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

ISO/IEC JTC1/SC29/WG11
MPEG2006/N8043
April 2006, Montreux, Switzerland

Source Video

Title **Status Report on Wavelet Video Coding Exploration**[1]

Status Output document

Authors R. Leonardi, T. Oelbaum, J.-R. Ohm, A. Signoroni

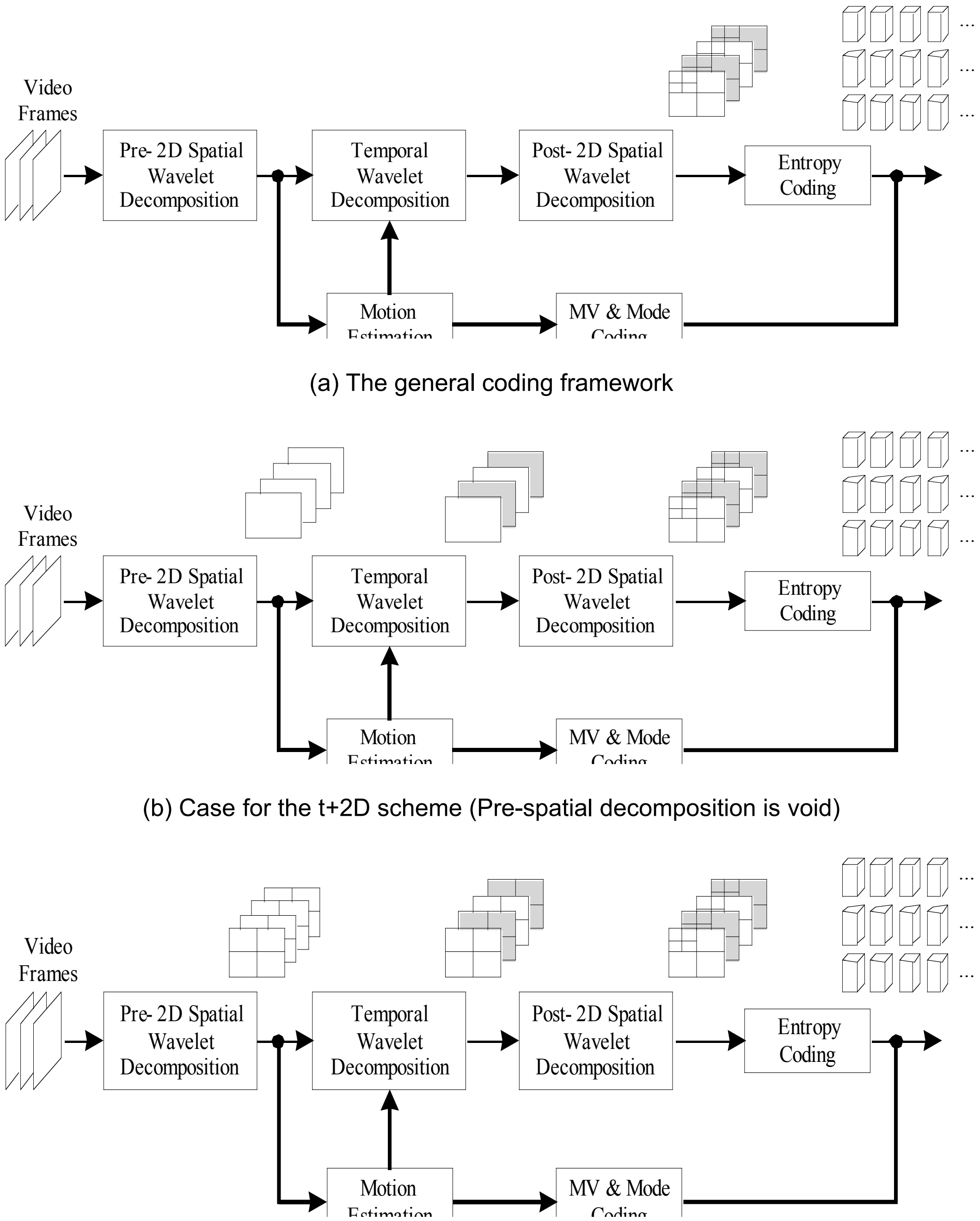
Contacts: riccardo.leonardi@ing.unibs.it, oelbaum@tum.de, ohm@ient.rwth-aachen.de, alberto.signoroni@ing.unibs.it

1 Video Coding with Wavelets

Current 3-D wavelet video coding schemes with Motion Compensated Temporal Filtering (MCTF) can be divided into two main categories. The first performs MCTF on the input video sequence directly in the full resolution spatial domain before spatial transform and is often referred to as spatial domain MCTF. The second performs MCTF in wavelet subband domain generated by spatial transform, being often referred to as in-band MCTF. Figure 1(a) is a general framework which can support both of the above two schemes. Firstly, a pre-spatial decomposition can be applied to the input video sequence. Then a multi-level MCTF decomposes the video frames into several temporal subbands, such as temporal highpass subbands and temporal lowpass subbands. After temporal decomposition, a post-spatial decomposition is applied to each temporal subband to further decompose the frames spatially.

In the framework, the whole spatial decomposition operations for each temporal subband are separated into two parts: pre-spatial decomposition operations and post-spatial decomposition operations. The pre-spatial decomposition can be void for some schemes while non-empty for other schemes. Figure 1(b) shows the case of the t+2D scheme where pre-spatial decomposition is empty. Figure 1(c) shows the case of the 2D+t+2D scheme where pre-spatial decomposition is usually a multi-level dyadic wavelet transform. Depending on the results of pre-spatial decomposition, the temporal decomposition should perform different MCTF operations, either in spatial domain or in subband domain.

[1] Editorial comment : this is an updated version of MPEG output document N7822.


Video Frames
Pre- 2D Spatial Wavelet Decomposition
Temporal Wavelet Decomposition
Post- 2D Spatial Wavelet Decomposition
Entropy Coding
Motion Estimation
MV & Mode Coding
...

(a) The general coding framework

(b) Case for the t+2D scheme (Pre-spatial decomposition is void)

(c) Case for the 2D+t+2D scheme (Pre-spatial decomposition exists)

Figure 1: Framework for 3-D wavelet video coding.

A first classification of SVC schemes according to the order the spatial and temporal wavelet transform are performed was introduced in the first Scalable Video Models [1], [2] on the base of the Call for Proposals responses at Munich meeting. The so called t+2D schemes (one example is [3]) performs first an MCTF, producing temporal subband frames, then the spatial DWT is applied on each one of these frames. Alternatively, in a 2D+t scheme (one example is [4]), a spatial DWT is applied first to each video frame and then MCTF is made on spatial subbands. A third approach named 2D+t+2D uses a first stage DWT to produce reference video sequences at various resolutions; t+2D transforms are then performed on each resolution level of the obtained spatial pyramid.

Each scheme has evidenced its pros and cons [5,6] in terms of coding performance. From a theoretical point of view, the critical aspects of the above SVC scheme mainly reside

- in the coherence and trustworthiness of the motion estimation at various scales (especially for t+2D schemes)
- in the difficulties to compensate for the shift-variant nature of the wavelet transform (especially for 2D+t schemes)
- in the performance of inter-scale prediction (ISP) mechanisms (especially for 2D+t+2D schemes).

An analysis of the differences between schemes is also reported in the sequel.

"t+2D"

A t+2D scheme acts on the video sequences by applying a temporal decomposition followed by a spatial transform. Earlier wavelet based coding systems was based on this scheme [7, 8]. Many wavelet based SVC systems are based on the t+2D spatiotemporal decomposition, and what follows is a partial reference list [3, 9–21]. Despite the t+2D is could seem simpler than other solutions, it presents some relevant issues especially for spatial scalability features.
When full spatial resolution decoding is required, the process is reversed until the desired fame-rate (partial vs complete MCTF inversion) and SNR quality; instead, if a lower spatial resolution version is needed the inversion process disclose an incoherence with respect to the forward decomposition. The problem consists in the fact that the inverse MCTF transform is performed on the lower spatial resolution (obtained by the partial inversion of the spatial DWT) of the temporal subband frames and inverse motion compensation uses the same (scaled) motion field estimated for the higher resolution sequence analysis. Because of the non ideal decimation performed by the low-pass wavelet decomposition (which generates spatial aliasing), a simply scaled motion field is, in general, not optimal to invert the temporal transform at lower resolution level. It can also be said that the motion vectors should be the same (scaled) for the various spatial resolutions since they simply record the actual physical motion at the different scales. Then the main problem seems to be spatial aliasing left in the lower resolution subbands by the non-ideal CDF 9/7 anaylsis/synthesis filters. This problem can be reduced for intermediate and lower resolutions by using (for that resolution) more selective wavelet filters [22] or locally adaptive spectral shaping acting on the quantization parameters inside each spatial subband [23]. However such approaches can determine coding performance loss at full resolution (because either wavelet filters or coefficient quantization laws are moved from coding performance *ideal* conditions).

Another relevant problem is represented by the ghosting artefacts that appears on the low pass temporal subbands when MC is not applied or when it fails due to unreliable motion vectors or to inadequate motion model. Such ghosting artefacts comes visible when high pass subbands are discarded that is when reduced framerate decoding is performed. A solution to this issue has been proposed under the framework of *unconstrained* MCTF (UMCTF) [24] which basically consists in omitting the "update" lifting step so that only the "prediction" one is performed in the lifting implementation of a MCTF. As usual in a wavelet transform framework the temporal update step is beneficial in that it smoothes the low-pass subbands (due to its action of temporal MC average) and reduces temporal aliasing. Then, omitting it cause a coding performance worsening on the low-pass temporal subbands (then on reduced frame rate decoding), however temporal averaging itself introduces ghosting artefacts where the MC model fails. A solution that try to adaptively weight the update step according to a motion field reliability model parameter has been proposed in [25, 26].

In the common motion compensated temporal filtering cases (e.g. with Haar or 5/3 kernels) an UMCTF approach actually lead to temporal open-loop versions of classical motion compensated (respectively uni- or bi-directional) temporal prediction schemes with eventually multiple reference frames, as supported in AVC. UMCTF is also used for low-delay and/or low-complexity SVC configurations (see e.g. [27]).

"2D+t"

In order to solve the problem of motion field scaling at different spatial levels a natural approach has been to consider a 2D+t scheme, where the spatial transform is applied before the temporal one. Motion information is structurally scalable here, but this does not automatically guarantee its efficient coding. The main problem of this approach is that it suffers from the shift-variant nature of wavelet decomposition, which leads to inefficiency in motion compensated temporal transforms on the spatial subbands. This problem has found a partial solution in schemes where the motion estimation and compensation take place in an overcomplete (translation invariant [28,29]) wavelet domain, but at the expense of an increasing complexity. Different coding systems have also been proposed [4, 30–39] which are based on a 2D+t wavelet spatio-temporal decomposition.

Pyramidal "2D+t+2D"

From the above discussion it comes clear that the spatial and temporal wavelet filtering cannot be decoupled because of the motion compensation. As a consequence it is not possible to encode different spatial resolution levels at once, with only one MCTF, and thus both lower and higher resolution sequences or subbands must be MCTF filtered.

In this perspective, a possibility for obtaining good performance in terms of bitrate and scalability is to use an Inter-Scale Prediction (ISP) schemes, which lead to the so called 2D+t+2D architectures. In [3] a ISP 2D+t+2D scheme has been described which derived from AVC/H.264 based proposals: prediction between the lower and higher resolutions is performed before applying spatio-temporal transform. Then the lower resolution sequence (or frame or block) is interpolated and used as prediction signal for the corresponding higher resolution signal (see Fig.2). The residual is then filtered both temporally and spatially. This architecture has a clear basis on what have been the first hierarchical representation technique, introduced for images, namely the Laplacian pyramid [40]. So, even if from an intuitive point of view the scheme seems to be well motivated, it has the typical disadvantage of overcomplete transforms, namely that of leading to a full size residual image. This way the information to be encoded as refinement is spread on a high number of coefficients and coding efficiency is hardly achievable.

Figure 2. 2D+t+2D pyramidal scheme: ISP with interpolation.

STool “2D+t+2D”

Another tool 2D+t+2D scheme, presented in [41, 42] with the name *STool* (Spatio-Temporal tool), combines a layered representation with ISP in the MCTF domain. The STool scheme is shown in Fig. 3. STool appears as a valid alternative approach to interpolation based schemes because it efficiently combines the idea of prediction between different resolution levels with the spatio-temporal wavelet transform framework. Compared with the previously described schemes it presents several advantages. First of all, the different spatial resolution levels have all undergone an MCTF, which prevents the problems of t+2D schemes. Furthermore, the MCTF are applied before spatial DWT, which solves the problem of 2D+t schemes. Moreover, the prediction is confined to the same number of transformed coefficients that exist in the lower resolution format so that there is a clear distinction between the coefficients that are associated to differences in the lowpass bands of high resolution format with respect to the low resolution ones and the coefficients that are associated to higher resolution details, and this constitutes an advantage between the prediction schemes based on interpolation in the original sequence domain. Another important advantage is that it is possible to decide which and how many temporal subbands to use in the prediction. So, one can for example discard the temporal highpass subbands if when a good prediction cannot be achieved for such “quick” details. Alternatively this allows for example a QCIF sequence at 15 fps to be efficiently used as a base for prediction of a 30 fps CIF sequence.

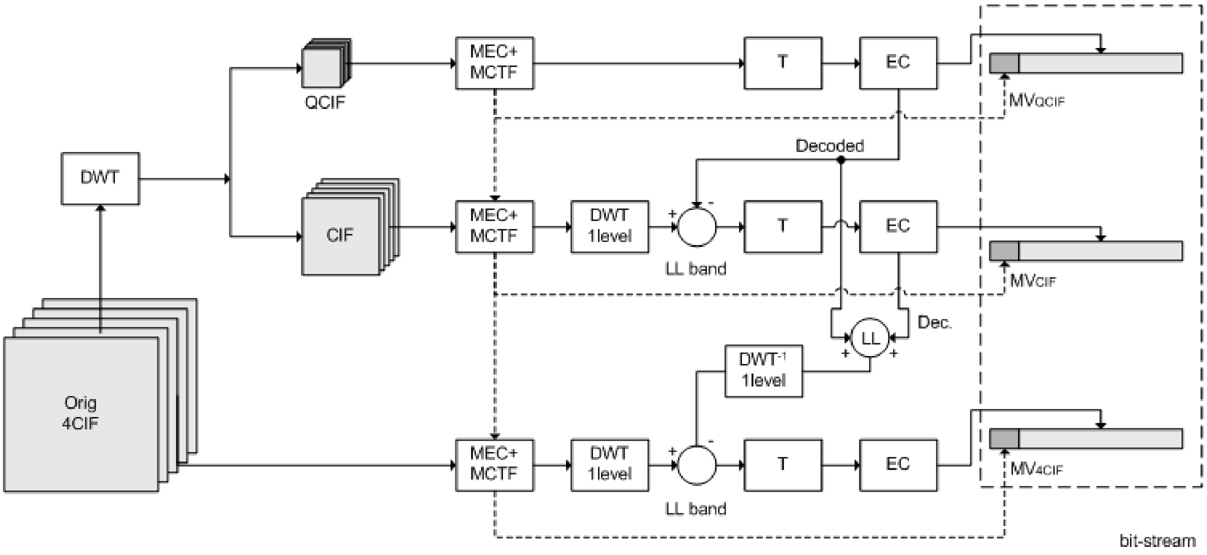


Figure 3. 2D+t+2D STool scheme: ISP without interpolation.

2D+t+2D architectures can be divided in open-loop ISP (the prediction signal is obtained from the original information) and closed-loop ISP solutions (the prediction signal is obtained from the decoded information). In a purely closed loop scheme the prediction signal used at a spatial level *s*+1 must collect all the decoded information coming from the previously coded prediction and residue signals. As both, the encoder and the decoder must use the same prediction, this could reduce scalability features. In a purely open loop scheme the signal at spatial resolution *s* is directly taken as the prediction signal, then prediction at spatial level *s*+1 only depends from the spatial level *s*. However, open loop schemes, especially at low bit-rates, undergo to the drift problems at the decoder side and then they are usually not considered. Solutions which blend the two extremes can be envisaged, as proposed in [43] within the asymmetric closed-loop prediction, in order to not sacrifice too much scalability features. Another solution which presents some similarities with respect to STool has been proposed in [44].

1.1 AhG on Further Exploration on Wavelet Video Coding (VidWav)

After the decision to proceed to the SVC standardization, jointly with ITU-T in the JVT group, with a scalable video solution based on the already mature and optimized MPEG AVC/H.264 technologies, an Ad-Hoc group on “further exploration on Wavelet Video Coding” was originated at the Palma Meeting in October 2004.

During the break-out sessions of the subsequent Meeting (Hong Kong, January 2005), several points have been discussed, which did lead to some directions for future work [45]. Because of their relevance such points are also reported in this document:

1- Goals and software issue
2- Functionalities
3- Synthesis of experiments in wavelets – starting point
4- Collaborative experiments

1) Goals and software issue. Discussions indicated that the first priority of the wavelet ad-hoc group is to collect evidence for the advantages and potential advantages which can be offered by wavelet transforms for scalable video compression. This includes both functionality and compression efficiency, both objective and subjective. The second priority of the group is to identify and evaluate the tools (technology components) which are responsible for providing the greatest coding efficiency and/or other functionalities of interest. The third priority of the group is to integrate those tools which show the greatest promise into a common software platform, which shall be based on the software offered by MSRA. MSRA software will be released 2 weeks after the Hong Kong meeting. Technical description of the software will be released 4 weeks after the Hong Kong meeting. The Software will be released under the new MPEG license policy and available through Aachen CVS repository.

