3 and STool.

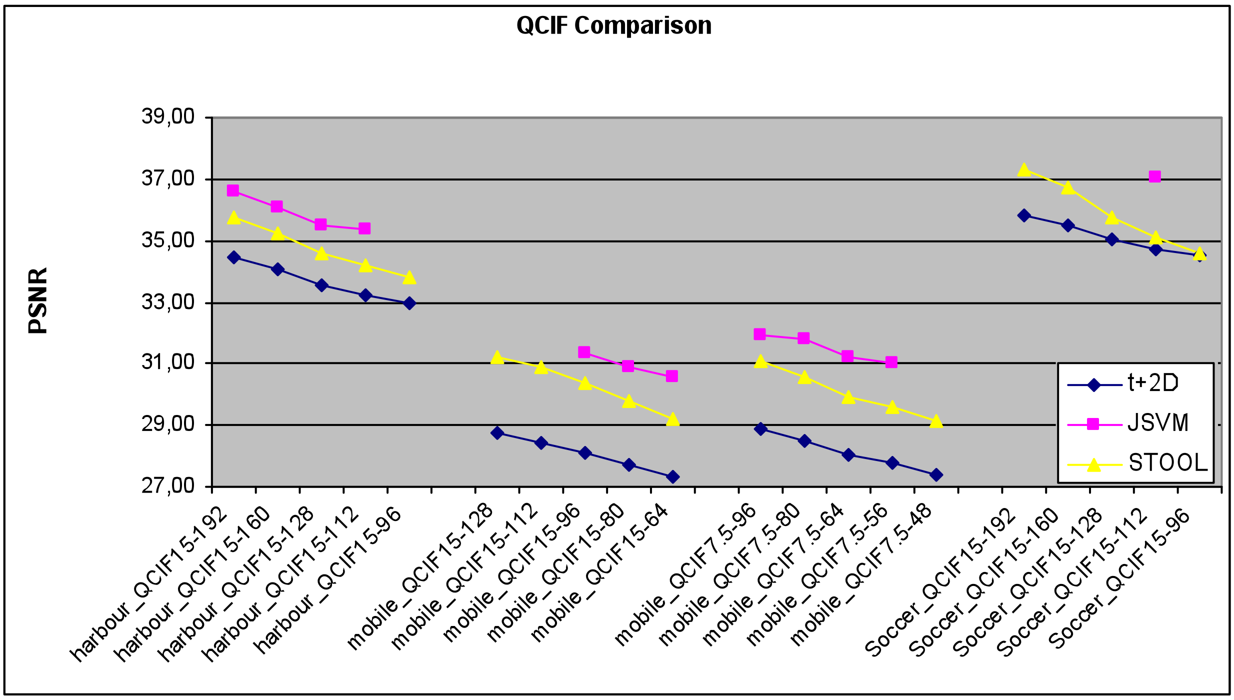
Results in Fig. 2 indicate that using a common reference, “2D+t+2D” configuration PSNR results are very close (and sometimes outperforms) those of JSVM3.



(a)



(b)



(c)

