Error-resilient scalable compression based on distributed video coding

Mourad Ouaret *, Frédéric Dufaux, Touradj Ebrahimi

MultiMedia Signal Processing Group (MMSPG), Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

ARTICLE INFO

ABSTRACT

Distributed Video Coding (DVC) is a new paradigm for video compression based on the information theoretical results of Slepian–Wolf (SW) and Wyner–Ziv (WZ). In this work, a performance analysis of image and video coding schemes based on DVC is presented, addressing temporal, quality and spatial scalability. More specifically, conventional coding is used to obtain a base layer while WZ coding generates the enhancement layers. At the decoder, the base layer is used to construct Side Information (SI) for the DVC decoding process. Initially, we show that the scalable DVC approach is codec-independent, which means that it is independent from the method used to encode the base layer. Moreover, the influence of the base layer quality on the overall performance of the schemes is studied. Finally, evaluation of the proposed schemes is performed in both cases, with and without transmission errors. The simulation results show that scalable DVC has a lower compression efficiency than conventional scalable coding (i.e. scalable video coding and JPEG2000 for video and image, respectively) in error-free conditions. On the other hand, the DVC-based schemes show better error resilience as they outperform conventional scalable coding in error-prone conditions. More specifically, the Rate Distortion (RD) performance of the proposed schemes for image coding is compared with respect to Reed Solomon (RS) protected JPEG2000. While the latter exhibits a cliff effect as its performance dramatically decreases after a certain error rate, the performance of the DVC-based schemes decreases in a steady way with error rate increase.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Scalable coding is becoming important nowadays in heterogeneous multimedia networks. Different clients on a network might require decoding the same multimedia content at different frame rates, qualities or resolutions, depending on the requirements and the available resources at the client side. For this purpose, scalable coding encodes the content once and enables decoding at different temporal, quality (or Signal to Noise Ratio (SNR)) or spatial resolutions. Scalable coding is attractive for several applications such as surveillance cameras and media browsing.

A performance analysis of scalable schemes for image and video coding based on Distributed Video Coding (DVC) is performed in this paper, dealing with temporal, quality and spatial scalability. The performance is evaluated for different base layer qualities. Furthermore, Scalable DVC (SDVC) is compared with scalable conventional coding in terms of compression efficiency.

Scalable Video Coding (SVC) [1,2] is introduced by the Moving Picture Experts Group (MPEG) for scalable video compression. SVC is an extension of the AVC/H.264 [3] standard using a lifting framework. The temporal scalability in SVC is achieved by the hierarchical B coding structure in AVC/H.264. For spatial scalability, the video is

* Corresponding author.
  E-mail address: mourad.ouaret@epfl.ch (M. Ouaret).

first downsampled at the required spatial resolutions. The corresponding enhancement layers are generated by predictive dependencies with respect to the lower layers. The encoding and decoding processes start at the base layer resolution and progress toward the higher layers. Further, the quality scalability is generated in the same way as the spatial scalability by creating predictive dependencies with respect to lower layers with the same spatial resolution. SVC achieves a very good compression efficiency as it exploits the correlation at the encoder side. In other words, SVC entails high complexity encoding. Furthermore, SVC is based on a deterministic predictive framework (i.e. prediction loop), which impairs the performance of the codec in error-prone conditions.

JPEG2000 [4] is a state-of-the-art standard for image coding. Among many features, it enables very efficient scalability. More specifically, spatial scalability is enabled thanks to the use of the Discrete Wavelet Transform (DWT), which results in a dyadic data structure. Quality scalability is ensured by the quality layers, where each one represents a quality increment.

DVC [5] is a new emerging paradigm for video compression. Unlike conventional video coding schemes (AVC/H.264 [3]) where the source statistics are exploited at the encoder, DVC shifts this task toward the decoder. This results in low complexity encoders and highly complex decoders. Therefore, DVC is suitable for applications where computational power is sparse at the encoder. In addition, DVC has a good resilience to transmission errors. This results from the fact that DVC is based on a statistical framework rather than the closed-loop prediction used in conventional video coding. A study of DVC's performance in the presence of transmission errors has been presented in [6], showing its good error resilience properties. Further, Error Concealed (EC) DVC is shown to outperform EC AVC/H.264 in different modes in [7]. The DVC concealment used spatial EC to improve the performance of temporal EC.

Little research has been performed on scalable schemes based on DVC and it mainly focused on video coding. Tagliasacchi et al. [8] implemented a scalable version of PRISM (Power-efficient, Robust, hIgh compression, Syndrome-based Multimedia coding) [9]. The approach enhances an AVC/H.264 base layer with a PRISM refinement bitstream resulting in a spatio-temporal scalable video codec. It focuses on the case where estimation and most of the motion compensation task is performed at the decoder. Results show that scalable PRISM outperforms non-scalable PRISM and H.263+ Intra, but has a poorer performance when compared to motion compensated H.263+. In fact, since the base layer used AVC/H.264, comparison should have been made with respect to the latter.

A solution to the problem of scalable predictive video coding is introduced in [10] by posing it as a variant of the Wyner–Ziv (WZ) Side Information (SI) problem. It discusses mainly quality scalability. Results show that the proposed codec is approximately 4.0 dB superior to a naive scalable codec based on a conventional codec. In addition, motion compensation is performed at the encoder which increases its complexity.

Finally, WZ codes are integrated into a standard video codec to achieve efficient and low complexity scalable coding in [11]. Experimental results show improvements in coding efficiency of $3.0-4.5$ dB over MPEG4 FGS [12] for video sequences with high temporal correlation.

Schemes for SDVC image and video coding dealing with temporal, spatial and quality scalability are introduced in this paper. The base layer is generated from the conventional part of the stream, which is used to generate the SI. This is done either temporally by motion compensated interpolation or spatially by a spatial bicubic interpolation. The enhancement layer is represented by the WZ bits. Further, the DVC decoding process generates the decoded image or video at full spatial resolution and enhanced quality. The performance of DVC SVC is compared to SVC in error-free conditions. Furthermore, the influence of the base layer quality on the schemes performance is also investigated. For image compression, the comparison is made with scalable JPEG2000. In error-prone conditions, the introduced schemes for image coding are compared to scalable JPEG2000. In this case, the different parts of the JPEG2000 stream are interleaved [13] and then protected with Reed Solomon (RS) [14] codes. Moreover, the error resilience tools of JPEG2000 are switched on to enable the recovery from errors present in the stream.

This paper is outlined as follows. First, the theoretical foundations of DVC as well as the practical DVC scheme used in this research are reviewed in Section 2. Application of DVC to scalable image and video coding is presented in Section 3 addressing different scalabilities. Then, the error resilience tools of JPEG2000 are described in Section 4. The test material and conditions in addition to the simulation results are discussed in Section 5. Finally, some conclusions are drawn in Section 6.

2. Distributed video coding (DVC)

2.1. Theoretical DVC

DVC is the result of the information-theoretic bounds established for Distributed Source Coding (DSC) by Slepian and Wolf [15] for lossless coding, and by Wyner and Ziv [16] for lossy coding with SI at the decoder. Lossless DSC refers to two correlated random sources separately encoded and jointly decoded by exploiting the statistical dependencies.