2) Functionalities. The group believes that there are some interests in using wavelets, in terms of coding efficiency but also in terms of functionalities. The aim of this group is then to assess these points. To this extent the AHG will make a list of:

1- Functionalities that can be fully addressed by SVC.

2- Functionalities that can be addressed by SVC but with limitations (or that wavelets could do in a better way)

3- Functionalities that can not be addressed by SVC, but wavelets

This discussion will be developed until the next meeting on the reflector and would lead to a presentation to the next meeting.

3) Synthesis/Starting point. The aim of this item is to make a synthesis of previous experiments lead so far on wavelet technologies. The aim is to help focusing new experimentations in this AHG according to actual knowledge (drawbacks and weaknesses identified). The following list of tools category has to be further refined for the next meeting after discussions on the reflector.

a)Temporal wavelet transforms

1. The motion compensated lifting structure has proven to be most effective as a means for constructing open-loop multi-scale transforms from wavelet transform kernels, so as to exploit inter-frame redundancy with motion.
2. Wavelet kernels which have proven to be interesting include the Haar and bi-orthogonal 5/3.
3. Update steps are known to have the potential to cause ghosting artifacts at reduced temporal resolutions. Various strategies for minimizing this effect have been proposed and shown to be effective:
    a. Eliminating the update steps – some loss of compression efficiency
    b. Attenuating the update steps in regions where motion modeling is less effective
    c. 5/3 transform with uniform direction motion fields
4. Prediction steps may be understood in terms of classical motion compensated prediction (uni-directional for the Haar and bi-directional for the 5/3), except that quantization is performed out of loop. Prediction mode decisions such as those used in common video standards (everything from MPEG-1) have been found to be effective also in the context of wavelet lifting. So far prediction mode switching has been investigated only in the context of block-based motion.

b) Motion Models

1. Block based motion compensation creates discontinuities which are not well suited to subsequent application of the spatial wavelet transform.
2. One way to mitigate the above problem is the use of OBMC/deblocking.
3. A second way to mitigate the problem of block discontinuities is to perform motion compensation in the subband domain; specifically the discontinuities can be made to appear in the subband domain rather than the image domain.

c) Motion Inversion

1. Where more than one lifting step is used, the motion fields required by one lifting step (most commonly the update step) can be derived from those used by another lifting step. This has generally proven to be more effective than explicitly signaling the motion fields for all lifting steps.
2. Two types of approaches for deriving the missing lifting steps can be classified as explicit and implicit. Implicit inversion is performed by "Barbell" lifting, while explicit inversion involves the derivation of an explicit (approximate) inverse of the signaled motion field.

d) Spatio-Temporal Transform Structures

1. When the motion is well modeled and estimated, the t+2D transform structure (or equivalent) yields the highest energy compaction and hence maximizes the compression efficiency of the full resolution video.
2. At reduced spatial resolutions, the t+2D structure can lead to the appearance of artifacts. At lower bit-rates, quantization errors may mask these artifacts; however, for a fully scalable scheme, such masking cannot be relied upon. These artifacts can be eliminated by resorting to a multi-resolution structure and excluding higher resolution subbands from the motion compensation of lower resolution subbands. However, such an approach necessarily reduces full resolution compression efficiency.
3. Schemes to blend the strategies described above have shown to be promising.

e) Visual Properties of Low Spatial Resolution Scales

1. It is known that spatial DWT kernels commonly used for image compression, such as the 9/7, produce significant levels of aliasing in the LL subband frames. The aliasing becomes particularly visible (as a non-shift invariant component) in the presence of motion, where it can be very disturbing.
2. One way to reduce the aliasing problem mentioned above is to use longer DWT kernels. In particular, 3 lifting step kernels have been shown to yield reduced levels of low-resolution aliasing with some small sacrifice in full resolution compression performance.
3. Another way to reduce the aliasing problem is to use the MPEG B filters or similar, but these essentially necessitate the use of a redundant spatial pyramid.
4. The t+2D structure also produces less aliasing power at reduced spatial resolutions than schemes which exclude higher frequency subbands from the motion compensation of lower frequency subbands (as in point 2 of the previous section).

f) Impact of Scalability

1. Spatial scalability presents probably the greatest difficulties for a fully embedded coder. One reason for this is that motion bit-rate must be scaled substantially to

accommodate the large range of bit-rates expected across different spatial resolutions. Another reason relates to the visual issues described above.

2. Motion scalability is also important at lower bit-rates even within a single spatial resolution.
3. Scalability appears to come with some cost, but we don't know how large this is at present. One difficulty presented by scalability is the selection of RD optimization operating points to balance the contributions of motion and texture information which interact in a non-linear way.

g) Interesting Technologies Proposed So Far

1. Entropy coding strategies: ESCOT, EBCOT, EZBC
2. Down sampling filters: 9/7; 3-lifting step filter; MPEG B-filter
3. Various intra-coding strategies
4. Motion compensation strategies: various forms of OBMC/deblocking; various approaches to in-band MC
5. Post-processing: deringing/deblocking filters
6. Various techniques for scalable motion
7. 3-band temporal decomposition and techniques for achieving more uniform quality from frame to frame

4) Collaborative work. In order to evaluate and improve tools in wavelet video coding, a first set of tools to be studied has been defined: a) motion estimation, b) entropy coding. To this extent collaborative work has to be done to:

1. Provide a means for consistent interchange of motion parameters between implementations, including the SVM, for the purpose of isolating inconsistencies which may be attributable to motion and identifying more carefully the benefits/weaknesses associated with various transform structures.
2. Provide a means for consistent interchange of spatio-temporal subband frames between different coder implementations, for the purpose of identifying the impact of different entropy coding strategies.

The AhG on VidWav also decided to continue explorative activities [46] and to adopt a reference model and software based on the MSRA SVC software [47].

In the reference model three working modalities for wavelet video coding have been considered [47, 48]:

- A t+2D architecture as described [11, 49]
- A 2D+t(+2D) architecture (In-band temporal filtering) as described in [39, 50]
- A 2D+t+2D ISP (STool) architecture as described in [41, 43]

2 Tailored Wavelet Video Coding applications and functionalities

Wavelet video coding appears promising for much functionality such as:

1. Targeting storage of high definition content (no delay constraint), with non predefined scalability range. Inbuilt scalability brought by wavelets allows a very high definition coding with quality up to lossless, and a very low definition decoding (in case of quick preview of the content).

2. Targeting a very high number of spatio-temporal decomposition levels. Scalability is mainly designed to encode once, serve all. Wavelets allow a single encoding, and can serve all spatio-temporal decomposition levels (from QQCIF to HD, and even higher resolution).

3. Targeting non dyadic spatial resolution. Basically, wavelets are interesting also for mobile video; one knows that mobile screens are not designed to fit CIF or even QCIF resolution. It would be interesting to allow a reshape of the video, in a non-dyadic fashion.

4. Targeting fast moving region of interest tracking over time. Extracting salient points using most important wavelet coefficients are now quite known methods. By extracting salient point, one can track region of interest during time. Another way would be to manually select an object in the video, to track it following the motion.

5. Extremely fine grain SNR scalability. This scalability is naturally implemented given the multiresolution framework enabled by wavelet representation. Depending on the chosen filters, one can start from perfect reconstruction to very low quality.

6. Enabling efficient similarity search in large video databases. Different method based on wavelets can be used. Instead of searching full resolution video, one can search low quality videos (spatially, temporally, and SNR reduced), to accelerate the search operation. Then, on low quality videos, and using salient points, similarity can be found in space and time.

7. Allowing better rate distortion performances for very high resolution material. DCT-based codec are limited to 8x8 transform. For high resolution materials, uniform regions can quickly become very visible when using DCT. This can be solved using wavelets, which are not limited to 8x8 blocks. Rate distortion performances would be in this case much more optimized.

8. Multiple Description Coding which would lead to better error-resilience. By using the lifting scheme, it is an easy way to separate video data to transmit two separate bit streams. Using intelligent splitting, one can decode independently and separately the two bit streams (spatially, temporally and/or SNR reduced) or reconstruct the whole bit stream using the two representations of the video.

9. Space variant resolution adaptive decoding. When encoding the video material, it is possible to decode a high spatial resolution only in some area, keep lower resolution in the surrounding areas. Multi resolution schemes can provide easy ways to separate important information in the scene from less important information.

10. Easily provides means to optimally prioritize temporal versus spatial information for fast decoding purposes. After some basic global motion analysis from the compressed portion of the bit-stream, one can skip high frequency content (space & time) in case of high motion; in case of very slow motion, it is possible to skip high temporal frequency content, in limited bandwidth conditions. This could also be extended for locally fast moving data, clearly by changing the prioritization of bits.

11. Obtain a full compatibility with J2K and MJ2K. MJ2K is only "intra" coded video using J2K. If J2K compatibility is obtained, consequently MJ2K compatibility is also obtained. Parsing video contents when only pointing on "intra" picture is a fast and efficient way to search into video database.

12. Digital watermarking and waterscrambling. Watermarking of wavelet coefficients (i.e. inserting hidden or logo information in the content) can be realized in many different ways. Using multi resolution representation, it is interesting to insert information in low frequency subbands. Waterscrambling concerns the video data encryption. A video can be previewed on a very low resolution (spatially, temporally or SNR reduced), and transmitted for a full view on higher resolution.

Much functionality has direct and concrete applications for:

- Digital Cinema

Using the functionality of high definition storage, using wavelets can offer more than 3 levels of spatial resolutions, in order to deliver very high quality content. Also, one can benefit of better reduction of spatial correlation beyond 8x8 blocks by functionality 7, or its neighbors. A better reduction of temporal correlation across large temporal intervals (in JVT like approaches, P-B3-B2-B3-B1-B3-B2-B3-P), provided local structure is consistently estimated from frames over several frames (minimize absolute difference over several frames at the same time)

- Surveillance

Much functionality allows wavelet video codec to be used for surveillance. For instance, surveillance can has benefit from a very high number of not only dyadic decomposition, tracking motion, extremely fine grain scalability and at last, a full compatibility with (M)JPEG2000. Example of surveillance can be given with car plate tracking (may be non MCTF, capture of salient points) and recognition of car plates, and finally, video surveillance (from high definition screen to mobile screen)

- Video editing

Functionality such as non dyadic decomposition, or high number of decomposition level can be interesting for video editing and video authoring, so to accelerate treatments. More generally, wavelets have multiple kinds of filters allowing denoising, restoration, etc.

- Conversion format

Making benefit of non dyadic decomposition, an example can be to convert SD contents to HDTV contents or adapting classical QCIF or CIF to mobile screen.

- Wireless broadcasting

Wireless transmission has growing interest, especially for mobile transmission. Protection and error resilience is a major issue, can be partially solved using multiple description coding (functionality 9). Separating a video content into two independent and fully decodable bit streams is very interesting in case of error prone environments. Mobile networks are subjects to bit error and packet losses. Having two different and complementary representations of a video content is easily achievable by separating the wavelet coefficients, in a lifting scheme fashion.

- Video indexing, browsing and information retrieval

Working on a small amount of wavelet coefficients can drastically reduce the processing time. Information can be found on low frequencies (allowed by the multi resolution representation). Actually, a high number of decomposition allows working on low resolution contents, to speed up the process time. Also, better RD performance speed up the treatments. Finally, a complete compatibility with JPEG2000 and MJPEG2000 gives efficiency to browse only “intra” pictures on videos.

- Medical imaging

Some applications for medical imaging can have benefit to work on very high definition content, or localized high definition. Storage of very high definition contents is also a major issue, which can be solved using wavelets. Recalling that on high definition contents, blocks can be very disturbing in this special application. Avoiding those artifacts can be very interesting for a good analysis of the contents.

- Data encryption

Contents delivery is often a problem considering illegal copy and peer to peer systems. In that sense, copyright and protection of data can be easily solved using wavelets. Watermarking and Waterscrambling are two news fields, that have proved efficiency thank to recent techniques.

3 Performance evaluation

3.1 Quality assessment in a scalable video coding framework

Problem statement

Objective decoding quality (PSNR) values to compare different coding systems are usually calculated at a certain resolution and for each considered system with respect to a) each system related reference video sequence at the considered resolution, b) a single reference video sequence. In both cases a) and b) the comparison is unfair. This is because each system differs in the way such references are calculated (MPEG downsampling filters for JSVM3.0, 9/7 wavelet filter bank for “t+2D” WVC configuration, 3-LS wavelet filters [22] for STool configuration) and then only PSNR trends referred to a single system are meaningful but not absolute PSNR comparison among systems. In particular due to poor half-band selectivity of wavelet low pass filters, WVC system references are in general more detailed and contain more or less visible spatial aliasing. This determines lower PSNR values with respect to those measured with respect to a smoother. Therefore, PSNR differences between the three coding schemes lose significance when lower or intermediate resolutions are considered. In the following we propose a possible solution to this problem which will allow us to “re-interpret” the PSNR results obtained with method a) or b) (see par. 3.1.2).

Due to the above difficulties and in order to select best SVC schemes quality assessment has been mainly done visually by a subjective evaluation method (see par. 3.1.3).

Objective measures with averaged reference

A method to create a fair reference between two systems with their own reference video V_1 and V_2 is to create a common weighted reference $V=\alpha_1 V_1+\alpha_2 V_2$. In particular, by selecting $\alpha_1=\alpha_2=1/2$ it can be easily verified that $PSNR(V,V_1)=PSNR(V,V_2)$. This means that V_1 and V_2 are each other equally disadvantaged by the creation of V. Moreover, signal V can be reasonably used as a common reference for PSNR comparisons of decoded sequences generated by different systems. In fact, even if a rigorously fair comparison of two very different coding systems each one using its own reference signal could be seen as an extremely complex and multi-parametric problem to solve, a simple analysis uniquely based on signals spectral content give us the possibility to confirm that the signal V can be used as common reference, as stated above.

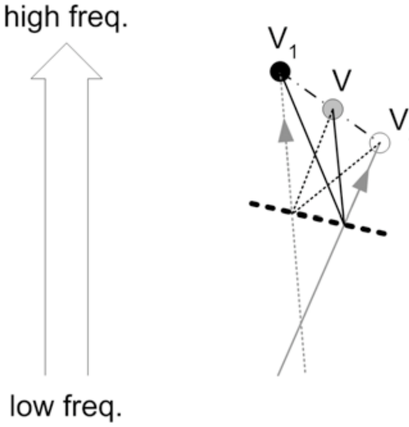


Figure 4. Common reference representation

Fig. 4 is a "projection" of the decoding process which evidences the video frame's frequency content. V_1 is the reference video sequence of WVC system, while V_2 is the one of JSVM, they are both created starting from an original video by means of LTI filters and decimation (MPEG downsampling generates smoother sequences than those generated by wavelet kernels; signal smoothing is also a strategy in AVC-H.264 based systems in order to reduce the visual impact of the artefacts related to the block based DCT and motion model). From a spectral point of view, V can be considered halfway, being it a simple average. As every transform based coding systems reconstructs first the lower spectral part of pictures, it is plausible, as shown in Fig. 4, that, at a certain bit rate (represented by the dashed bold line), the WVC reconstructed signal is nearer to the JSVM reference than to its own reference. The converse is not possible and V actually compensates for this disparity, making it possible a fair comparison on a common reference.

Visual tests method

The evaluation of coded video in absence of an unimpaired reference, demands for the usage of a particular test method, i.e. the Single Stimulus MultiMedia (SSMM) test method. The Single Stimulus MM test method is basically derived from the Single Stimulus method, as described in ITU-R rec. BT 500-11, and the Single Stimulus with two repetitions, as it was used in the MPEG-4 1995 Competition test. This method has been used also for SVC system comparisons as described in [5,6].

3.2 Recent performance results

A comparison among the decoded sequences by JSVM3.0, VidWav reference software WVC 2.0 in “t+2D” working condition and AVC base-layer (with configuration files, provided by MSRA) and Vidwav reference software WVC 2.0 in “2D+t+2D” working condition (configured as described in the m12642 document [43]) is reported. All the points have been extracted following the Palma extraction path and the bitstream size have been verified. It has to be mentioned that the results, concerning JSVM3.0 absolute performance, presented in this section has been obtained by using configuration files not fully optimized for this codec version.

The main purpose of presenting the above comparison, is to have an indication on the PSNR curves shifts, occurring when a common reference is used, as described in Subsection 3.1. All PSNR results considered in this section are reported in the excel file attached to document m12643 [51]. Slightly better PSNR results can be obtained with optimized configurations of JSVM and by enabling closed loop hierarchical B-frames prediction [61]. Such results and related visual comparisons are reported in Sec. 3.4 of this Status Report.