Figure 1: (a-c) PSNR comparison at QCIF resolution

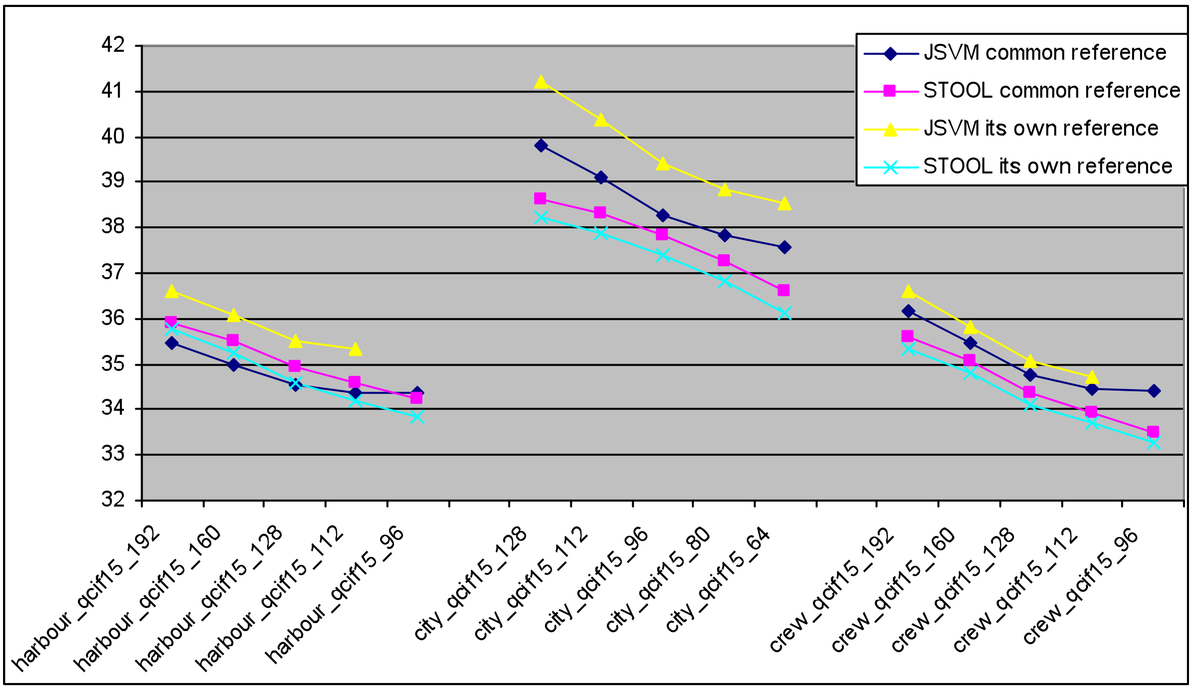


Figure 2: PSNR at QCIF resolution: common reference usage

2.1.3 Visual comparison at QCIF resolution

We show a visual comparison among some sample frames. In Fig.3 some 15fps 128kpbs decoded frames of the CREW sequence are displayed, and in Fig.4 a representative frame of the 7.5fps decoded FOOTBALL sequence is shown for 2 different bit-rates.

	JSVM	STool	“t+2D”
Fr 17			
Fr 31			
Fr 83			

Figure 3: visual comparison on CREW QCIF 15fps 128kbps

	JSVM	STool	“t+2D”
128 kbps			
256 kbps			

Figure 4: visual comparison on FOOTBALL QCIF 7.5 (frame 17)

2.2 Intermediate spatial resolution

This is the case of CIF sequences extracted from 4CIF coded bit-streams. In this situation the STool interscale prediction is applied once while the VidWav "t+2D" applies the inverse MCTF using one level downscaling of the motion field. Figure 5 shows a visual comparison on the CITY sequence. All three sequences are visually close. In Fig.6 we show some PSNR results with or without using a common reference as defined in sec. 2.2.2. Similar remarks on the common reference usage , previously made for QCIF resolution, apply also in this case.

2.3 Highest spatial resolutions

2.3.1 CIF originals

We propose a visual comparison for the sequences FOOTBALL (Fig.7) and MOBILE (Fig.8). In these cases the CIF resolution is the highest one. We remarked that from a visual point of view the decoded sequences are very close.

Figure 5: City_CIF15-192: (top) STool (192kbps) mean PSNR 34.05dB, (mid) “t+2D” ref sw (195kbps) mean PSNR 33.43dB, (bottom) JSVM3 (192kbps) mean PSNR 36.76dB