If we consider two statistically dependent random sequences X and Y, rates R_X and R_Y can be achieved by entropy coding such that $R_X \geqslant H(X)$ and $R_Y \geqslant H(Y)$, where $H(X)$ and $H(Y)$ are the entropies of X and Y, respectively. The Slepian–Wolf (SW) theorem proves that a better rate can be achieved with joint decoding and gives tighter bounds for the total rate $R_X + R_Y$. The admissible rate region established by SW, which corresponds to the shaded area depicted in Fig. 1, is defined by

$$R_X \geqslant H(X|Y), \quad R_Y \geqslant H(Y|X)$$

$$R_X + R_Y \geqslant H(X,Y) \tag{1}$$

Decoding with SI is considered as a special case of DSC. In this case, the source X depends on some SI Y, which corresponds to the black dot on the region border shown in Fig. 1. Later on, Wyner and Ziv established bounds for lossy compression with SI at the decoder as an extension to the SW theorem. In this case, the source X is encoded without having access to the SI Y. On the other hand, the decoder has access to the SI to produce X with a certain distortion D.

2.2. Practical DVC

Fig. 2 shows the DVC architecture used in this work [17].

At the encoder, the frames are separated into two sets. The first one is the key frames, which are fed to a conventional AVC/H.264 Intra encoder. The second set is the WZ frames. The latter are transformed and then quantized prior to WZ encoding. The same 4×4 separable integer transform as in AVC/H.264 is used with properties similar to the Discrete Cosine Transform (DCT) [18]. Then, the same bands are grouped together and the different bitplanes are extracted and then fed to a turbo encoder [19]. The latter offers near-channel capacity error correcting capability. Furthermore, a Cyclic Redundancy Check (CRC) [20] is computed for each quantized bitplane and transmitted to the decoder. The frequency of the key frames is defined by the Group Of Pictures (GOP).

At the decoder, the key frames are conventionally decoded and then used to generate the SI for the WZ decoder. In the monoview case, Motion Compensation Temporal Interpolation (MCTI) [21] is used to generate the SI. For this purpose, MCTI uses the key frames to perform motion estimation. The resulting motion vectors are interpolated at midpoint as illustrated in Fig. 3.

A virtual channel is used to model the correlation between the DCT coefficients of the original and SI frames.
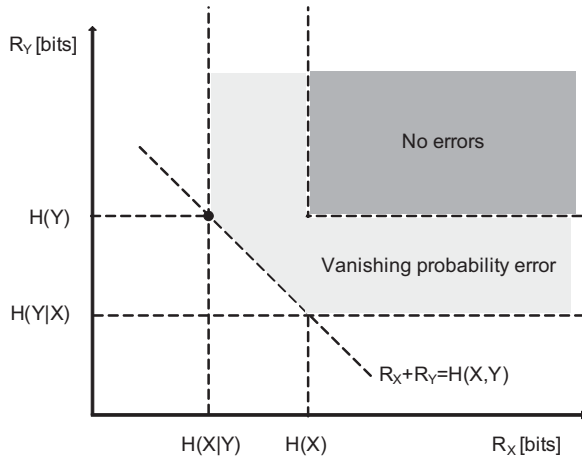

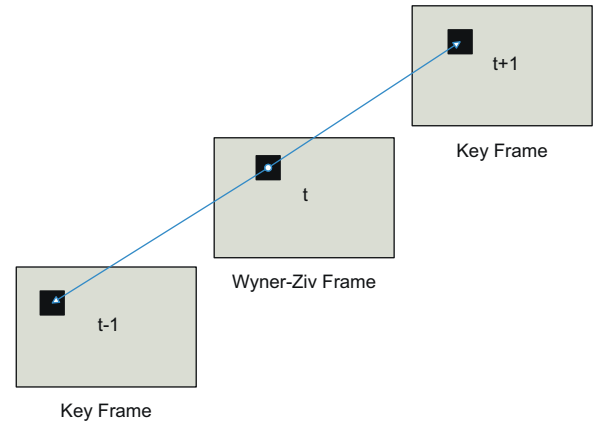Fig. 1. Achievable rate region defined by the Slepian–Wolf bounds.

Fig. 3. Motion compensation temporal interpolation (MCTI). MV is a motion vector in the forward direction.
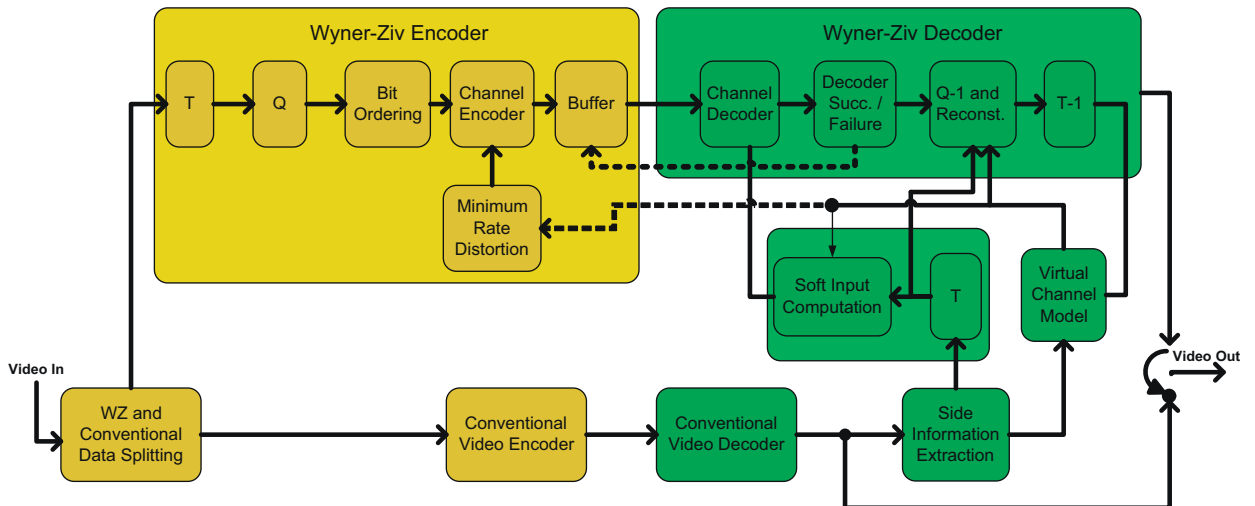

Fig. 2. Conventional DVC architecture.

It is shown that the residual of the DCT coefficients follows a Laplacian distribution [5]. The reconstruction process [22] uses the SI along with decoded bins to recover the original frame up to a certain quality. The decoder accepts the SI DCT value as a reconstructed one if it fits into the quantization interval corresponding to the decoded bin. Otherwise, it truncates the DCT value into the quantization interval. This DVC scheme is decoder driven as the request for parity bits from the encoder is performed via a feedback channel until successful decoding. The decoding is considered successful if the decoded bitplane's error probability is lower than 10^{-3} and its CRC matches the one received from the encoder.

3. Scalable DVC (SDVC)

In this section, DVC is used to encode images and video in a scalable stream. This approach is shown to be base layer codec-independent [23]. In other words, the base layer can use any conventional codec. In addition, DVC is attractive for error resiliency [6,24]. These advantages are due to its statistical framework.