Lowest spatial resolution results

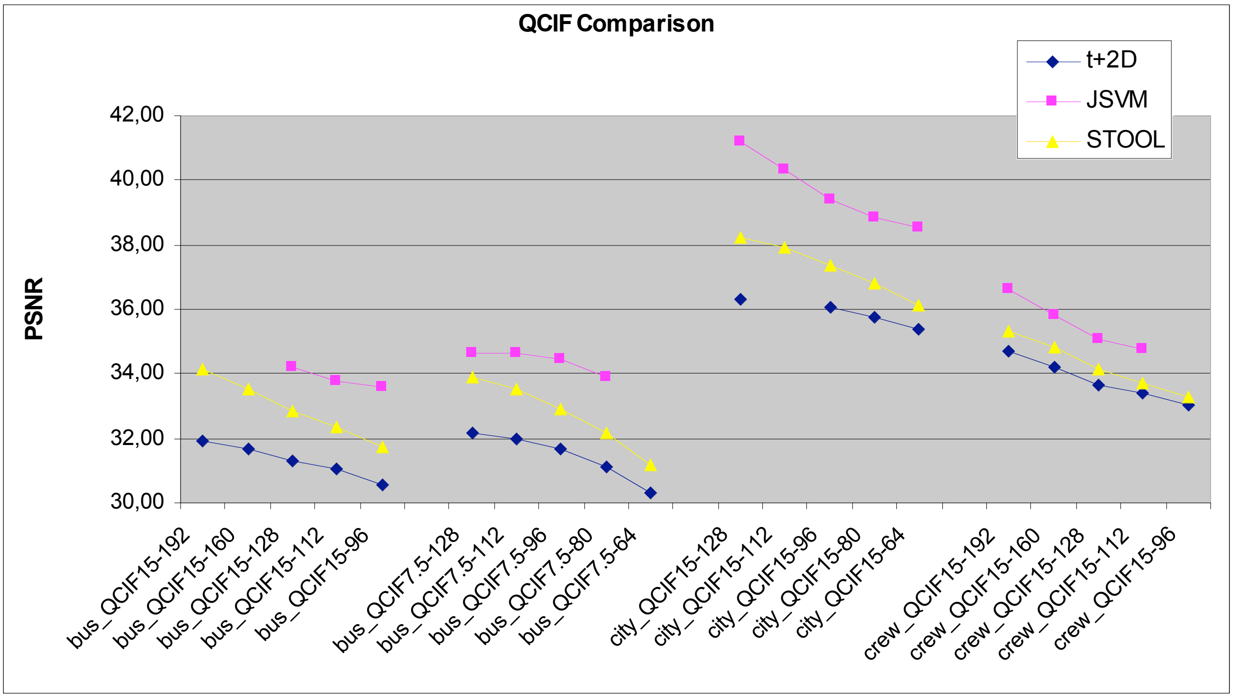
3.2.1.1 PSNR comparison with original references

Figure 5 presents 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.

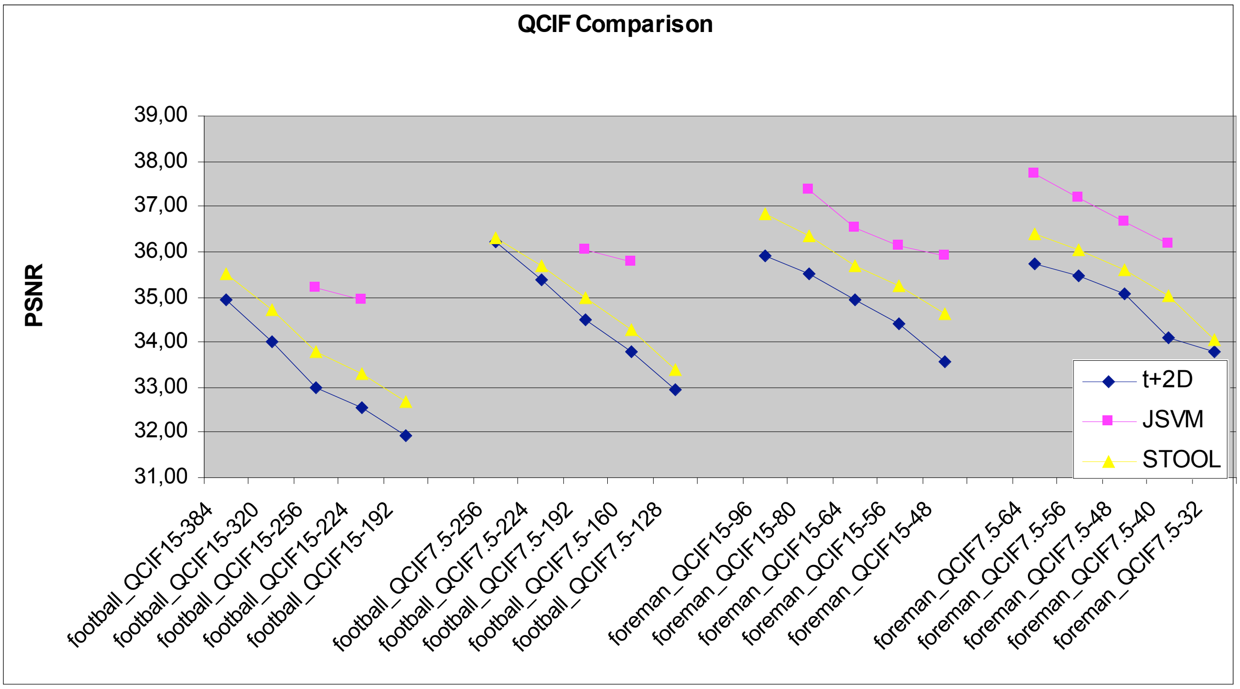
3.2.1.2 PSNR comparison with averaged references

In Figure 6 we compare the PSNR results obtained on two sequences using both system related references and a common reference for JSVM3 and STool.

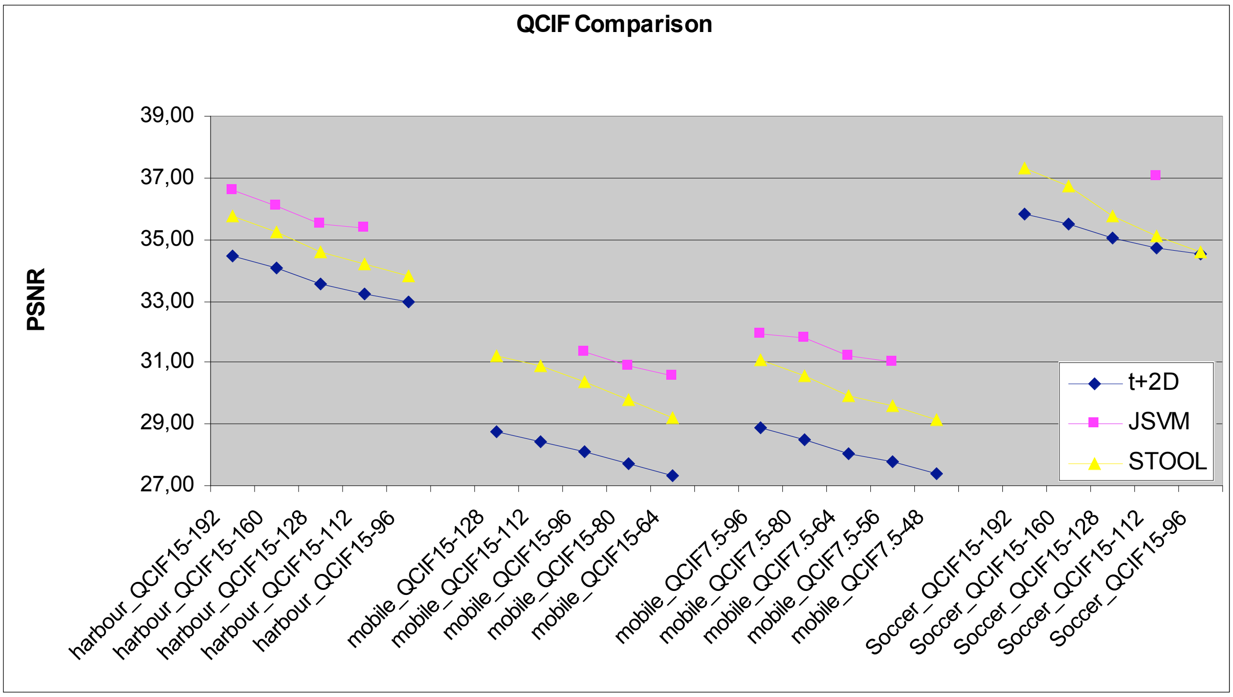
Results in Fig. 6 indicate that using a common reference, “2D+t+2D” configuration PSNR results are very close (and sometimes outperforms) those of JSVM3.


QCIF Comparison
t+2D
JSVM
STOOL
PSNR
42,00
40,00
38,00
36,00
34,00
32,00
30,00
bus_QCIF15-192
bus_QCIF15-160
bus_QCIF15-128
bus_QCIF15-112
bus_QCIF15-96
bus_QCIF7.5-128
bus_QCIF7.5-112
bus_QCIF7.5-96
bus_QCIF7.5-80
bus_QCIF7.5-64
city_QCIF15-128
city_QCIF15-112
city_QCIF15-96
city_QCIF15-80
city_QCIF15-64
crew_QCIF15-192
crew_QCIF15-160
crew_QCIF15-128
crew_QCIF15-112
crew_QCIF15-96

(a)


QCIF Comparison
PSNR
39,00
38,00
37,00
36,00
35,00
34,00
33,00
32,00
31,00
t+2D
JSVM
STOOL
football_QCIF15-384
football_QCIF15-320
football_QCIF15-256
football_QCIF15-224
football_QCIF15-192
football_QCIF7.5-256
football_QCIF7.5-224
football_QCIF7.5-192
football_QCIF7.5-160
football_QCIF7.5-128
foreman_QCIF15-96
foreman_QCIF15-80
foreman_QCIF15-64
foreman_QCIF15-56
foreman_QCIF15-48
foreman_QCIF7.5-64
foreman_QCIF7.5-56
foreman_QCIF7.5-48
foreman_QCIF7.5-40
foreman_QCIF7.5-32

(b)



(c)

Figure 5. (a-c) PSNR comparison at QCIF resolution

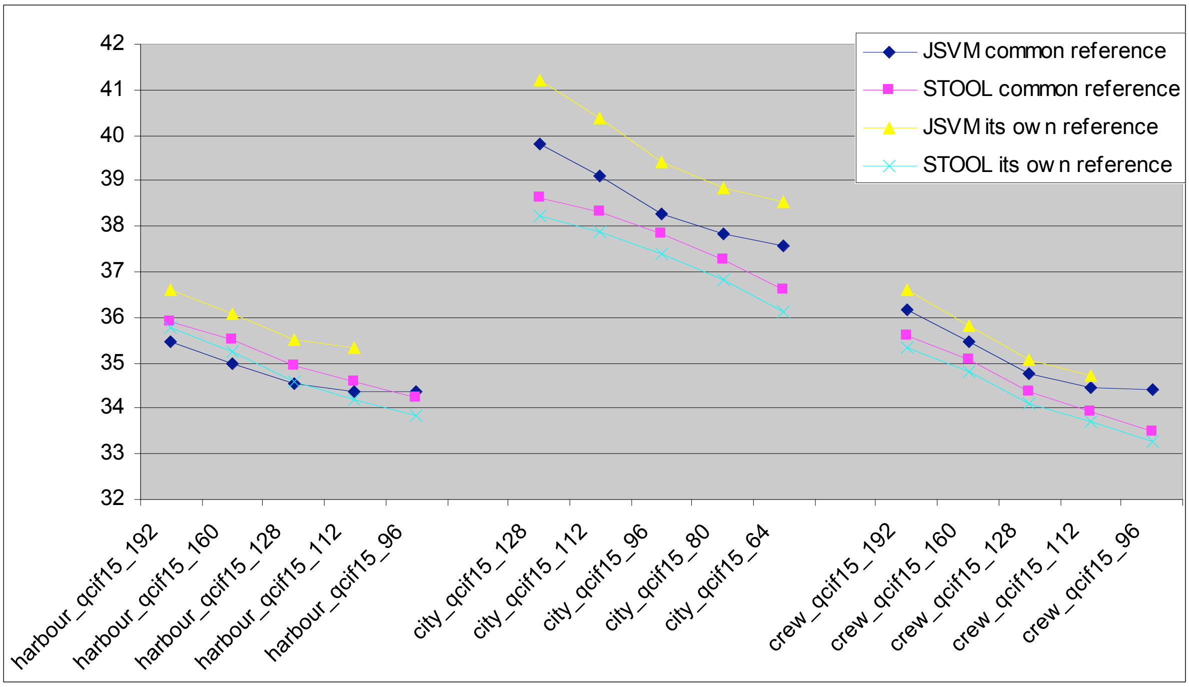


Figure 6. PSNR at QCIF resolution: common reference usage

3.2.1.3 Visual comparison at QCIF resolution

We show a visual comparison among some sample frames. In Fig.7 some 15fps 128kpbs decoded frames of the CREW sequence are displayed, and in Fig.8 a representative frame of the 7.5fps decoded FOOTBALL sequence is shown for 2 different bit-rates.

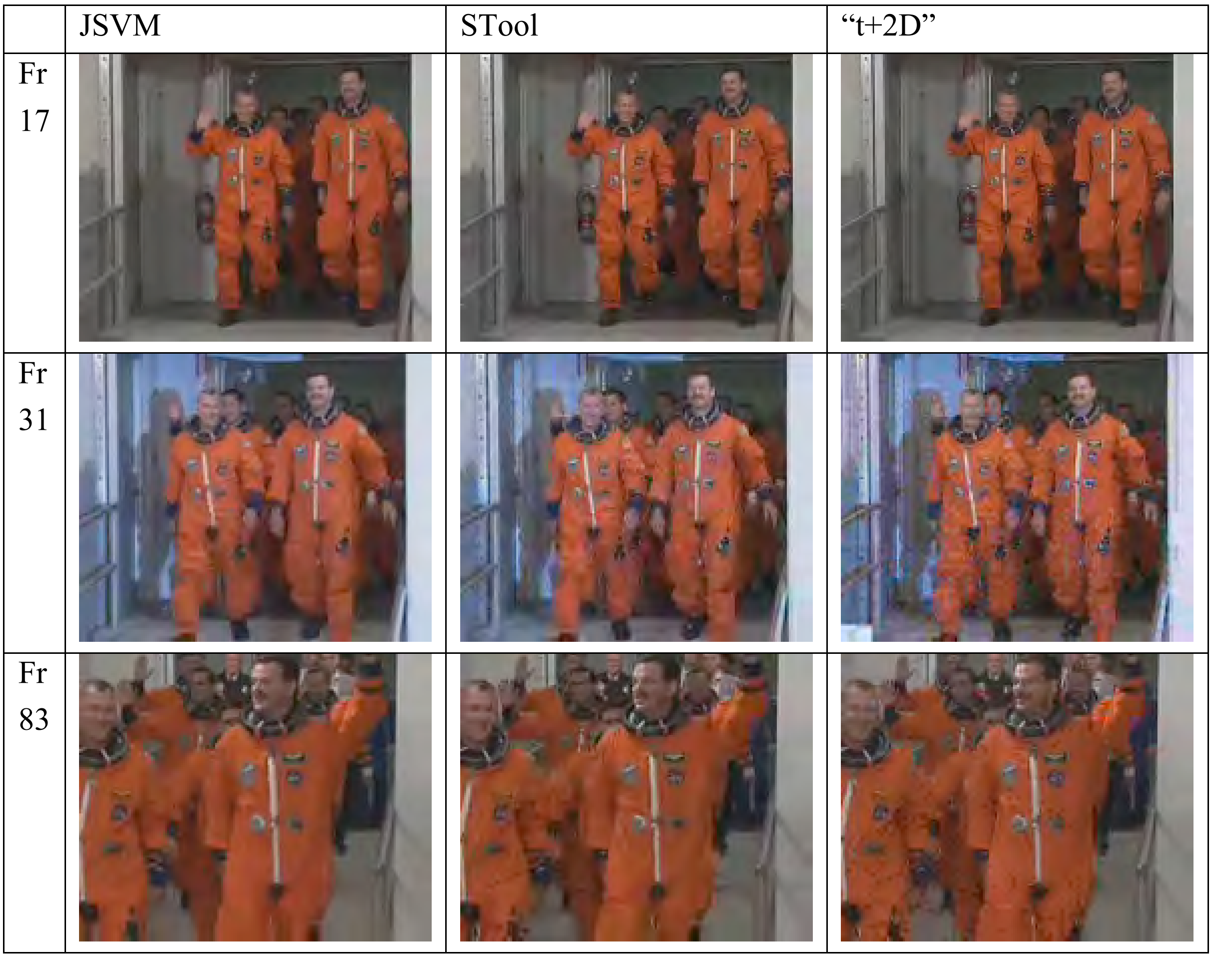


Figure 7. visual comparison on CREW QCIF 15fps 128kbps

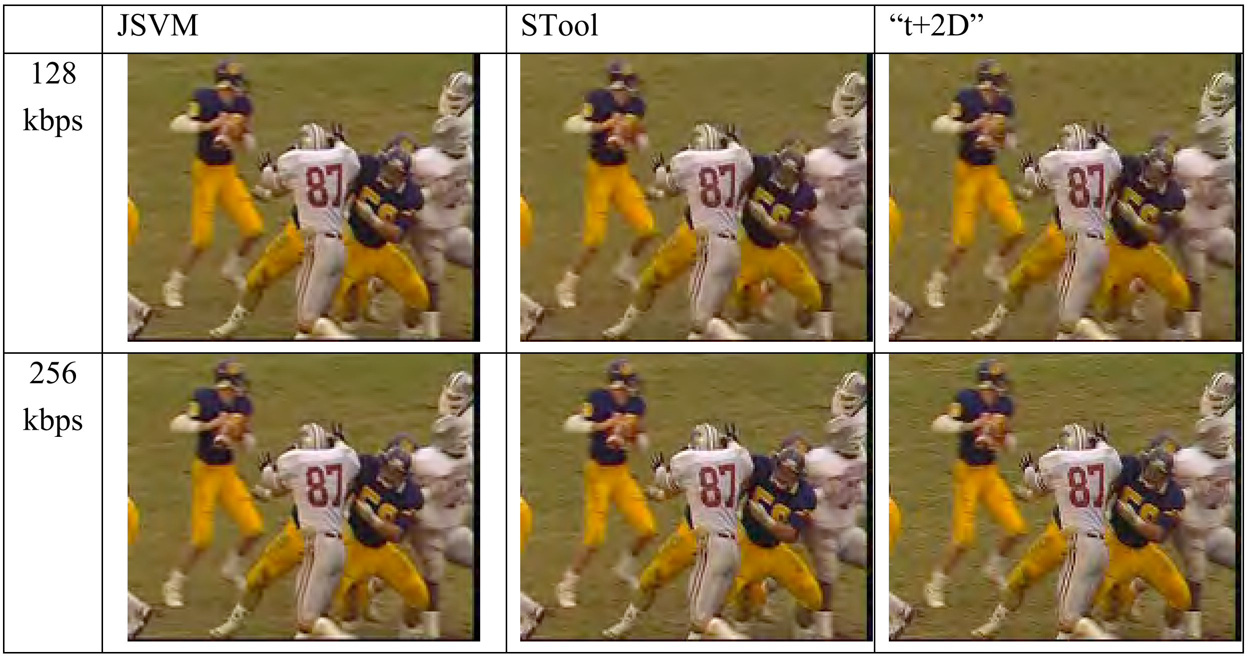


Figure 8. visual comparison on FOOTBALL QCIF 7.5 (frame 17)