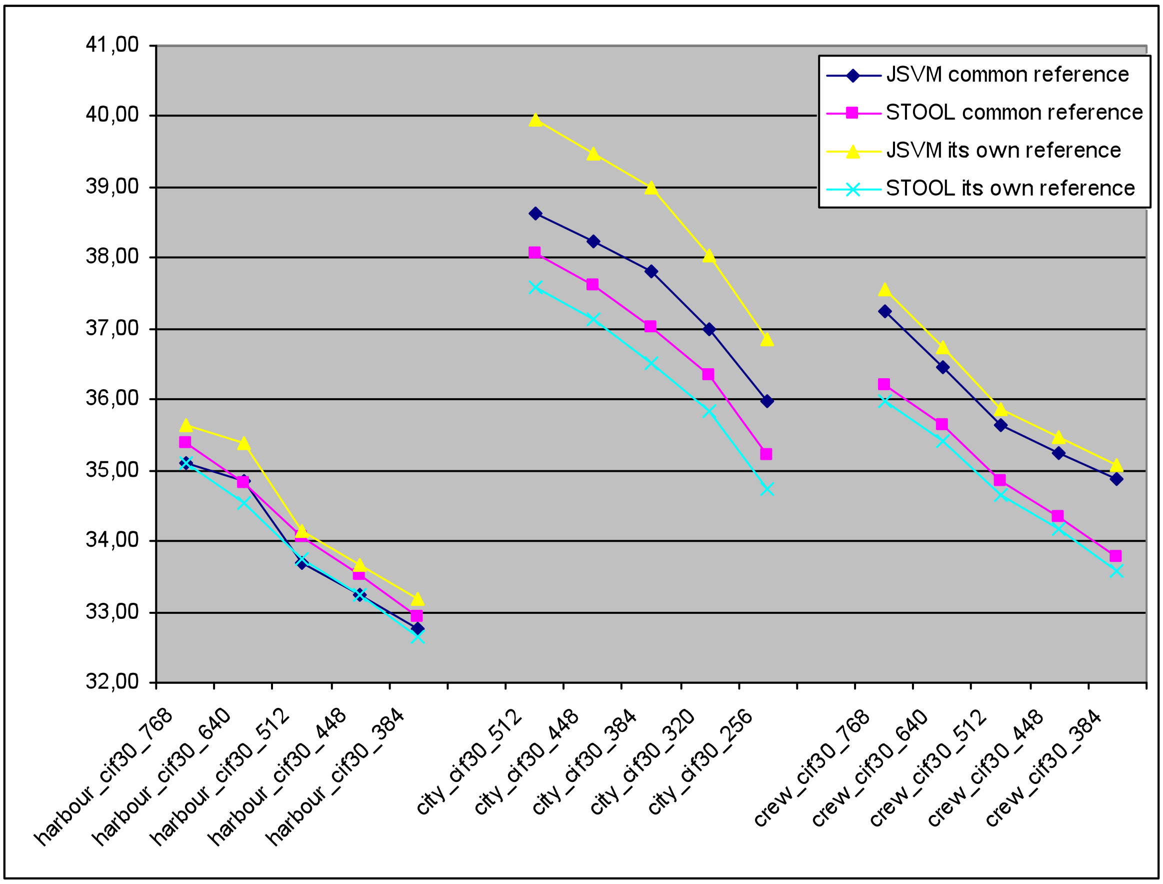


Figure 6: some PSNR results at CIF resolution (with and without using a common reference)

(a)

(b)

(c)

Figure 7: Football_CIF30-1024: (a) JSVM mean PSNR 35.95dB (b) STool 34.62dB (c) “t+2D” (1.128Mbps) 36.0db


(a)


(b)


(c)

Figure 8: Mobile_CIF30-384: (a) JSVM (384kbps) mean PSNR 31.04dB (b) STool (384kbps) 29.63dB (c) “t+2D” (**429**kbps) 31.26dB

2.3.2 4CIF originals

In this case, even if the current “2D+t+2D” STool VidWav implementation suffer from the redundancy of the motion vector representation (this find correspondence in terms of PSNR performance) visual performance remains inferior but comparable with respect to the other schemes.

(a)

(b)

(c)

Figure 9: HARBOUR 4CIF 30fps 1024kbps: (a) STool PSNR 33.02dB, (b) “t+2D” PSNR 34.45dB, (c) JSVM3 PSNR 32.58dB

3 Evaluation summary

The results can be summarized as follows:

- At highest spatial resolution experiments, “t+2D” working condition of VidWav reference software outperforms most of the time JSVM3 in terms of PSNR, but it shows drastic drop in performance at lower spatial resolution (see for example in QCIF and CIF soccer sequence artefacts).
- At lower spatial resolutions “2D+t+2D” working conditions of Vidwav reference software shows competitive performance with JSVM3.0 (visual comparison is suggested).

At the highest spatial resolution, the current “2D+t+2D” working condition of Vidwav reference software does not perform comparably since in the current implementation, motion remains independently encoded for all different spatial resolution layers, without exploiting the existing correlation between layer motions. The negative effect of this fact are more visible for low bit-rates operating points. A differential predictive scheme for motion vector coding across layers according to the STool architecture is under investigation. In addition, the current Vidwav reference software doesn’t allow to extract the lowest temporal resolution subband, so it is difficult to incorporate a good interband prediction scheme in a 2D+t+2D working condition. Finally, a entropy coding is not currently tailored to individual temporal resolutions of information detail.

4 References

[1] N. Adami, M. Brescianini, R. Leonardi and A. Signoroni, "SVC CE1: STool - a native spatially scalable approach to SVC", ISO/IEC JTC1/SC29/WG11, M11368, 70th MPEG Meeting, Palma de Mallorca, Spain, Oct. 2004.

[2] 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 Proc. of VCIP 2005, SPIE vol. 5960 (nr.58), Beijing, China, July 2005.

[3] N. Adami, M. Brescianini, R. Leonardi and A. Signoroni, "New prediction schemes for scalable wavelet video coding", ISO/IEC JTC1/SC29/WG11, M12642, 74th MPEG Meeting, Nice, France, Oct. 2005.

[4] ISO/IEC JTC1/SC29/WG11, "Wavelet Codec Reference Document and Software Manual", N7334, 73th MPEG Meeting, Poznan, Poland, July 2005.

[5] N. Adami, M. Brescianini and R. Leonardi, "Edited version of the document SC 29 N 7334", ISO/IEC JTC1/SC29/WG11, M12639, 74th MPEG Meeting, Nice, France, Oct.'05.

[6] ISO/IEC JTC1/SC29/WG11, "Description of Core Experiments in MPEG-21 Scalable Video Coding," N6521, Redmond, July 2004.