Fig. 4 shows a scenario where codec-independent scalability can be a very interesting feature. Suppose that we have three content providers A, B and C. A and B will encode the input using two different conventional coders $coder_A$ and $coder_B$ generating streams S_A and S_B. Provider C generates the WZ refinement stream S_C for the same input. In network Net_1, clients 1 and 2 have different conventional decoders dec_A and dec_B, respectively. Instead of sending two completely different streams which are

both codec A and B compliant, both clients will decode the different base layers and use the same refinement stream. This would reduce traffic in the corresponding network. In network Net_2, clients 3, 4 and 5 have the conventional decoders dec_A, dec_B and dec_A, respectively. Clients 3 and 4 would, for example, just decode at the base layer quality retrieved from their subnetwork Net_{21}. Further, client 5 receives both base and enhancement layers from its subnetwork Net_{22}. At the same time, client 4 has access to subnetwork Net_{22} via which it will receive the enhancement layer at a later stage. Therefore, the codec-independent scalability feature offers high flexibility in the way the stream is distributed in the network and helps reducing the traffic at the same time in certain cases.

3.1. SDVC for image coding

3.1.1. Quality scalability

Fig. 5 shows the DVC image coding scheme used for quality scalability. At the encoder side, the input image is simultaneously encoded by WZ and conventional coding. At the decoder side, the conventional part of the stream is decoded generating the base layer. Further, the latter is used as SI along with the parity bits in the WZ decoding process to increase the quality of the base layer.

There are two ways to control the quality of the decoded image, which are the number of decoded frequency bands and bitplanes. For band level scalability, as more bands are decoded, the quality of the decoded image increases. On the other hand, if all bands are decoded, the quality of the decoded image increases with
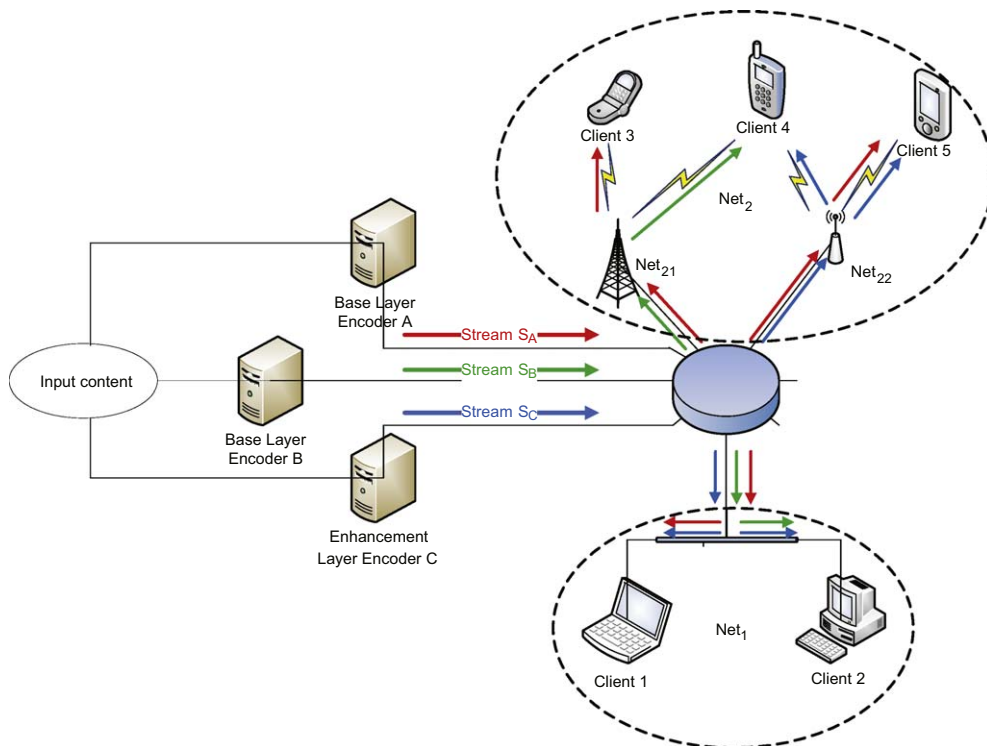

Fig. 4. Codec-independent scenario.

the number of corrected bitplanes. Finally, the best scalability granularity is achieved when a combination of both decoded bands and bitplanes is used.

3.1.2. Spatial and quality scalability

A scheme for image coding ensuring both spatial and quality scalability is illustrated in Fig. 6. It is an extension of the quality scalability scheme. Therefore, a spatial downsampling/upsampling is introduced prior/after the conventional encoding/decoding in the lower branch of the scheme. In parallel, a quality refinement stream is sent for the enhancement of the base layer. Then, the base layer or its enhanced version is used to generate SI to decode the image at the full spatial resolution. Note that the downsampling is preceded by a convolution with a

Fig. 5. DVC architecture for image coding for quality scalability.

Fig. 6. DVC architecture for image coding for spatial and quality scalability.

Gaussian kernel to reduce the effect of aliasing, and the upsampling is a bi-cubic interpolation performed in both vertical and horizontal directions.

3.2. SDVC for video coding

In this section, schemes based on DVC for video coding are described dealing with temporal, spatial and quality scalability. In this case, the DVC schemes for spatial and quality scalability are identical to the ones used for image coding as the same encoding and decoding process is applied to each frame (i.e. image) of the video.

3.2.1. Temporal scalability

The DVC scheme used in this research already ensures temporal scalability. By decoding only key frames, the decoded video has half the original temporal resolution ($GOP = 2$). This makes temporal scalability straight forward. Furthermore, the scheme can be easily extended to n temporal enhancement layers. In this case, a GOP of 2^n is chosen. This results in a base layer with $1/2^n$ the full temporal resolution, whereas the decoding of each enhancement layer doubles the temporal resolution. Fig. 7 illustrates the decoding process for two enhancement layers ($GOP = 4$). Nevertheless, a higher GOP size generally leads to worse compression efficiency for DVC.