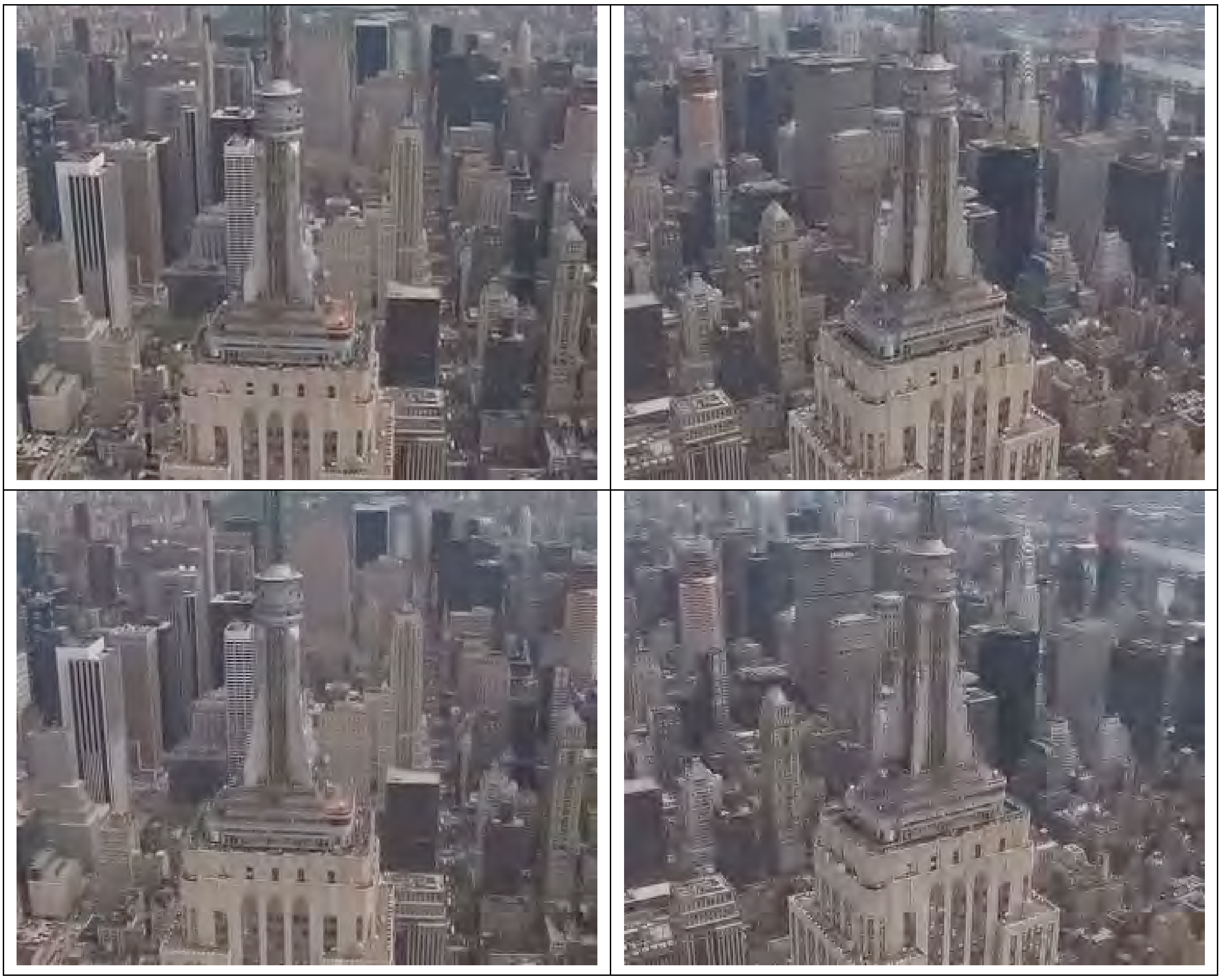
Intermediate spatial resolution

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 9 shows a visual comparison on the CITY sequence. All three sequences are visually close. In Fig. 10 we show some PSNR results with or without using a common reference. Similar remarks on the common reference usage , previously made for QCIF resolution, apply also in this case.

Highest spatial resolutions

3.2.1.4 CIF originals

We propose a visual comparison for the sequences FOOTBALL (Fig.11) and MOBILE (Fig.12). 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.



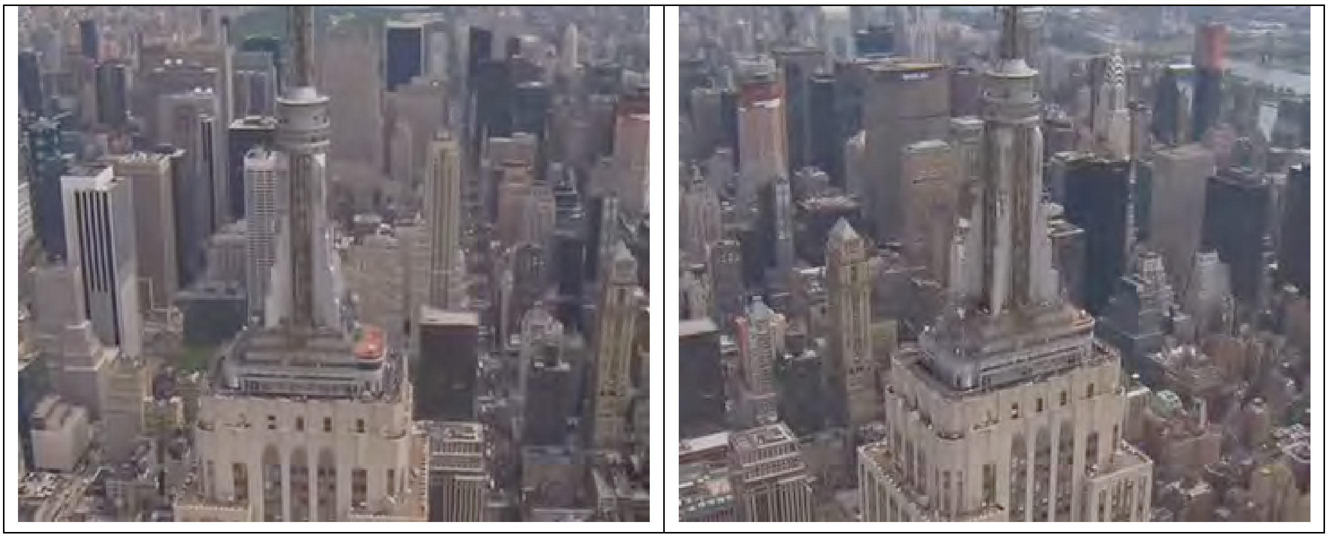


Figure 9. City_CIF15-192: (top) STool (192kbps) mean PSNR 34.05dB, (mid) "t+2D" ref sw (195kbps) mean PSNR 33.43dB, (bottom) JSVM3 (192kbps) mean PSNR 36.76dB

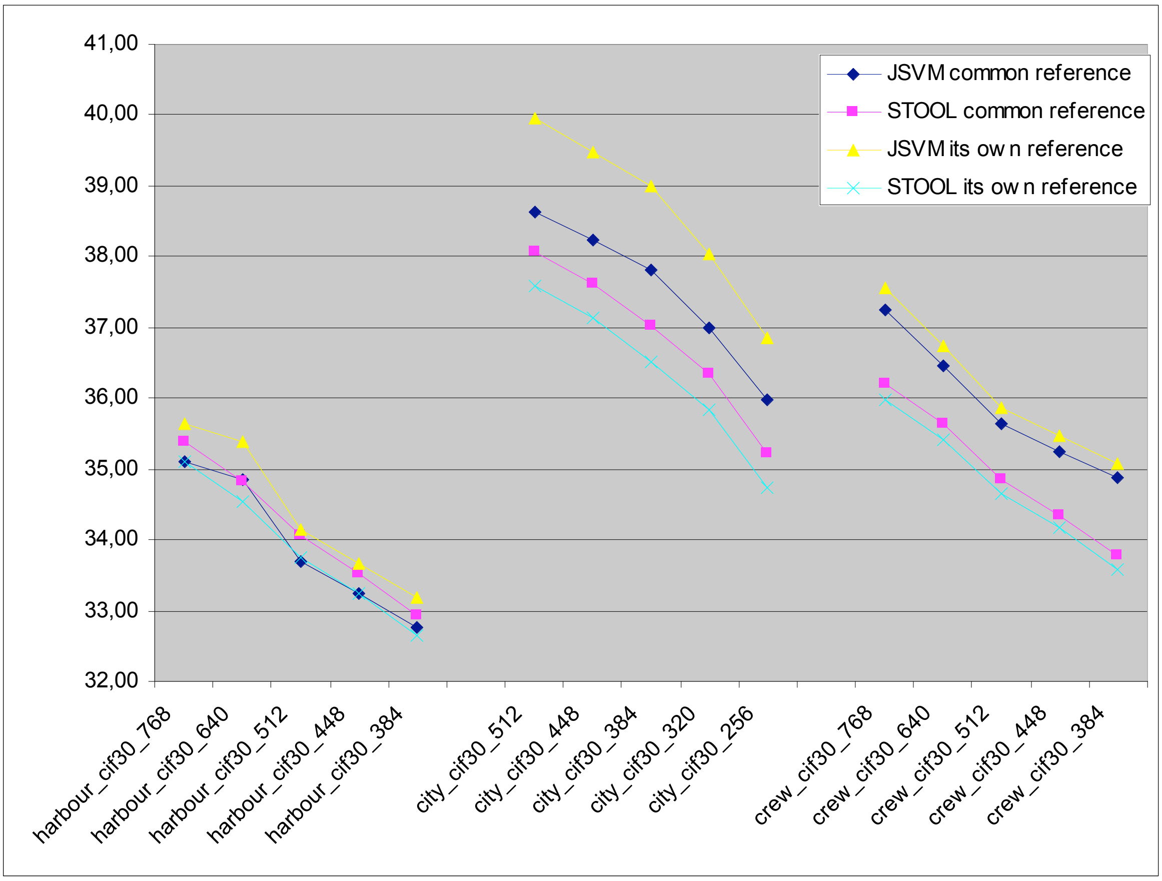


Figure 10. some PSNR results at CIF resolution (with and without a common reference)

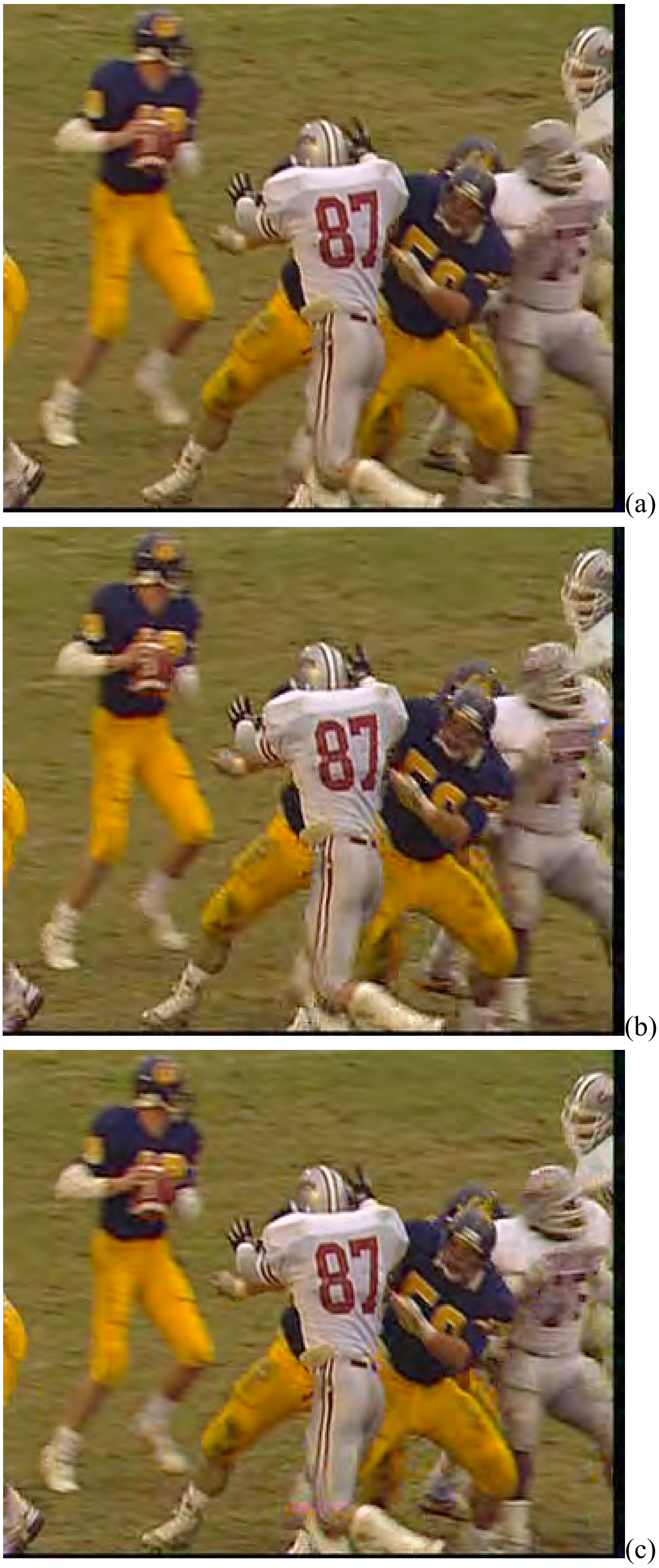


Figure 11: Football_CIF30-1024: (a) JSVM mean PSNR 35.95dB (b) STool 34.62dB (c) “t+2D” (1.128Mbps) 36.0db