3.2.2. Temporal, spatial and quality scalability

The scheme is depicted in Fig. 8, by combining the three schemes previously described. First, the odd frames are spatially downsampled and then conventionally encoded in block **A**. When decoded, the latter produce the base layer, $Y_B(2i - 1)$ and $Y_B(2i + 1)$, which has half the spatio-temporal resolution of the original video. At the same time, WZ bits are sent in parallel to enable the quality enhancement of the base layer to generate $Y_L(2i - 1)$ and $Y_L(2i + 1)$. Then, temporal side information and DVC decoding are used to generate video at full frame rate and half spatial resolution in block **B**, i.e. $Y_L(2i)$. Furthermore, $Y_L(2i - 1)$, $Y_L(2i + 1)$ (or $Y_B(2i - 1)$, $Y_B(2i + 1)$) and $Y_L(2i)$ are spatially interpolated and used as side information to generate video at the full spatio-temporal resolution in blocks **C** and **D**. Note that the video can be decoded at half its original temporal resolution and full spatial resolution if blocks **B** and **D** are skipped. This scheme can be easily extended to n layers, since it is a combination of the previous ones.
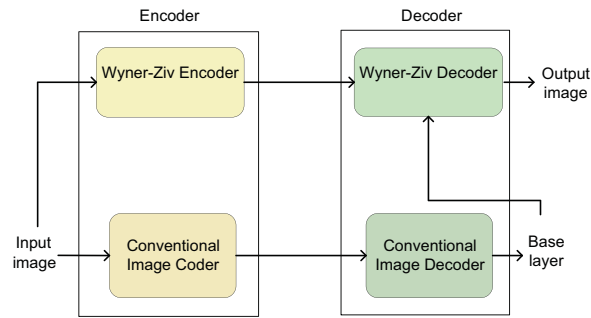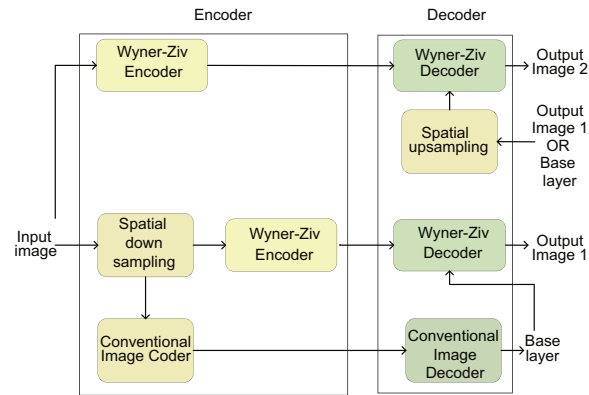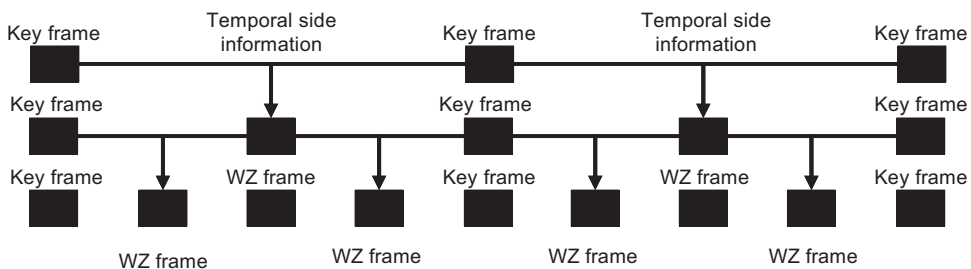

Fig. 7. Temporal scalability for two enhancement layers ($GOP = 4$).
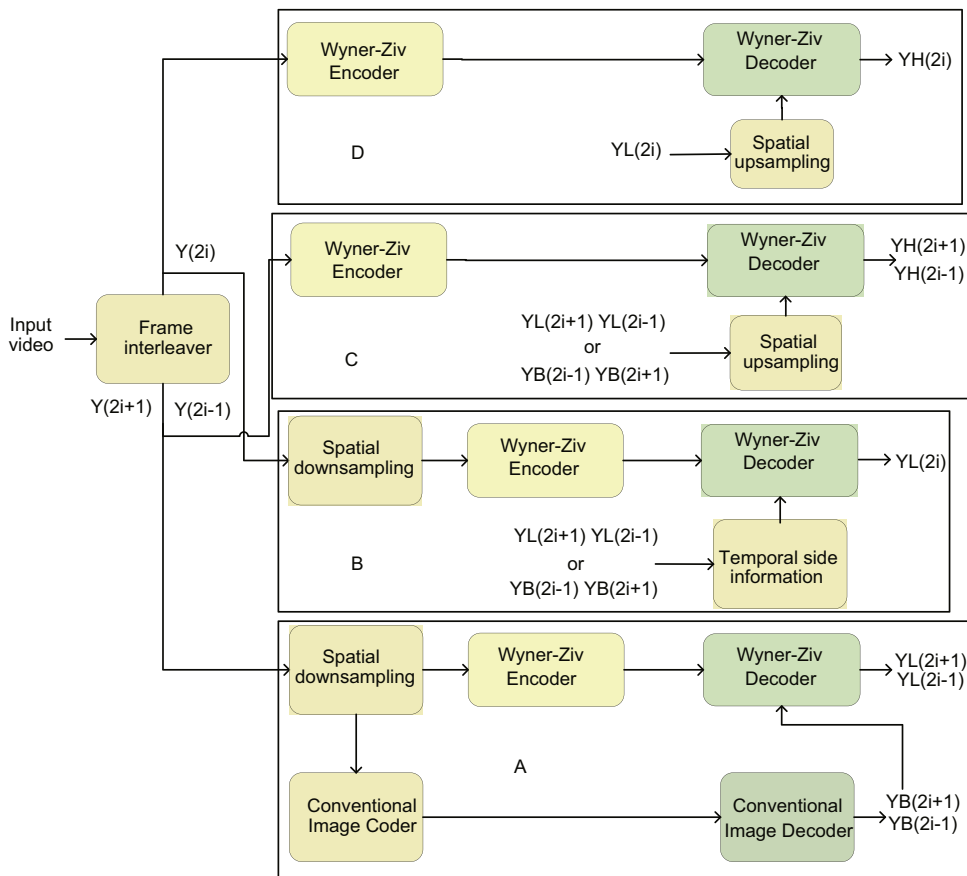
Fig. 8. DVC scheme for temporal, spatial and quality scalability.

4. JPEG2000

The JPEG2000 standard makes use of the DWT. JPEG2000 supports some important features such as improved compression efficiency, lossless and lossy compression, multi-resolution representation, Region Of Interest (ROI) coding, error resilience and a flexible file format. Fig. 9 depicts the JPEG2000 fundamental building blocks.

In JPEG2000, arithmetic coding is performed for each code-block in three coding passes, the significance propagation pass, the magnitude refinement pass and the cleanup pass. During the significance propagation pass, a bit is coded if its location is not significant but at least one of its eight connect neighbors is significant. The second pass is the magnitude refinement pass, where all bits that became significant in a previous bitplane are coded. The magnitude refinement pass includes the bits from coefficients that are already significant. Finally, the cleanup pass encodes all bits, which were not encoded during the previous passes. For more details on the JPEG2000 standard, refer to [4].

To protect the JPEG2000 stream against transmission errors, the bitstream is protected with RS [14] codes. Since packet losses are simulated, it is well known that Forward Error Correcting (FEC) combined with interleav-

ing provides more robustness against packet losses. The interleaving prevents from loosing important chunks of the bitstream at the same position but rather bursts the errors over the whole bitstream. The interleaving technique [13] is depicted in Fig. 10, where each ith word of the interleaved bitstream is constructed from the ith byte of each word from the original bitstream.

Furthermore, the error resilience tools of JPEG2000 are switched on. A more thorough description of these tools is given in [25] and [26], whereas detailed performance evaluations are presented in [25,27,28]. Note that, although JPEG2000 defines error resilient tools, the procedure that the decoder shall follow in order to cope with the possible presence of errors is not standardized.

JPEG2000 is relying on both resynchronization markers and data partitioning to limit the impact of transmission errors. More specifically, the codestream is composed of packets, with each packet corresponding to a quality layer, a resolution, a component and a precinct. As these packets constitute independently coded units, this data partitioning limits the spread of transmission errors to a great extent. In addition, Start Of Packet (SOP) resynchronization markers can be optionally inserted in front of every packet, as illustrated in Fig. 11. These markers enable the decoder to resynchronize in the presence of errors. Moreover, the quantized wavelet coefficients are
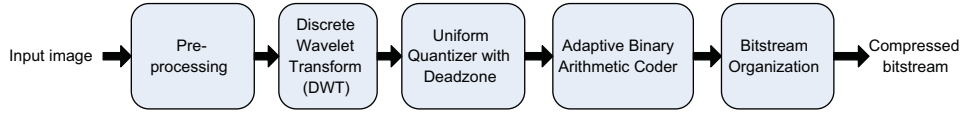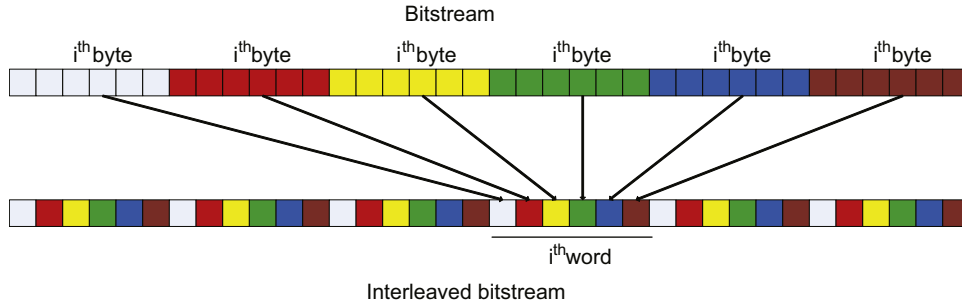
Fig. 9. The JPEG2000 fundamental building blocks.
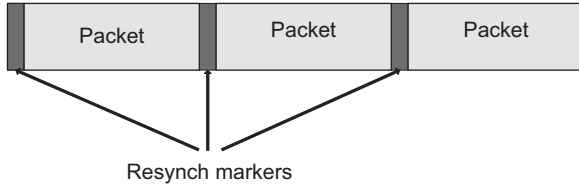

Fig. 10. JPEG2000 codestream interleaver.

Fig. 11. SOP resynchronization markers in JPEG2000.

partitioned into code-blocks, each code-block being independently coded using an MQ arithmetic coder. A number of options can be used to strengthen the resilience of the arithmetic coder. First, the arithmetic coder can be required to use a predictable termination procedure at the end of each coding pass. In addition, the arithmetic coder can be restarted at the beginning of each coding pass. Finally, a segmentation symbol can be encoded at the end of each bitplane. In this case, if the segmentation symbol is not correctly decoded at the decoder side, an error is flagged in the preceding coding passes. As a direct consequence of these mechanisms, shorter coding passes will entail better error resilience. Therefore, small code-blocks tend to lead to better performance in the presence of errors, unlike the error-free case. While the above tools detect where errors occur, conceal the erroneous data, and resynchronize the decoder, they do not correct transmission errors. Furthermore, these tools do not apply to the image header despite the fact that it is the most important part of the codestream.

5. Simulation results

5.1. Test material and conditions

The test material consists of the video sequences *Foreman* and *Soccer* with CIF resolution at 15 fps in addition to two monochromatic images *P04* and *P07* with a spatial resolution of 2256×1504. The following codecs are used to run the simulations:

- For the AVC/H.264 Intra coding, the publicly available reference software (JM 11.0) [29] is used with the following settings:
  - High profile encoding.
  - CABAC for high profile.
  - The 8×8 transform enabled for high profile in addition to the original 4×4 transform in AVC/H.264. The encoder chooses adaptively between both transforms. In uniform regions, the 8×8 transform is used while the 4×4 transform is used in highly textured parts of the video.
  - Disable transform coefficients thresholding. This is enabled by default in the software and might lead to a loss in Rate Distortion (RD) performance.
  - Enable the use of explicit lambda parameters and set the weight of the I slice to 0.5. The default value in the software assumes the presence of P and B frames, which is not the case in this work.
  - AdaptiveRounding is enabled. This parameter is used in the quantization process to adjust the rounding offset to maintain an equal expected value for the input and output of the quantization process for the absolute value of the quantized data. It is recommended to use AdaptiveRounding when encoding with high quality.
  - Enable RD optimization for the mode selection.
- For SVC, the publicly available reference software (JSVM 5.1) [30] is used with the following settings:
  - Three quality layers.
  - Two spatial layers (QCIF and CIF).
  - Two temporal layers (7.5 and 15 fps).
- KAKADU version 5.2 [31] is used for JPEG2000 with the following settings:
  - Codeblock size of 64×64.
  - Three decomposition levels.

○ Visual frequency weighting is switched off. This parameter is used to give good visual appearance. On the other hand, it reduces the RD performance.

○ Target rate control is switched on. It is used when transcoding the AVC/H.264 stream to JPEG2000 such that both streams have a similar rate.

○ Cuse_sop is set true. Use SOP markers in front of each packet at the encoder.

○ Cmodes set to RESTART|ERTERM|SEGMARK. Enable the different error resilience features of JPEG2000 at the encoder.

○ Use the option—resilient_sop at the decoder so that it recovers from and/or conceals errors to the best of its ability.

● The transform-domain DVC codec proposed in [17] is used. For temporal scalability, a *GOP* of 2 is used and the QP is chosen such that the decoded key and WZ frames have a similar quality.

Initially, AVC/H.264 Intra is used to encode the base layer. Then, the required rate to decode the different WZ layers is estimated in the error-free case using the feedback channel. Further, the base layer is decoded and re-encoded using JPEG2000. The WZ layers are used to enhance the JPEG2000 encoded base layer. This is performed to show the codec-independent scalability feature of the DVC-based schemes. In the error-prone case, the feedback channel is switched off and the rates computed in the error-free case are maintained, which means that a prefect encoder rate control is assumed. In other words, we assume the availability of a DVC scheme with an encoder rate control, which is as efficient as the one with a decoder rate control. If the bitplane does not converge during the decoding process (i.e. the error probability of the decoded bitplane does not reach a value below 10^{-3}) with the available WZ bits, it cannot use the feedback channel to request additional WZ bits and just uses the bitplane from the SI.

The packet losses are introduced into the base layer and WZ streams with 11 and 51 different packet error patterns [32] for video and image, respectively. The final result is taken as the median of all trials per rate. Moreover, Packet Loss Rates (PLRs) of 5%, 10% and 20% are considered with a packet size of 256 bytes.

Finally, different possible predictors are possible as SI in the presented DVC schemes. In our simulations, the base layer is directly used as SI in the quality and spatial scalability schemes.

5.1.1. Video

For video, the base layer is encoded using AVC/H.264 Intra at different Quantization Parameter (QP) values as lower QP means higher rate and quality. Then, the DVC decoder is run with the corresponding base layer in the error-free case to define the minimum amount of WZ bits for successful decoding. The total rate is computed as the sum of the base layer and WZ rates.

At the encoder, the WZ bits are generated for the quantized DCT coefficients. The quantization is performed

using the matrix QI_8

$$QI_8 = \begin{pmatrix} 128 & 64 & 32 & 16 \\ 64 & 32 & 16 & 8 \\ 32 & 16 & 8 & 4 \\ 16 & 8 & 4 & 0 \end{pmatrix}$$

Three RD points are computed per plot. The first RD point corresponds to the base layer and first enhancement layer bits. The second RD point corresponds to the previous RD point bits and the second enhancement layer, and so on. The scalability is achieved by initially decoding the first four bands to obtain the first quality layer, which corresponds to the first RD point. Then, the second layer corresponds to decoding four additional bands (i.e. a total of eight bands). This gives the second RD point. Finally, the third layer corresponds to decoding all the bands to obtain the third RD point. The performance of SDVC for video coding is compared to SVC with three quality layers, two spatial layers and two temporal layers. Moreover, the influence of the base layer quality on the performance of the DVC schemes in error-free and error-prone conditions is studied. This is only performed for the video schemes as the behavior is similar for images since it is a special case of video, where the number of frames is equal to 1.