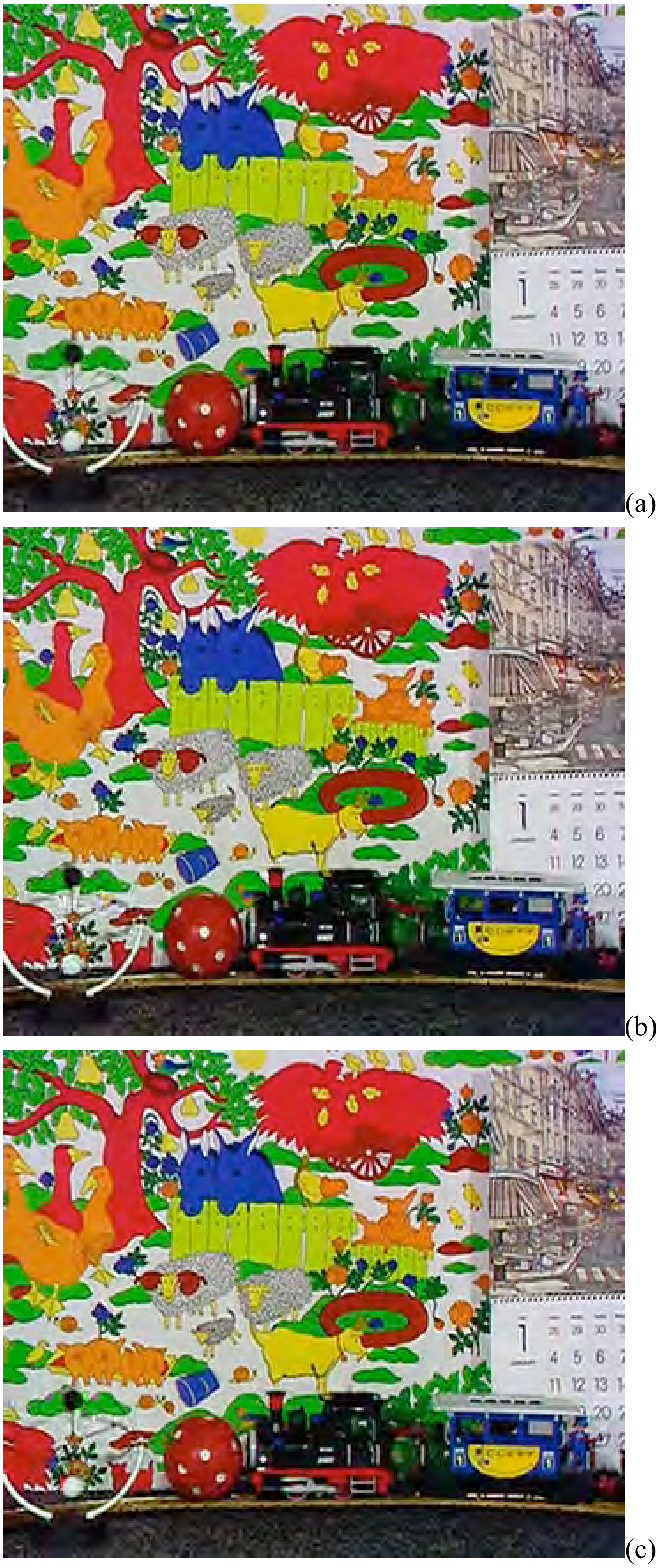
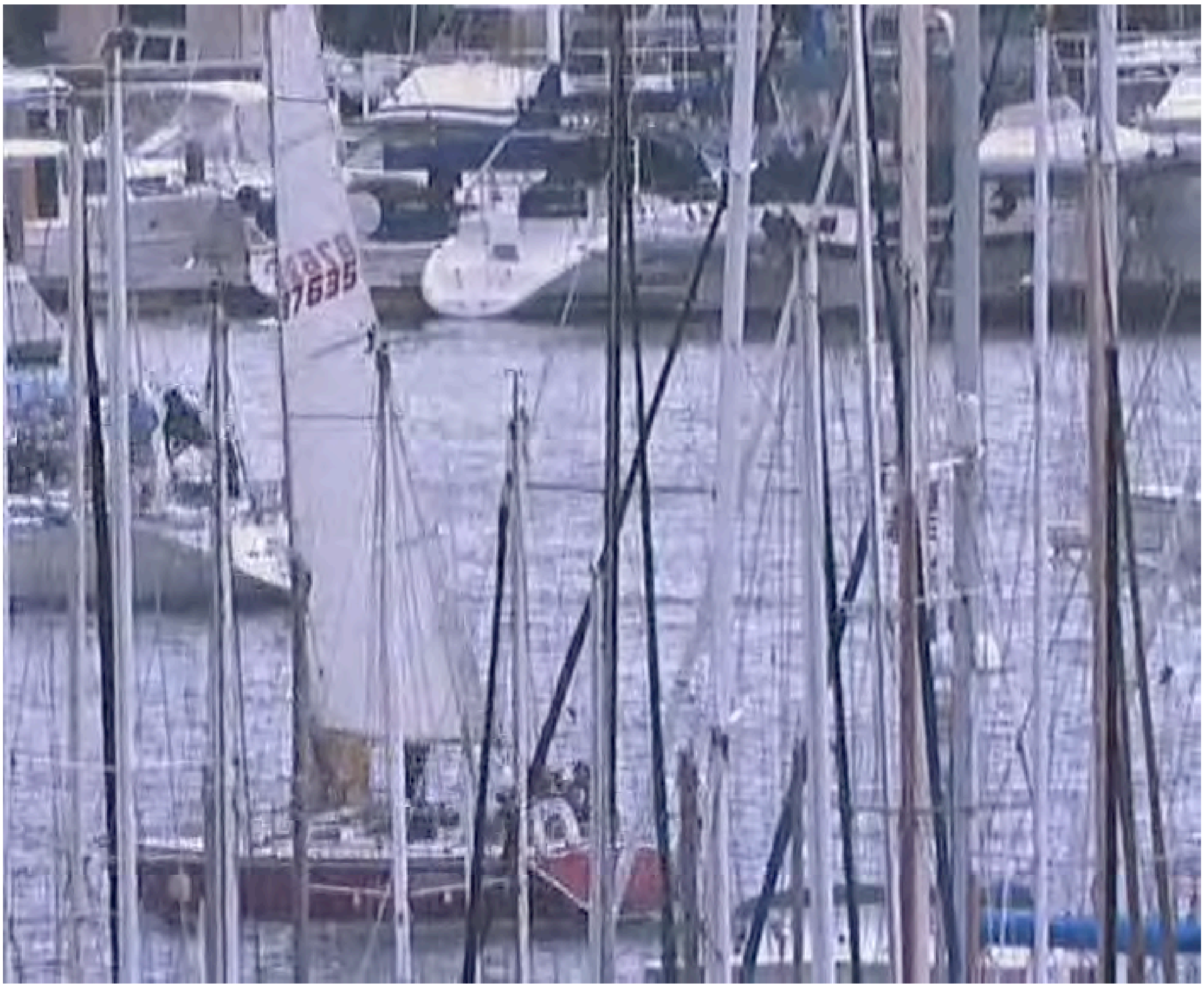


Figure 12: Mobile_CIF30-384: (a) JSVM (384kbps) mean PSNR 31.04dB (b) STool (384kbps) 29.63dB (c) “t+2D” (**429**kbps) 31.26dB

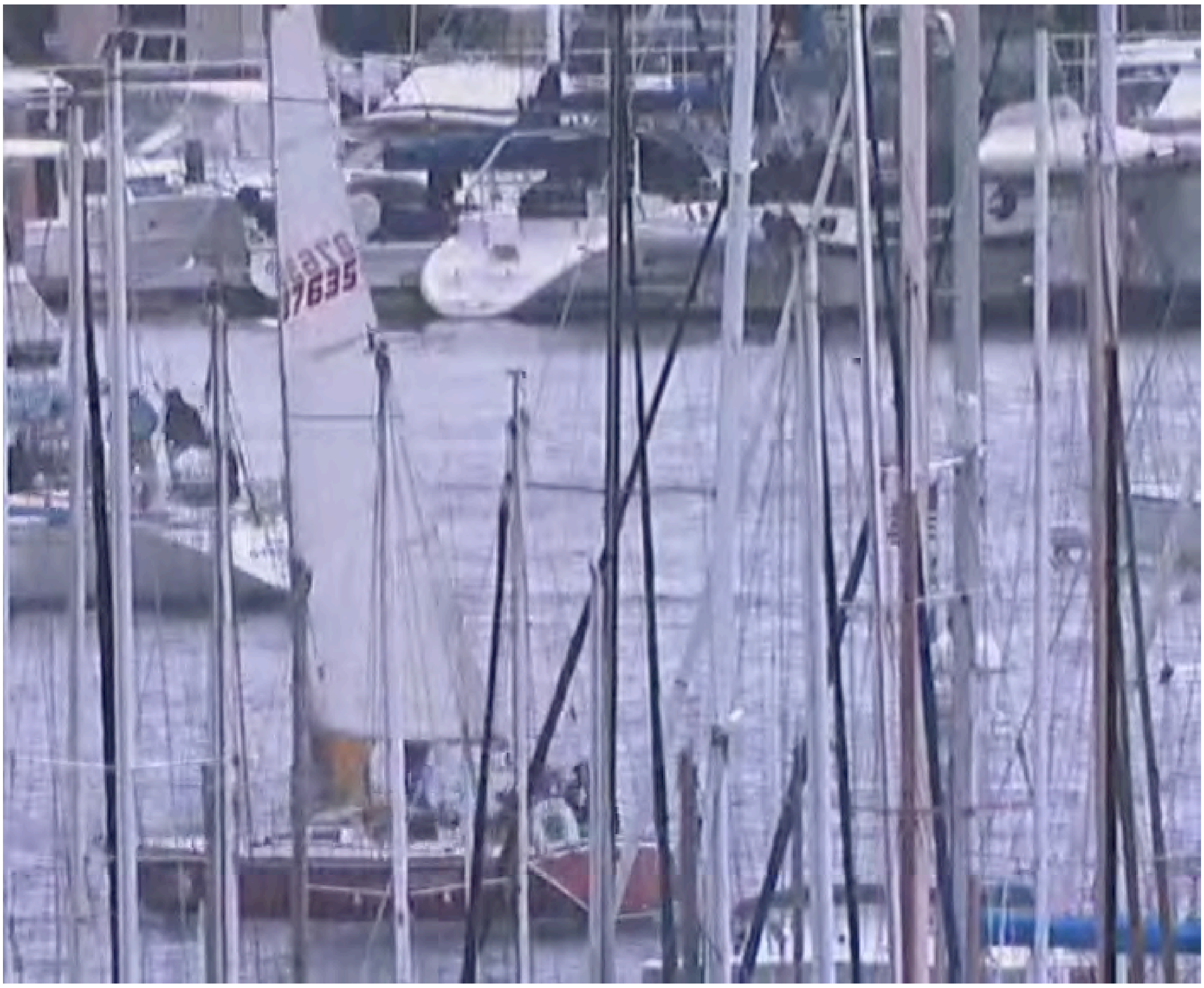
3.2.1.5 4CIF originals

In this case (Fig. 13), 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 13: HARBOUR 4CIF 30fps 1024kbps: (a) STool PSNR 33.02dB, (b) "t+2D" PSNR 34.45dB, (c) JSVM3 PSNR 32.58dB

Extended SNR scalability results

A series of tests has been made in [57] [58] in order to compare JSVM1 with respect to the MSRA codec which is the former version of the VidWav reference software WVC 1.0 and from which VidWav reference software WVC 2.0 mainly derives.

Conclusions drawn in [57] evidenced that the two compared systems had comparable coding performance and that the supported scalability range of MSRA codec was much larger than the JSVM1.0 codec and also presented more elegant quality degradation when bit rate is reduced.

In Fig.14 we report the 4CIF PSNR results for the sequence "Crew".

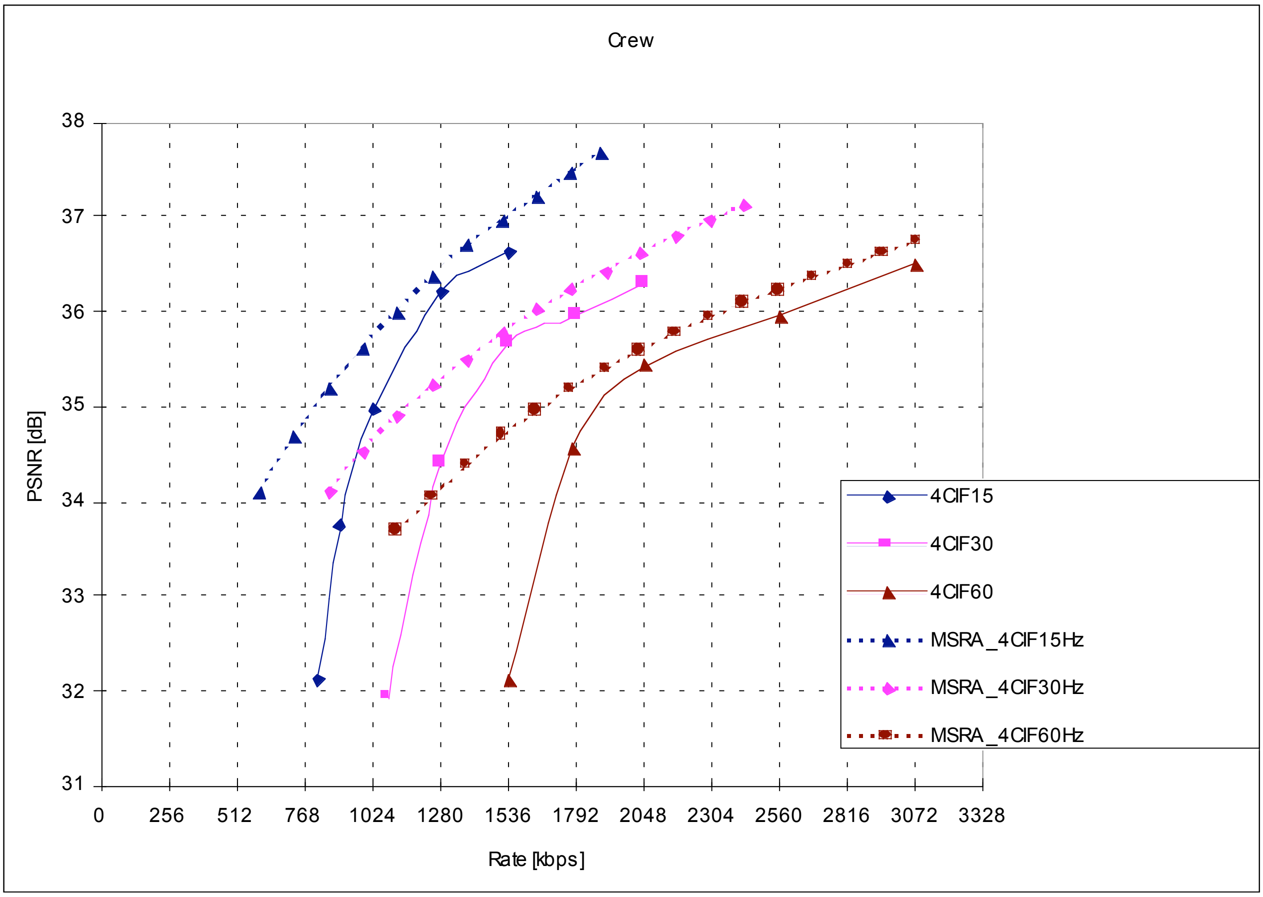


Figure 14: PSNR comparison at maximum 4CIF resolution. Solid lines refers to JSVM 1.0 while dashed lines to MSRA software.

Here we also report the main observations the authors of [57] derived:

> From the PSNR results, we can see that, the performance of wavelet-based codec is comparable to the DCT-based H.264 scalable extension. For many CIF sequences, the PSNR performance can be better than JSVM1.0. For HARBOUR sequence, the gain is very significant. There are also some sequences with which JSVM1.0 is better than MSRA codec, especially for CITY. For low bit rate video, the JSVM also often demonstrates its advantage. As a conclusion, we do not think that the wavelet-based codec always has a performance inferior to that of DCT-based JSVM.
>
> As another important observation, we find that the SNR scalablity range for narrow at each spatial layer under current tesing condition. For some sequences, the five tesing points at QCIF7.5 and QCIF15 resolution merge to only three points. This is due to the redundancy and near-simucast property among the various spatial layers. If the SNR scalability range is enlarged for QCIF, the CIF bitstream can not obtain good quality at low bit rate.

> The above phenomenon is more obvious in 4CIF sequence where three resolutions are involved. We can see the PSNR performance of 4CIF15Hz, 4CIF30Hz and 4CIF60Hz drop very quickly when bit rate decreases to its first two testing points. This is due to a large CIF bitsteam in 4CIF bitstream and the two bitstream is not embedded well.
>
> However, for wavelet-based codec adopting non-redundant representation, the SNR scalability range for each spatial layer can be much larger and has more elegant quality degradation when bit rate is reduced.

Further experiments on extension of the SNR scalability bit-rate range was made in an exploratory experiment and reported in [58] where there was an evidence that for the JSVM1, which uses a layered technique to support the spatial scalability, the performance of a resolution is highly dependent with the supported SNR bit-rate range of all lower resolutions. On the base of a preliminary set of experiments on the CIF scenario aiming to test JSVM 1.0 software for different low-resolution highest bit-rates on a determined extraction path, the following observations was made:

> When we want to support a wide SNR scalability range at all resolutions, that is, when the “Maximum FGS Bit-Rate” at lower resolutions are increased, the JSVM1.0 performance at higher resolution will loss. (Of course, it is not necessary for the “Maximum FGS Bit-Rate” at lower resolutions to achieve a very high bit-rate. This depends on applications and users. Maybe a quality of 34dB~38dB is good enough.)
>
> There are two kind of performance loss phenomenon for high resolution.
>
> 1. At low bit rate of high resolution, the extraction bit rate is not high enough so that the referenced FGS layers at lower resolution layers are partly or completely unavailable at decoder. This is a drifting problem.
> 2. At high bit rate of high resolution, the extraction bit rate is high enough that the referenced FGS layers at lower resolution layers are completely available at decoder. There is no drifting problem. But the performance will still lose when the “Maximum FGS Bit-Rate” of low resolution is increased. This is due to the larger sub-bitstream for low resolution video which can not be fully exploited by high resolution bitstream. This is a penalty of simulcast.

A sample of that experiments is presented here in Figure 15.