5.1.2. Image

The images undergo the same DVC encoding/decoding scenario except that the base layer is encoded using non-scalable JPEG2000 and the comparison is made with respect to scalable JPEG2000 in terms of compression efficiency. In error-prone conditions, the JPEG2000 code-stream is interleaved and then protected by RS codes. Three protection modes, Weak Protection (WP) (i.e. RS(255,179), around 30% overhead of parity bits), Average Protection (AP) (i.e. RS(255,153), around 40% overhead of parity bits) and Strong Protection (SP) (i.e. RS(255,113), around 56% overhead of parity bits), are considered. The stronger the RS code, the better it recovers from errors with the cost of a larger parity bits overhead.

5.2. Error-free conditions

5.2.1. Video

First, simulation results are presented to illustrate the idea of codec-independent scalability. More specifically, the DVC scalable schemes are run with an AVC Intra coded base layer requesting a certain amount of parity bits. This same amount is used to enhance a JPEG2000 coded base layer, which is generated by transcoding the AVC base layer to JPEG2000. More specifically, the base layer is decoded with an AVC decoder and then encoded using a JPEG2000 encoder.

Fig. 12 depicts the RD performance of the SDVC schemes for *Foreman* and *Soccer*. It is obvious that the performance drops as the number of scalability layers increases. This results in the quality scheme having the best performance and the one ensuring all three scalabilities having the worst performance. Further, it shows that scalability is enabled by DVC for different base layers by
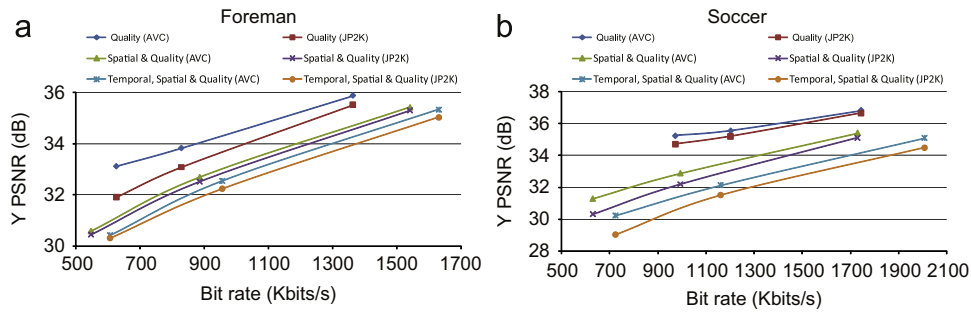
Fig. 12. The DVC scalable schemes with different base layers and same enhancement layer for *Foreman* and *Soccer*. AVC and JP2K stand, respectively, for AVC and JPEG2000 encoded base layers. (a) *Foreman*. (b) *Soccer*.
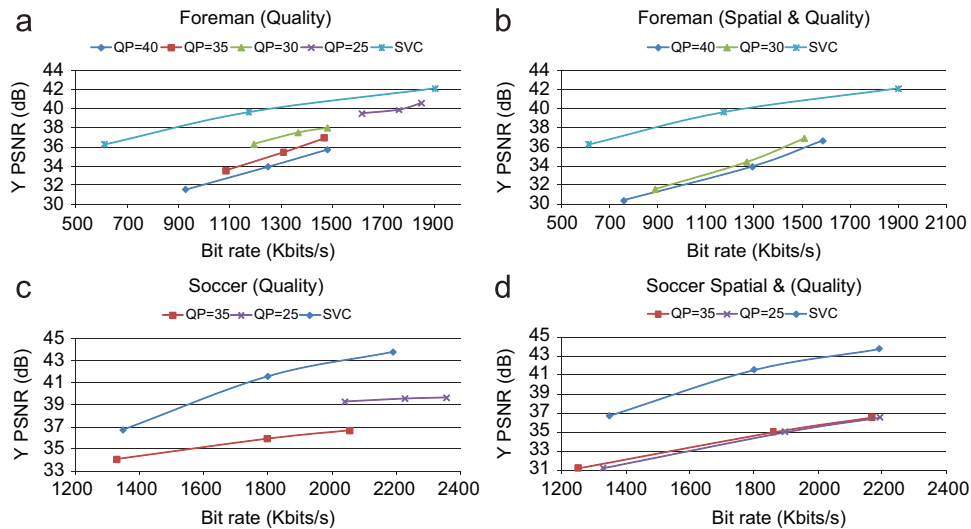
Fig. 13. The performance of the scalable DVC video coding schemes for different base layer qualities in the error-free case. QP is the quantization parameter of AVC/H.264 Intra used to encode the base layer. (a) Foreman (quality scalability). (b) Foreman (spatial quality scalability). (c) Soccer (quality scalability). (d) Soccer (spatial quality scalability).

Table 1
Base and enhancement layers bit rates for *Foreman* for quality scalability.

AVC QP	Base layer (Kbps)	1st layer (Kbps)	2nd layer (Kbps)	3rd layer (Kbps)
40	255.26	703.29	322.46	232.84
35	471.76	614.35	225.5	159.72
30	843.64	351.5	173.78	115.4
25	1451.62	163.62	146.01	89.32

Table 2
Base and enhancement layers bit rates for *Foreman* for quality and spatial scalability.

AVC QP	Base layer (Kbps)	1st layer (Kbps)	2nd layer (Kbps)	3rd layer (Kbps)
40	85.12	675.98	533.53	292.44
30	252.36	639.18	380.84	239.7

Table 3
Base and enhancement layers bit rates for *Soccer* for quality scalability.

AVC QP	Base layer (Kbps)	1st layer (Kbps)	2nd layer (Kbps)	3rd layer (Kbps)
35	430.43	897.51	470.32	256.91
25	1685.42	354.02	187.3	129.22

the same enhancement layer (i.e. codec-independent scalability).

Hereafter, the base layer quality is varied for the spatial and quality scalability schemes for video as shown in Fig. 13. Moreover, Tables 1–4 contain the bit rates for the different base layers and the corresponding enhancement layers for *Foreman* and *Soccer*. As the amount of base layer bits is increased (i.e. better SI quality), fewer WZ bits are used and therefore the better the performance. Moreover, the spatial scalability scheme is outperformed by the

quality scalability scheme. The reason is that the SI quality for the quality scalability scheme increases rapidly with the base layer rate increase. On the other hand, the SI

quality for the spatial scalability scheme saturates rapidly due to the spatial interpolation. It is obvious that SDVC lacks compression efficiency when compared to SVC, which is expected as DVC is known for being inferior to conventional coding for the moment.

5.2.2. Image

The codec-independent scalability feature is shown for images P04 and P07 as illustrated in Fig. 14.