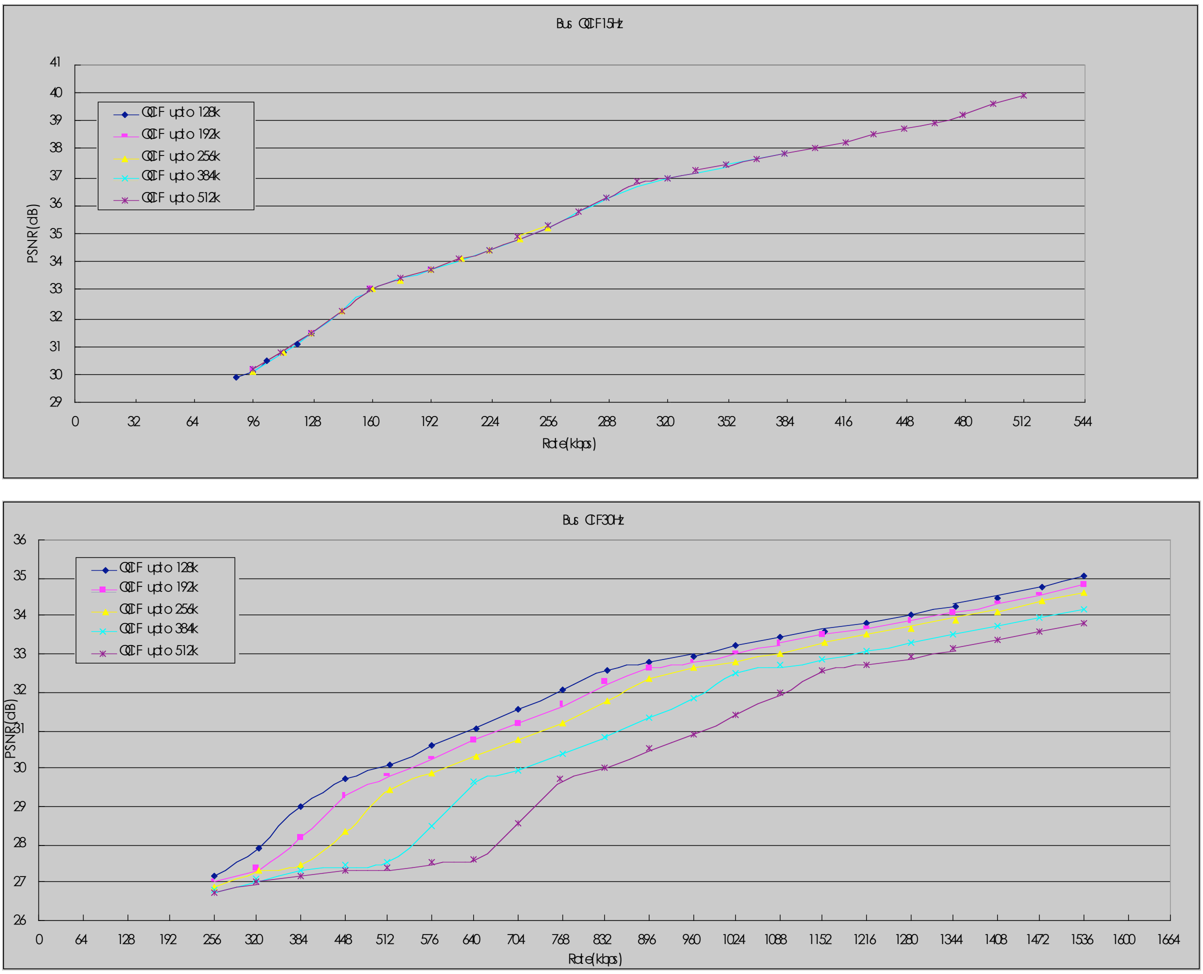


Figure 15: PSNR results at CIF resolution on the "Bus" sequence: tests on several higher bit-rate configuration for lower (QCIF) resolution. Default config: Maximum FGS Bit-Rate at QCIF15Hz = 128k. 192k, 256k, 384k, 512k have also been tested.

Moreover, in Figure 16 we report form [58] a JSVM 1.0 vs MSRA comparison graph were JSVM 1.0 is run with default and extended maximum FGS Bit-Rate at QCIF15Hz.

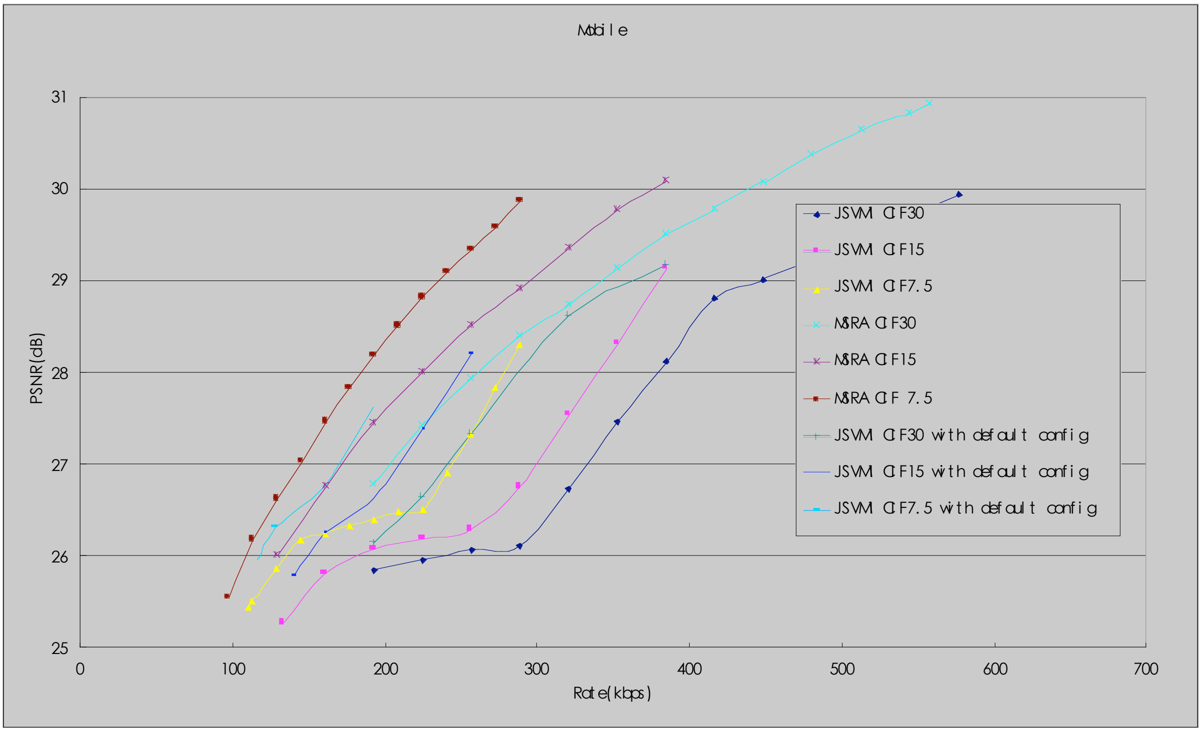


Figure 16: PSNR comparison at CIF resolution between MSRA and JSVM software on the "Mobile" sequence.

3.3 STool improvements

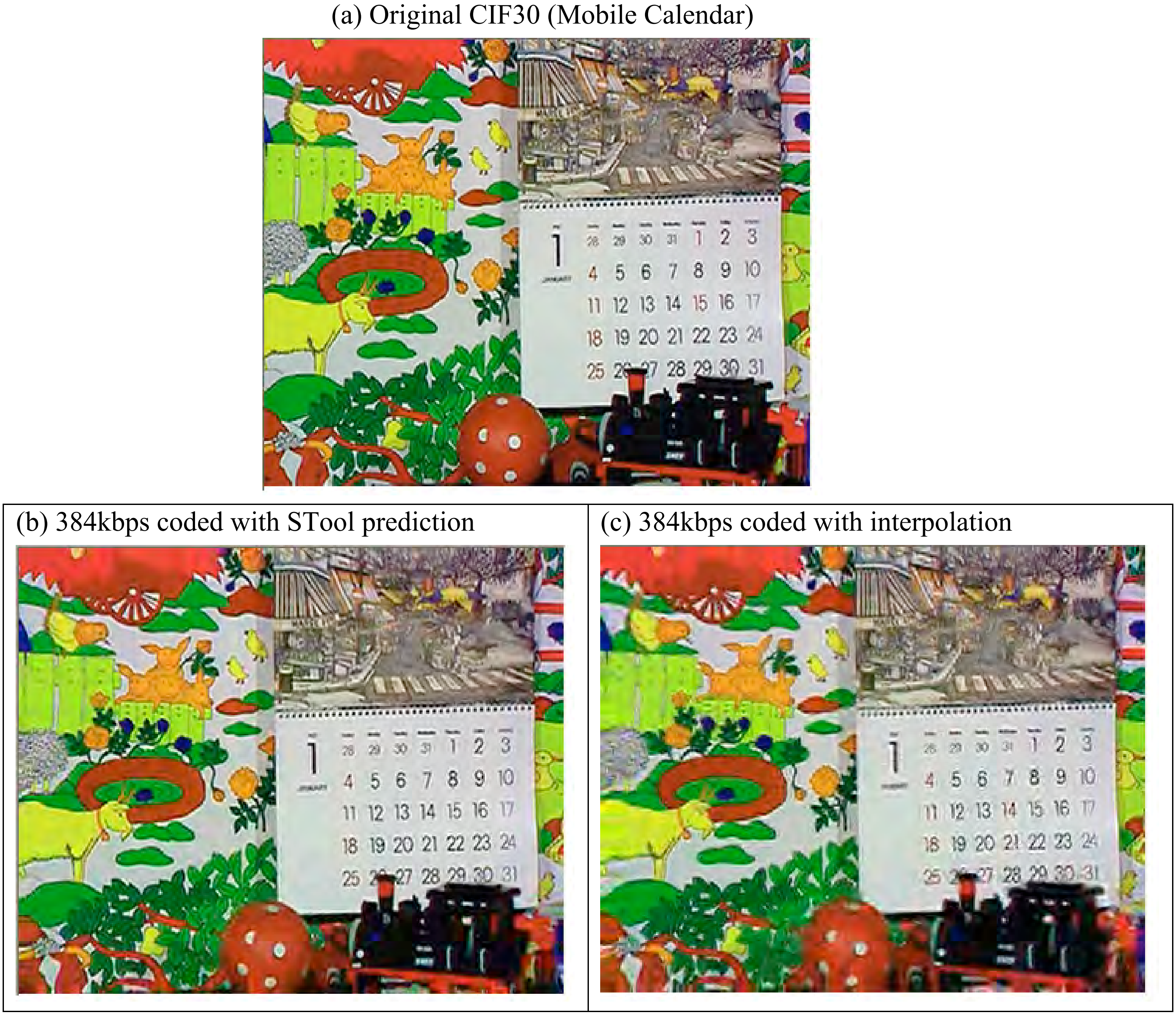
Improvements with respect to the pyramidal 2D+t+2D scheme

Table 1 reports the average luminance PSNR for the interpolation based pyramidal 2D+t+2D scheme of Figure 3 in comparison with the scheme presented in Figure 2. *Mobile Calendar* CIF sequences at 30fps are coded at 256 and 384kbps and predicted from a QCIF video coded at 128kbps (all headers and coded motion vectors included). We also compare different configurations of the STool architecture in order to highlight its versatility: 1) STool prediction made only from the lowest temporal subband of the QCIF video (in this case, which results to be the best case, only the 79kbps of the lowest temporal subband, without motion vectors, are extracted from the 128kbps coded QCIF, then 256-79=177kbps or 384-79=305kbps can be used for CIF resolution data); 2) like 1) but including all the QCIF sequence to enable multiple adaptations, i.e. extraction of a maximum quality QCIF 30fps from each coded CIF video.

Table 1. PSNR comparison among different kind of inter-scale predictions

Sequence	Format	Bitrate (kbps)	PSNR_Y pyramidal	PSNR_Y STool (mult. adapt. disabled)	PSNR_Y STool (mult. adapt. enabled)
Mobile	CIF 30fps	256	23.85	27.62	26.51
		384	25.14	29.37	28.81

Figure 17 shows an example of visual results at 384 Kbps. The STool with multiple adaptation disabled case is compared against the interpolation based ISP (also without multiple adaptation). The latter scheme generates an overall more blurred image, and the visual quality gap with respect to our system is clearly visible.



(a) Original CIF30 (Mobile Calendar)

(b) 384kbps coded with STool prediction

(c) 384kbps coded with interpolation

Figure 17. Visual comparison at 384kbps on Mobile Calendar CIF 30fps: (a) original frame CIF30 (Mobile Calendar), (b) coded at 384kbps with the STool scheme of Figure 2, (c) coded at 384kbps with the interpolation pyramidal scheme of Figure 3.

Improvements with respect to the 70th meeting, Palma (10/2004)

One year improvement of the STool and of the JSVM schemes on the lower resolution. We compare today results (current document and [52] respectively) with the results presented at the MPEG Palma Meeting in Oct.2004 ([41] System 1 based on the MSRA SVC software and HHI SVC proposal and software respectively). In Tab. 2 we calculated, for each test sequence, a PSNR measure which is the average PSNR on the whole set of QCIF multiple extracted Palma points allowable for each sequence. PSNR are calculated with respect to each system reference i.e. 3-LS filtered and MPEG downsampling filtered sequences respectively. The PSNR improvements (difference) are free from the bias related to the different reference sequence.

Sequence	PSNR Palma Stool	PSNR Palma JSVM	PSNR Nice Stool	PSNR Nice JSVM	Difference Stool	Difference JSVM
Bus	31,49	33,96	32,34	34,02	0,85	0,06
Foreman	33,46	36,52	35,17	36,64	1,71	0,12
Football	32,23	35,91	33,94	36,04	1,71	0,13
Mobile	27,45	30,83	29,77	30,89	2,32	0,06
Harbour	34,69	36,06	34,73	36,06	0,04	0
City	37,07	38,92	37,23	39,73	0,16	0,81
Soccer	35,66	36,71	35,89	37,02	0,23	0,31
Crew	34,09	35,86	34,24	35,84	0,15	-0,02

Table 2: PSNR improvements on the QCIF resolution

However, the PSNR gains shown in the above table are not all reflected by corresponding improvements in visual quality.

Indeed, visual tests performed at the 75th MPEG meeting in Bangkok indicated that a significant improvement in performance could only be obtained for the crew sequence at the 4CIF resolution.

3.4 Latest Performance Results: Montreux comparison

Further experiments have been conducted following the 75th MPEG meeting in Bangkok, under testing conditions described in document W7823 [59]. For COMBINED scalability, all tested sequences and extraction path have been left unchanged with respect to the previous experiments. The comparison has been performed using JSVM 4.0, the WCS 2.0 in "2D+t+2D+ working condition, and the aceSVC wavelet codec[2].

Input documents [61,62,63,64] report the complete set of results for the 3 considered codecs with respect to testing conditions defined in W7823 [59].

Annex 1, 2 to this Status Report show the Y component only R-D curves at 4CIF spatial resolution for the 3 codecs under comparison for the *combined* and *spatial* scalability conditions. Annex 3 show the Y component only R-D curves at all spatial resolution for the 3 codecs under comparison for the *SNR* scalability conditions.

On PSNR grounds, comparisons among the decoded sequences by JSVM3.0, JSVM 4.0 and the WT codecs indicate, on average:

- slightly superior performance of WT based codecs under *SNR* scalability conditions
- slightly superior performance of JSVM 4.0 codec under *spatial* and *combined* scalability conditions

[2] The aceSVC codec has been presented to the Vidwav group providing new functionalities for the exploration activity [60]. The aceSVC software consists of three modules: encoder, extractor and decoder. Its design is based on the wavelet transform performed in temporal and spatial domains. In temporal domain the MCTF with adaptive choice of wavelet filters is used. In spatial domain different 2D wavelet transforms can be applied, including motion adaptive spatial transform [54]. The software supports temporal, spatial and fine-grane quality scalabilities, based on a generalised spatio-temporal decomposition architecture. The aceSVC software performance has been verified by crosschecking the proposed results in terms of extracted bitstream length and PSNR values of the decoded sequences.

On the basis of the conducted visual tests for SNR and combined scalability at the Montreux meeting, performed with the help 12 test expert viewers, at +/- _ sigma with 95% confidence intervals, comparisons between VidWav reference system and JSVM 4.0 appear on average superior for JSVM 4.0, with marginal gains in SNR conditions, and superior gains in *combined* scalability settings. A full description of such test can be found in Annex 4 of the current document.

4 Decoder-side Reduction of Artefacts

A video that has been coded and decoded using motion-compensated 3D-Wavelets, usually suffers from three different types of artefacts. Firstly, when small coefficients in higher-frequency sub-bands are quantized to zero, this can result in a blurred impression due to the loss of high-frequency content. This blurring can only be minimized by investing more bits in these coefficients – if such bits are available. Secondly, block-based motion-compensation (MC) often results in a blocky prediction at diverging motion, and the quantized reconstruction may contain visible blocking as well. Adaptive filtering over block-boundaries, overlapping MC, etc., are tools that have improved decoding results for this artefact. Thirdly, spatial ringing is introduced through quantization of the wavelet coefficients. These coefficients represent the amplitudes for oscillating basis-functions that the reconstruction is built from. Consequently, additive noise in the coefficients affects these oscillating basis functions as well. Little activity had been devoted to reduce this type of artefact although it can have strong impact on the overall visual impression of the decoded sequence. The table below summarizes the above paragraph.

Artefact	Artefact Description	Tools
Blurring	Loss of high-frequency coefficients	Rate/Distortion-Optimization, Rate-Allocation
Blocking	Block-MC, block-wise mode decision	De-Blocking Filter, OBMC, Transition-Filters
Wavelet-Ringing	Quantized coefficients for basis-functions	**De-Ringing Filter**

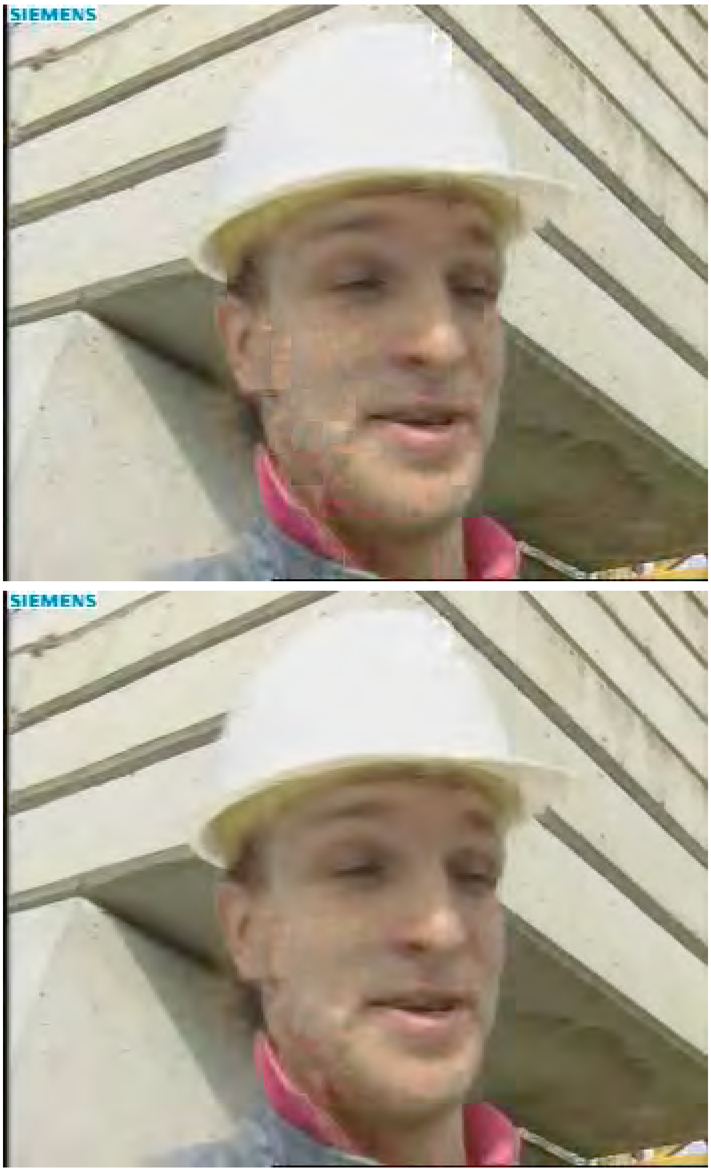
It is important to note that all three types of artefacts are a result of quantization. Without quantization, none of the artefacts is observed. The magnitudes of the artefacts are related to the quantization step-size.

In the open-loop structure that the VidWav concept represents, artefact reducing decoder-side filtering can be viewed in two ways: either as an additional, optional filtering tool or as an integral part of the reconstruction filtering. Optional filtering has generally not been part of a standard specification. Also, the open-loop structure of MCTF allows for diverging reconstruction filter implementations. However, achievable quality may only be judged during the evaluation of the codec design by including reconstruction filter tools. More importantly, reconstruction quality in an application may only be guaranteed by including a specified filtering.

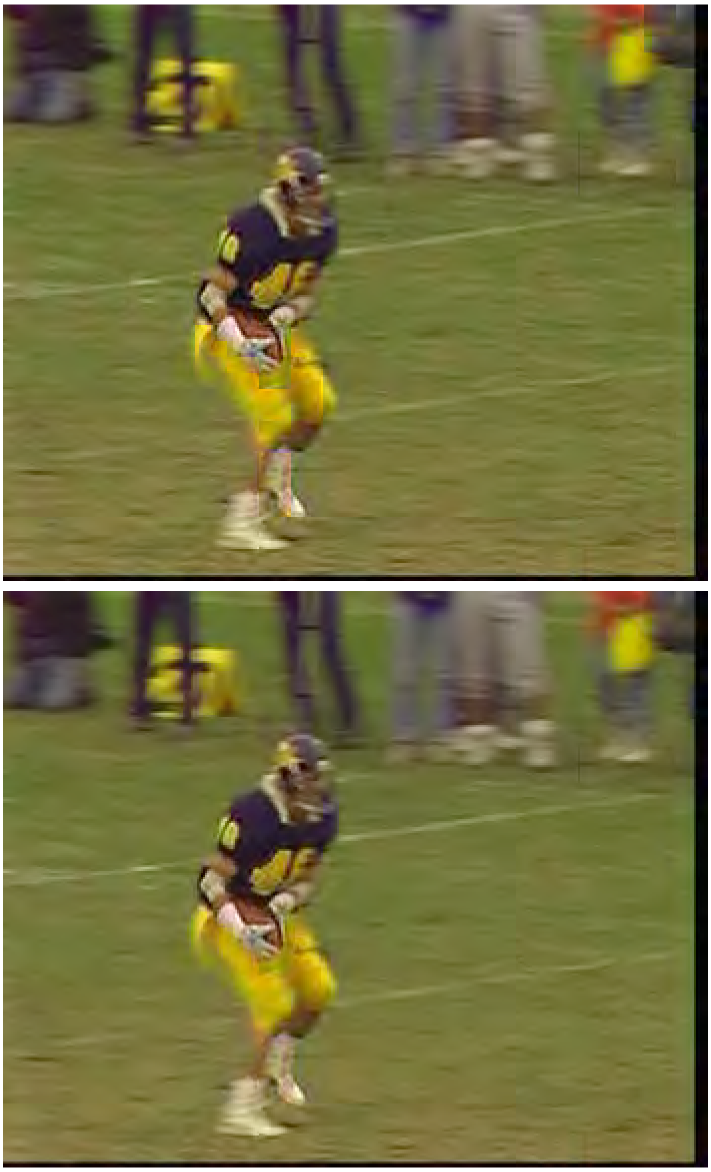
4.1 De-Ringing Filter results

A technical description of the de-ringing filter adopted for VidWav can be found in [53]. The quantization-adaptive artefact-removing filter presented in [53] has a subjectively very pleasing effect on the reconstructed video (see Fig. 18(a)-(c)). Its de-ringing as well as de-blocking properties add together in their beneficial impact. The most important property of the filter is that all decoded structures are preserved. Its smoothing effect is limited to the artificial structures which are a manifestation of quantization noise. When trained for PSNR-optimum performance, gains of more than 0.4 dB are typically observable. For visually best performance, PSNR gains are typically smaller but always existent. Some qualitative visual examples are given below. Respective upper images are without artefact-removing filtering, lower images have been filtered. Both respective results have been decoded from the same bit-stream.

(a) FOREMAN CIF 15Hz 96kbit/s



(b)FOOTBALL CIF 15Hz 384kbit/s



(c)MOBILE CIF 15Hz 128kbit/s

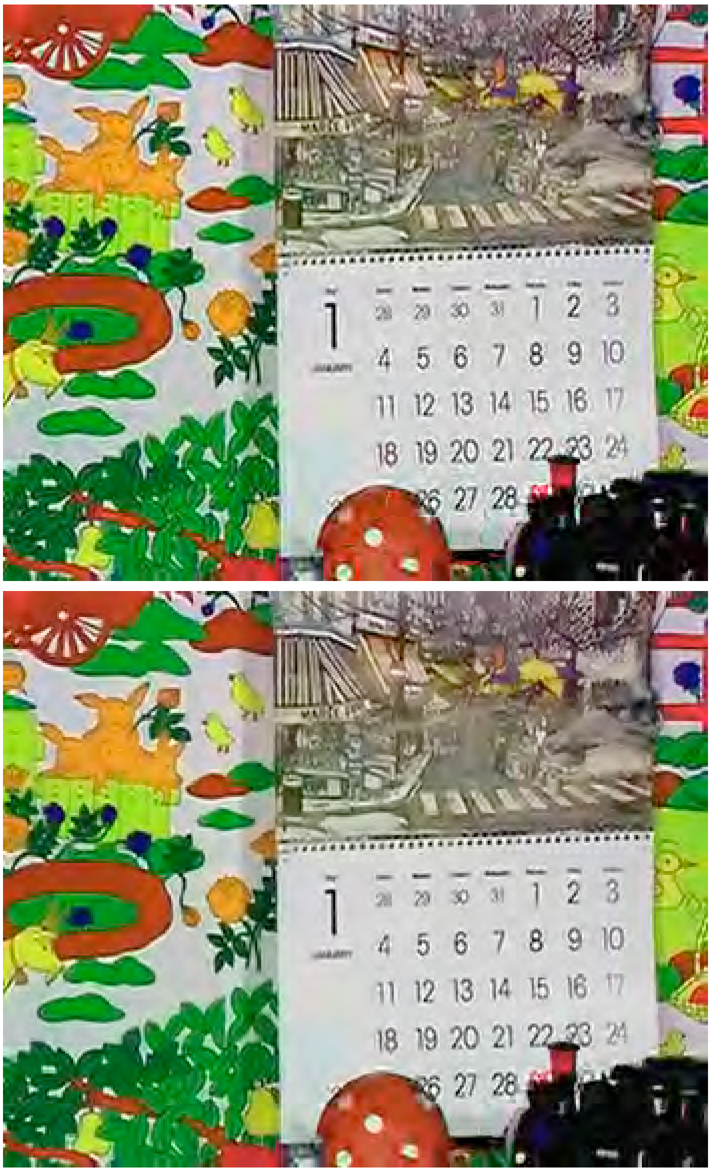


Figure 18 (a)-(c) Visual results with (bottom) and without (top) the use of the de-ringing filter

5 Perspectives towards future improvements

Some ideas towards future improvements of Wavelet based SVC solutions are reported:

Motion estimation resolution problem in current t+2D implementation

In order to support spatial resolution scalability, temporal levels that will be decoded on targeted lower spatial resolutions must use large macroblocks (64x64, 32x32), since from an implementation point of view, decoding is not working at low resolution for smaller size blocks.

2D+t+2D (Stool) inter-layer issues (currently not supported)

- consistent mode decision (e.g. intrablock) across spatial resolution layers
- consistent motion estimation across spatial resolution layers, for ensuring:
  - good prediction of LL on higher spatial resolution
  - optimal coding of motion field

New tools

Replace Intra-coding mode with Motion Adaptive Transform (better tuned to small areas of uncovered background)

Entropy coding (both for t+2D and 2D+t+2D architectures)

- Same scale temporal and spatial subbands appear to be coded separately (which means at the level of individual subband level), but given the 3D EBCOT used, good context requires the use of motion information which is not available within any given subband, and should be predicted or estimated to take into account the advantage of context information.

Temporal transform

- *Temporal filter*
  The application of the update step in MCTF is not justified in all applications. Therefore the application of delta filters low-pass filters, such as 1/3 (5/3 without the update step), should be supported. This feature can be useful in scenarios where lower-complexity is needed or in cases when key frames, which in specific cases should not be different than original frames, have to be accessible without applying IMCTF. Although avoiding the update step can lead to lower compression efficiency on the original sequence, quality of temporal scalability can be improved.
- *Motion block size*
  Current evaluation software uses 8x8 motion blocks and their multiples (16x16, 32x32,... size blocks) as the basic motion units. In previous standards and research results it has been shown that flexible motion block size, specifically the possibility of using smaller blocks such as 4x4 blocks, can improve coding efficiency. Therefore more flexible motion model is needed. Moreover, when intra blocks, as blocks that cannot be predicted from previous frames, are employed finer partitioning of frames is needed as these blocks usually correspond to smaller areas.
- *Scalable motion information*
  Various spatio-temporal decomposition schemes requires different types of motion-information scalability. Specifically, in 2D+t and 2D+t+2D schemes motion estimation is performed on different spatial resolution levels. In such scenarios the obtained motion information on different spatial levels is highly correlated and therefore its embedded, i.e. scalable coding can provide further compression gain.

Spatial transform

Spatial wavelet transform has traditionally been performed in a non-adaptive way. Lifting implementation of wavelet transform enables low-complexity adaptation according to spatial signal characteristics. Recently presented technique [54] uses adaptation on intra-inter coded block boundaries which avoids the application of intra prediction. Future applications of this approach can be based on other available information, such as motion vector gradient, that drives the adaptation.

6 *VidWav history*

This section provides an overview of the history of VidWav AhG from its establishment during the 70th MPEG meeting (Palma, ES). In the first subsection all the documents produced within the VidWav are summarised, while in the second subsection the participants are listed.

Requirements reference documents for SVC VidWav AhG are [55,56].

6.1 Meetings and input documents

Meeting 71 Hong-Kong, China:

10 input documents:

11680	Ruiqin Xiong Jizheng Xu Feng Wu Dongdong Zhang	Studies on Spatial Scalable Frameworks for Motion Aligned 3D Wavelet Video Coding
11681	Dongdong Zhang Jizheng Xu Hongkai Xiong Feng Wu	Improvement for In-band Video Coding with Spatial Scalability
11713	Markus Beermann Mathias Wien	Application of the Bilateral Filter for Quality-Adaptive Reconstruction
11732	Christophe Tillier Beatrice Pesquet-Popescu	CBR 3-band MCTF
11738	Gregoire Pau Beatrice Pesquet-Popescu	Optimized Prediction of Uncovered Areas in Wavelet Video Coding
11739	Gregoire Pau Beatrice Pesquet-Popescu	Four-Band Linear-Phase Orthogonal Spatial Filter Bank in Wavelet Video Coding
11741	Gregoire Pau Jerome Vieron Beatrice Pesquet-Popescu	Wavelet Video Coding with Flexible 5/3 MCTF Structures for Low End-to-end Delay
11748	G.C.K. Abhayaratne Ebroul Izquierdo	Wavelets based residual frame coding in t+2D wavelet video coding
11750	Marta Mrak Nikola Sprljan G.C.K. Abhayaratne Ebroul Izquierdo	Scalable motion vectors vs unlimited precision based motion compensation at the decoder in t+2D wavelet video coding
11757	Woo-Jin Han Kyohyuk Lee	Comments on wavelet-based scalable video coding technology

1 output document:

6914	Description of Exploration Experiments in Wavelet Video Coding

During the 71st meeting, wavelet based software from Microsoft Research Asia (MSRA) has been chosen as the common software for the investigation and evaluation within the VidWav.