Fig. 15 shows a comparison between the proposed scalable schemes for image coding and scalable JPEG2000. Moreover, the invested bits in the base and enhancement layers are depicted in Tables 5 and 6.

With no errors (i.e. 0% PLR), JPEG2000 should be superior to the DVC-based schemes if the RS parity bits are omitted as they are useless in this case. On the other hand, if the parity bits overhead is large enough, the

performance of JPEG2000 can be inferior to SDVC as illustrated in Figs. 19 and 20.

5.3. Error-prone conditions

5.3.1. Video

Next, the influence of base layer bit allocation is studied for the proposed SDVC schemes for video in error-prone conditions. It is observed that the slope at which the performance decreases with PLRs is greater for

Table 5
Base and enhancement layers bits for P04.

	Base layer (bpp)	1st layer (bpp)	2nd layer (bpp)	3rd layer (bpp)
Quality	0.42	0.35	0.19	0.18
Spatial and quality	0.15	0.68	0.23	0.2

Table 6
Base and enhancement layers bits for P07.

	Base layer (bpp)	1st layer (bpp)	2nd layer (bpp)	3rd layer (bpp)
Quality	0.4	0.38	0.26	0.19
Spatial and quality	0.15	0.64	0.3	0.21

Table 4
Base and enhancement layers bit rates for *Soccer* for quality and spatial scalability.

| AVC QP	Base layer (Kbps)	1st layer (Kbps)	2nd layer (Kbps)	3rd layer (Kbps)
35	117.83	1135.35	607.3	306.86
25	405.08	924.9	566.81	297.16
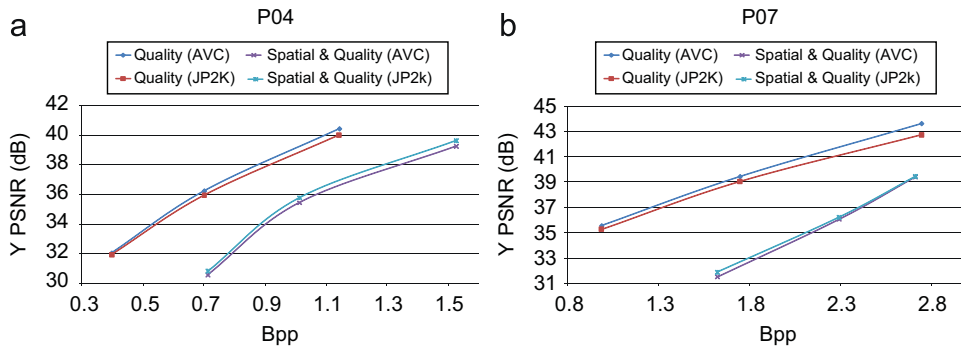
Fig. 14. The DVC scalable schemes with different base layers and same enhancement layer for P04 and P07. AVC and JP2K stand, respectively, for AVC and JPEG2000 encoded base layers.
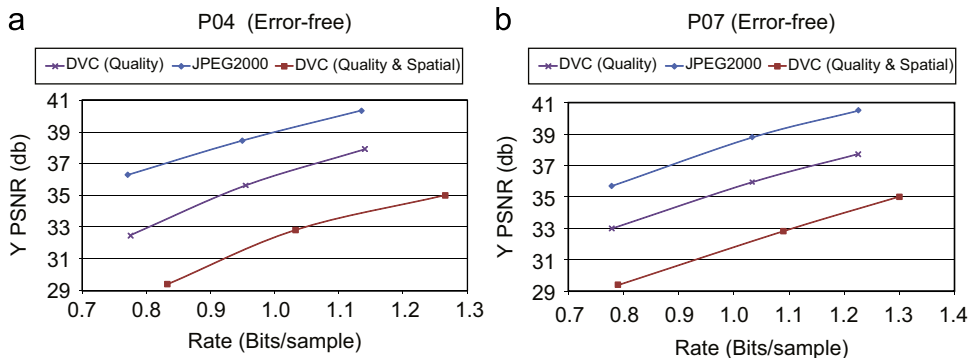
Fig. 15. Comparison of the DVC scalable schemes with scalable JPEG2000 for P04 and P07. (a) P04. (b) P07.
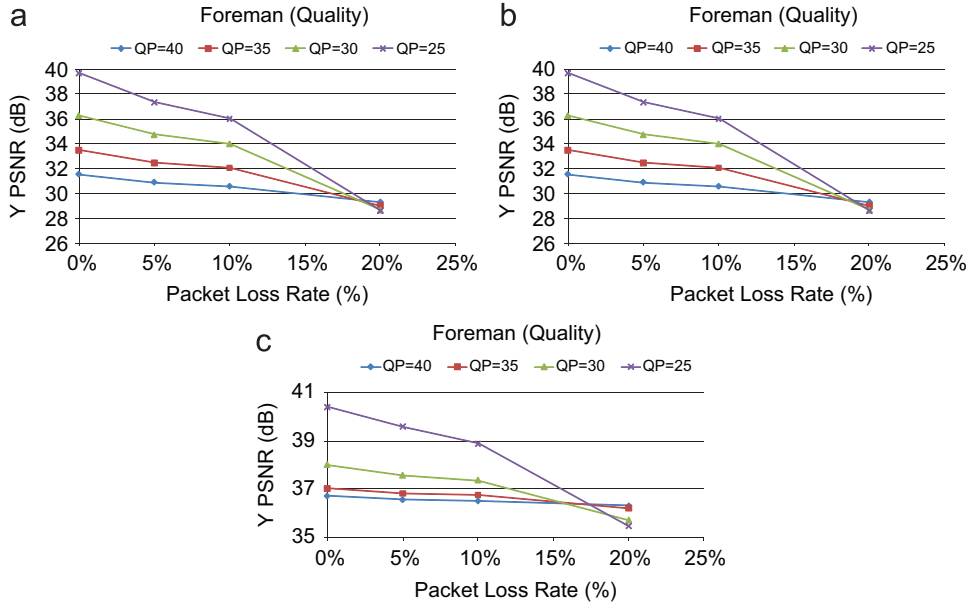
Fig. 16. The performance of the DVC quality video coding scheme for different base layer qualities. QP is the quantization parameter of AVC/H.264 Intra used to encode the base layer. (a) Foreman after first quality layer. (b) Foreman after second quality layer. (c) Foreman after third quality layer.
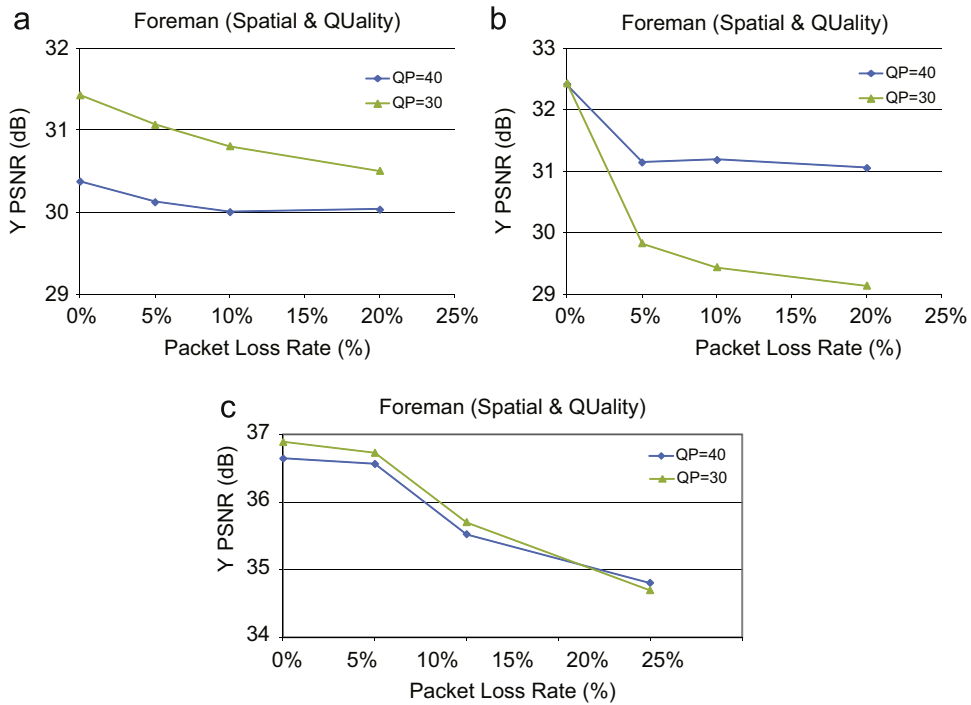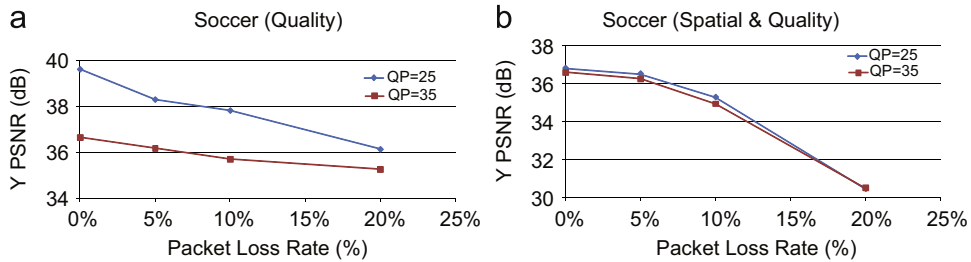