Meeting 72 Busan, Korea :

7 input documents:

11844	Z. K. Lu W. S. Lin Z. G. Li K. P. Lim X. Lin S. Rahardja E. P. Ong S. S. Yao	Perceptual Region-of-interest (ROI) based Scalable Video Coding
11952	ChinPhek Ong ShengMei Shen MenHuang Lee Yoshimasa Honda	Wavelet Video Coding - Generalized Spatial Temporal Scalability (GSTS).
11975	Ruiqin Xiong Jizheng Xu Feng Wu	Coding Perfromance Comparison Between MSRA Wavelet Video Coding and JSVM
11976	Yihua Chen Jizheng Xu Feng Wu Hongkai Xiong	Improvement of the update step in JSVM
12008	Markus Beermann Mathias Wien	De-ringing filter proposal for the VIDWAV Evaluation software
12056	Christophe Tillier Grégoire Pau Béatrice Pesquet-Popescu	Coding performance comparison of entropy coders in wavelet video coding
12058	Grégoire Pau Béatrice Pesquet-Popescu	Comparison of Spatial *M*-band Filter Banks for *t+2D* Video Coding

1 output document:

7098	Description of Exploration Experiments in Wavelet Video Coding

Meeting 73 Poznan, Poland:

7 input documents:

12176	Vincent Bottreau Grégoire Pau Jizheng Xu	Vidwav evaluation software manual
12286	Ruiqin Xiong Jizheng Xu Feng Wu	Responses to Vidwav EE1
12303	Grégoire Pau Maria Trocan Béatrice Pesquet-Popescu	Bidirectional Joint Motion Estimation for Vidwav Software
12339	Ruiqin Xiong Xiangyang Ji Dongdong Zhang Jizheng Xu Grégoire Pau Maria Trocan Vincent Bottreau	Vidwav Wavelet Video Coding Specifications
12374	Markus Beermann	Joint reduction of ringing and blocking for VidWav
12376	Yongjun Wu John Woods	Aliasing reduction for subband/wavelet scalable video coding
12410	Soroush Ghanbari Leszek Cieplinski	Results of Vidwav Exploration Experiment 3

2 output documents:

7334	Wavelet Codec Reference Document and Software Manual
7333	Description of Exploration Experiments in Wavelet Video Coding

Meeting 74 Nice, France:

7 input documents:

12616	Gregoire Pau Beatrice Pesquet-Popescu	Proposal of Vidwav OBMC bug fix
12633	Nikola Sprljan Marta Mrak Naeem Ramzan Ebroul Izquierdo	Motion Driven Adaptation of Spatial Wavelet Transform
12639	Nicola Adami Michele Brescianini Riccardo Leonardi	Edited version of the document SC 29 N 7334
12640	Markus Beermann Mathias Wien	Wavelet Video Coding EE4: Joint Reduction of Ringing and Blocking

12642	Nicola Adami Michele Brescianini Riccardo Leonardi Alberto Signoroni	New prediction schemes for scalable wavelet video coding
12643	Nicola Adami Michele Brescianini Riccardo Leonardi Alberto Signoroni	Performance evaluation of the current Wavelet Video Coding Reference Software
12699	Ruiqin Zhong	Verification of Vidwav EE4 results of RWTH

3 output documents:

7571	Draft Status Report on Wavelet Video Coding Exploration
7572	Description of Exploration Experiments in Wavelet Video Coding
7573	Wavelet Codec Reference Document and Software Manual V2.0

Meeting 75 Bangkok, France:

4 input documents:

12941	Nikola Sprljan, Marta Mrak, Toni Zgaljic, Ebroul Izquierdo	Software proposal for Wavelet Video Coding Exploration Group
12960	Nicola Adami Michele Brescianini Riccardo Leonardi, Livio Lima, Alberrto Signoroni	Report on Wavelet Video Coding EE5: Visual Performance Evaluation
12970	Riccardo Leonardi, Alberrto Signoroni, Sebastien Brangoulo	Proposed Status Report on Wavelet Video Coding Exploration
13011	Grégoir Pau, Sébastien Bragoulo, Beatrice Pesquet-Popescu	Integration of Bidirectional Joint Motion Estimation for Vidwav Software

3 output documents:

7822	Status Report on Wavelet Video Coding Exploration Version 1
7823	Description of Testing in Wavelet Video Coding
7824	Wavelet Video Coding : an Overview

Meeting 76 Montreux, Switzerland:

5 input documents:

13146	Marta Mrak Nikola Sprljan Ebroul Izquierdo	Performance evidence of software proposal for Wavelet Video Coding Exploration group

13246	Mathias Wien	JSVM-4.0 bitstreams for VIDWAV visual evaluation
13294	Riccardo Leonardi Michele Brescianini Hassan Khalil Ji-Zheng Xu Sébastien Brangoulo	Report on Testing in Wavelet Video Coding
13295	Riccardo Leonardi Michele Brescianini Hassan Khalil	Extended Scalability Performance of Wavelet Video Coding
13301	Nicola Adami, Alberto Signoroni, Riccardo Leonardi	Verification of proposal: "Performance evidence of software proposal for Wavelet Video Coding Exploration group"

6.2 VidWav participation

Academic Institutions

- ENST Paris
- University of Brescia, Italy
- RWTH Aachen University
- Queen Mary, University of London, United Kingdom.
- Rensselaer Polytechnic Institute
- Institute of Computing Technology, Chinese Academy of Sciences
- Image Communication Institute, Shanghai Jiao Tong University
- University of New South Wales, Australia
- University of Sheffield, United Kingdom

Research Institutions and Industry

- Institute for Infocomm Research, Singapore
- IRISA/INRIA Rennes
- Microsoft Research Asia
- Mitsubishi Electric ITE-VIL
- Samsung Electronics
- Thomson R&D

7 References

[1] ISO/IEC JTC1/SC29/WG11, "Scalable Video Model V 1.0," M6372, 68th MPEG Meeting, Münich, Germany, Mar. 2004.

[2] ISO/IEC JTC1/SC29/WG11, "Scalable Video Model 2.0," N6520, 69th MPEG Meeting, Redmond, WA, USA, Jul. 2004.

[3] S.-T. Hsiang and J.W. Woods, "Embedded Video Coding Using Invertible Motion Compensated 3-D Subband/Wavelet Filter Bank," Signal Processing: Image Communication, vol. 16, pp. 705-724, May 2001.

[4] Y. Andreopoulos, M. van der Schaar, A. Munteanu, J. Barbarien, P. Schelkens and J. Cornelis, “Complete-to-overcomplete discrete wavelet transform for fully scalable video coding with MCTF,” in Proc. of VCIP 2003, SPIE vol. 5150, pp. 719-731, Lugano (CH), July 2003.

[5] Subjective test results for the CfP on Scalable Video Coding Technology, ISO/IEC JTC1/SC29/WG11, M10737, 68th MPEG Meeting, Munich, Germany, Mar. 2004.

[6] Report of the Subjective Quality Evaluation for SVC CE1, ISO/IEC JTC1/ SC29/ WG11, N6736, 70th MPEG Meeting, Palma de Mallorca, Spain, Oct. 2004.

[7] J.R. Ohm, “Three-dimensional subband coding with motion compensation,” IEEE Trans. Image Process., vol. 3, no. 5, pp. 559–571, Sept. 1994.

[8] S.-J. Choi and J.W. Woods, “Motion-compensated 3-D subband coding of video,” IEEE Trans. Image Process., vol. 8, no. 2, pp. 155–167, Feb. 1999.

[9] A. Secker and D. Taubman, "Lifting-Based Invertible Motion Adaptive Transform (LIMAT) Framework for Highly Scalable Video Compression," IEEE Trans. Image Processing, vol. 12, no. 12, pp. 1530-1542, Dec. 2003.

[10] V. Bottreau, M. Benetiere, B. Felts, and B. Pesquet-Popescu, "A fully scalable 3D subband video codec," in Proc. IEEE Int. Conf. on Image Processing (ICIP 2001), vol. 2, pp. 1017-1020, Oct. 2001.

[11] Jizheng Xu, Ruiqin Xiong, Bo Feng, Gary Sullivan, Ming-Chieh Lee, Feng Wu, Shipeng Li: “3-D Subband Video Coding Using Barbell Lifting”, ISO/IEC JTC1/SC29/WG11, M10569/S05, 68th MPEG Meeting, Münich, Germany, Mar. 2004.

[12] G. Pau,C. Tillier, B. Pesquet-Popescu and H. Heijmans, “Motion Compensation and Scalability in Lifting-Based Video Coding” Signal Processing: Image Communication, special issue on Wavelet Video Coding, Elsevier/EURASIP, Vol. 19, p. 577-600, August 2004.

[13] B. Kim, Z. Xiong, and W. Pearlman, “Low bit-rate scalable video coding with 3d set partitioning in hierarchical trees (3d SPIHT),” IEEE Trans. Circ. Syst. for Video Tech., vol. 10, pp. 1374–1387, dec 2000.

[14] B. Pesquet-Popescu and V. Bottreau, “Three dimensional lifting schemes for motion compensated video compression,” IEEE Int. Conf. Accoust. Speech and Signal Proc., pp. 1793–1796, 2001.

[15] V. Bottreau, M. Benetiere, B. Felts, and B. Pesquet-Popescu, “A fully SCalable 3d subband video codec,” IEEE Int. Conf. Image Proc., pp. 1017–1020, 2001.

[16] K. Ho and D. Lun, “Efficient wavelet based temporally scalable video coding,” IEEE Int. Conf. Signal Proc., pp. 881–884, 2002.

[17] A. Secker and D. Taubman, “Lifting based invertible motion adaptive transform, LIMAT, framework for highly scalable video compression,” IEEE Trans. Image Proc., vol. 12, pp. 1530–1542, Dec. 2003.

[18] M. V. der Schaar and D. Turaga, “Unconstrained motion compensated temporal filtering (UMCTF) framework for wavelet video coding,” IEEE Int. Conf. Accoust. Speech and Signal Proc., pp. 81–84, 2003.

[19] P. Chen and J. W. Woods, “Bidirectional mc-ezbc with lifting implementation,” IEEE Transactions on Circuit and Systems for Video Technology 14, pp. 1183–1194, October 2004.

[20] Nathalie Cammas, “Codage video scalable par maillages et ondelettes t+2D,” PhD Thesis, France Telecom, IRISA,Université Rennes 1, 2004.

[21] Markus Flierl and Bernd Girod "Video Coding with Motion-Compensated Lifted Wavelet Transforms," EURASIP Journal on Image Communication, Special Issue on Subband/Wavelet Interframe Video Coding, vol. 19, no. 7, pp. 561-575, Aug. 2004.

[22] 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.

[23] Y. Wu and J.W. Woods, "Aliasing reduction for scalable subband/wavelet video coding," ISO/IEC JTC1/SC29/WG11, M12376, 73rd MPEG Meeting, Poznan, Poland, July 2005.

[24] M. van der Schaar and D. Turaga, "Unconstrained motion compensated temporal filtering (UMCTF) framework for wavelet video coding," in Proc IEEE Int. Conf. Acoust. Speech and Signal Proc., pp. 81–84, , Hong-Kong, China, Apr. 2003.

[25] Mehrseresht and D. Taubman, "Adaptively weighted update steps in motion compensated lifting based on scalable video compression," in Proc Int. Conf. on Image Processing, Barcelona, Spain, Sept. 2003.

[26] D. Taubman, D. Maestroni, R. Mathew and S. Tubaro, "SVC Core Experiment 1, Description of UNSW Contribution", ISO/IEC JTC1/ SC29/ WG11, M11441, 70th MPEG Meeting, Palma de Mallorca, Spain, Oct. 2004.

[27] D.S. Turaga, M. van der Schaar and B. Pesquet-Popescu, "Complexity Scalable Motion Compensated Wavelet Video Encoding", IEEE Trans. on Circuits and Syst. for Video. Technol., vol. 15, no. 8, pp. 982-993, Aug. 2005.

[28] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, San Diego, CA, 1998.

[29] Y. Andreopoulos, A. Munteanu, G. Van Der Auwera, J. Cornelis and P. Schelkens "Complete-to-overcomplete discrete wavelet transforms: theory and applications," IEEE Transactions on Signal Processing, vol. 53, nr. 4, pp. 1398-1412, 2005.

[30] T. Kimoto and Y. Miyamoto, "Multi-resolution motion compensated temporal filtering for 3d wavelet coding," ISO/IEC JTC1/SC29/WD11 M10569/S09, March 2004, munich, Germany.

[31] G. Baud, M. Duvanel, J. Reichel, and F. Ziliani, "VisioWave scalable video CODEC proposal," ISO/IEC JTC1/SC29/WG11M10569/S20, March 2004, munich, Germany.

[32] Y. Andreopoulos, A. Munteanu, J. Barbarien, M. van der Schaar, J. Cornelis, and P. Schelkens, "Inband motion compensated temporal filtering," Signal Processing: Image Communication 19, pp. 653–673, August 2004.

[33] Y. Wang, S. Cui, , and J. E. Fowler, "3D video coding using redundant-wavelet multihypothesis and motioncompensated temporal filtering," in Proceedings of the International Conference on Image Processing, 2, pp. 755–758, (Barcelona, Spain), September 2003.

[34] Davide Maestroni, Marco Tagliasacchi and Stefano Tubaro, In-band adaptive update step based on local content activity in Proc Visual Comm. and Image Proc. 2005, SPIE vol. 5960 (nr.19), Beijing, China, July 2005.

[35] Y. Wang, S. Cui, , and J. E. Fowler, "3D video coding using redundant-wavelet multihypothesis and motioncompensated temporal filtering," in Proceedings of the International Conference on Image Processing, 2, pp. 755–758, (Barcelona, Spain), September 2003.

[36] H.-W. Park and H.-S. Kim, "Motion estimation using low-band-shift method for wavelet-based movingpicture coding," IEEE Transactions on Image Processing 9, pp. 577–587, April 2000.

[37] J. C. Ye and M. van der Schaar, "Fully scalable 3-D overcomplete wavelet video coding using adaptive motion compensated temporal filtering," in Visual Communications and Image Processing, T. Ebrahimi and T. Sikora, eds., pp. 1169–1180, Proc. SPIE 5150, (Lugano, Switzerland), July 2003.

[38] X. Li, "Scalable Video Compression via Overcomplete Motion Compensated Wavelet Coding", Signal Processing: Image Communication, vol. 19, no. 7, pp. 637-651, 2004

[39] D. Zhang, S. Jiao, J. Xu, F. Wu, W. Zhang and H. Xiong, "Mode-based temporal filtering for in-band wavelet video coding with spatial scalability" in Proc Visual Comm. and Image Proc. 2005, SPIE vol. 5960 (nr.38), Beijing, China, July 2005.

[40] P.J. Burt and E.H. Adelson, "The Laplacian pyramid as a compact image code", IEEE Trans. on Communications, vol. 31, pp.532-540, Apr. 1983.

[41] N. Adami, M. Brescianini, R. Leonardi, 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.

[42] 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 Visual Comm. and Image Proc. 2005, SPIE vol. 5960 (nr.58), Beijing, China, July 2005.

[43] 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.

[44] R. Xiong, J. Xu, F. Wu, S. Li, "Studies on spatial scalable frameworks for motion aligned 3D wavelet video coding" in Proc Visual Comm. and Image Proc. 2005, SPIE vol. 5960 (nr.21), Beijing, China, July 2005.

[45] ISO/IEC JTC1/SC29/WG11, "Report of the 71st meeting", N6872, 71st MPEG meeting, Hong Kong, China, January 2005.

[46] ISO/IEC JTC1/SC29/WG11, "Exploration experiments on tools evaluation in Wavelet Video Coding", N6914, 71st MPEG meeting Hong Kong, January 2005.

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

[48] 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. 2005.

[49] R. Xiong, J. Xu, F. Wu and D. Zhang, "Studies on Spatial Scalable Frameworks for Motion Aligned 3D Wavelet Video Coding", ISO/IEC JTC1/SC29/WG11, M11680, 71th MPEG Meeting, Hong Kong, China, Jan. 2005.

[50] D. Zhang, J. Xu, H. Xiong and F. Wu, "Improvement for In-band Video Coding with Spatial Scalability" , ISO/IEC JTC1/SC29/WG11, M11681, 71th MPEG Meeting, Hong Kong, China, Jan. 2005.

[51] N. Adami, M. Brescianini and R. Leonardi, "Performance evaluation of the current Wavelet Video Coding Reference Software", ISO/IEC JTC1/SC29/WG11, M12643, 74th MPEG Meeting, Nice, France, Oct. 2005.

[52] ISO/IEC-JTC1 and ITU-T, "Joint Scalable Video Model (JSVM) 3.0 Reference Encoding Algorithm Description", ISO/IEC JTC1/SC29/WG11, N7311, 73th MPEG Meeting, Poznan, Poland, July 2005.

[53] M. Beermann, M. Wien, “Wavelet Video Coding, EE4: Joint Reduction of Ringing and Blocking”, ISO/IEC JTC1/SC29/WG11, M12640, 74th MPEG Meeting, Nice, France, Oct. 2005.

[54] N. Sprljan, M. Mrak, N. Ramzan, and E. Izquierdo, “Motion Driven Adaptation of Spatial Wavelet Transform”, ISO/IEC JTC1/SC29/WG11, M12633, 74th MPEG Meeting, Nice, France, Oct. 2005.

[55] ISO/IEC JTC1/SC29/WG11, “Requirements and Applications for Scalable Video Coding v.5,” N6505, Redmond, July 2004.

[56] ISO/IEC JTC1/SC29/WG11, “Applications and Requirements for Scalable Video Coding,” N6880, Hongkong, China, January 2005.

[57] R. Xiong, J. Xu and F. Wu, “Coding Perfromance Comparison Between MSRA Wavelet Video Coding and JSVM”, ISO/IEC JTC1/SC29/WG11, M11975, 72nd MPEG Meeting, Busan, Korea, Apr. 2005.

[58] R. Xiong, J. Xu and F. Wu, “Responses to VidWav EE1”, ISO/IEC JTC1/SC29/WG11, M12286, 73rd MPEG Meeting, Poznan, Poland, July 2005.

[59] ISO/IEC JTC1/SC29/WG11, “Description of Testing in Wavelet Video Coding,” ISO/IEC JTC1/SC29/WG11, N7823, 75th MPEG Meeting, Bangkok, January, 2004.

[60] Nikola Sprljan, Marta Mrak, Toni Z., and E. Izquierdo, “Software proposal for Wavelet Video Coding Exploration group,” ISO/IEC JTC1/SC29/WG11, M12941, 75th MPEG meeting, Bangkok, January 2006.

[61] M. Wien, “JSVM bitstreams for VIDWAV visual evaluation,” ISO/IEC JTC1/SC29/WG11, M13264, , 76th MPEG Meeting, Montreux, Switzerland, April 2006.

[62] R. Leonardi, M. Brescianini, H. Khalil, J.-Z. Xu and S. Brangoulo, “Report on Testing in Wavelet Video Coding,” ISO/IEC JTC1/SC29/WG11, M13294, , 76th MPEG Meeting, Montreux, Switzerland, April 2006.

[63] R. Leonardi, M. Brescianini and H. Khalil, “Extended Scalability Performance of Wavelet Video Coding,” ISO/IEC JTC1/SC29/WG11, M13295, , 76th MPEG Meeting, Montreux, Switzerland, April 2006.

[64] M. Mrak, N. Sprljan, T. Zgaljic, N. Ramzan, S. Wan and E. Izquierdo, Performance evidence of software proposal for Wavelet Video Coding Exploration group, ISO/IEC JTC1/SC29/WG11, M13146, , 76th MPEG Meeting, Montreux, Switzerland, April 2006.