Fig. 17. The performance of the DVC spatial and quality video coding scheme for different base layer qualities. QP is the quantization parameter of AVC/H.264 Intra used to encode the base layer. (a) Foreman after first quality layer. (b) Foreman after second quality layer. (c) Foreman after third quality layer.

higher base layer quality as illustrated in Figs. 16, 17 and 18. This is explained by the fact that higher base layer quality corresponds to fewer parity bits. Per consequent, the scheme shows less resistance to the increase in PLR.

5.3.2. Image

The scalability DVC schemes for image coding are studied hereafter in the error-prone conditions by comparing them to scalable JPEG2000 protected with RS codes. The rate for JPEG2000 is computed differently in

Fig. 18. The performance of the DVC scalable video coding schemes for different base layer qualities. QP is the quantization parameter of AVC/H.264 Intra used to encode the base layer. (a) Soccer after third quality layer. (b) Soccer after third quality layer.
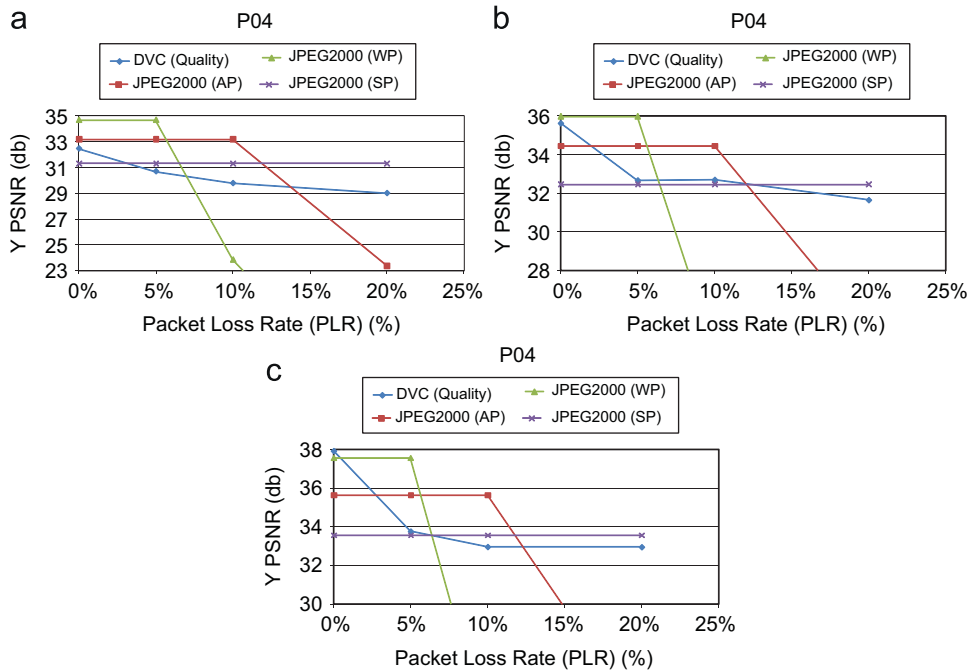
Fig. 19. Comparison of the DVC quality image coding scheme with RS-protected JPEG2000 for different PLRs at different rates for *P*04. WP, AP and SP stand for weak, average and strong RS protection, respectively. (a) P04 after first quality layer. (b) P04 after second quality layer. (c) P04 after third quality layer.

the error-prone case as the overall rate has to account for the RS parity bits. More specifically, if a RS code (n, k) (i.e. $(n - k)$ length parity bits are generated for each k length data) is used and a rate R is targeted, the rate control of JPEG2000 is set to achieve a rate k/nR. Thus, the RS parity bits rate is $(n - k)/nR$.

It is depicted in Fig. 19 the comparison between the DVC quality scheme and the scalable JPEG2000 with three layers (i.e. three RD points) when packet losses are simulated for *P*04.

First, it is observed that for JPEG2000 WP and AP, the performance tends to rapidly drop (i.e. cliff effect) after a certain PLR. This is not the case for DVC as it decreases steadily with error rate increase. More specifically, JPEG2000 WP and AP tend to drop rapidly at 10% and 20% PLR, respectively, as the error correcting capability limit of the RS code is reached. However, the RS code seems to correct all the errors at all PLRs for strong RS protection with the cost of having a huge parity bits

overhead. Similar behavior is observed for the spatial and quality scheme as shown in Fig. 20. The difference in the behavior between the DVC-based schemes and the RS-protected JPEG2000 can be explained as well by the nature of the protection in each case. More specifically, RS bits are applied to protect the JPEG2000 bitstream after the compression process. On the other hand, the WZ bits are part of the compression process of DVC as they protect the DCT coefficient of the WZ frames. Even though the SI information is corrupted further due to the simulated errors, some frequencies are preserved and might be corrected by the WZ bits.

Next, samples of the decoded images using RS-protected JPEG2000 and the scalable schemes are illustrated in Fig. 21 for a PLR of 10%. The DVC-based schemes produce images with fewer artifacts. The spatial-quality scheme performs much better than RS-protected JPEG2000 as the latter produces a highly blurred image. For the quality scheme, the difference is less perceptible.
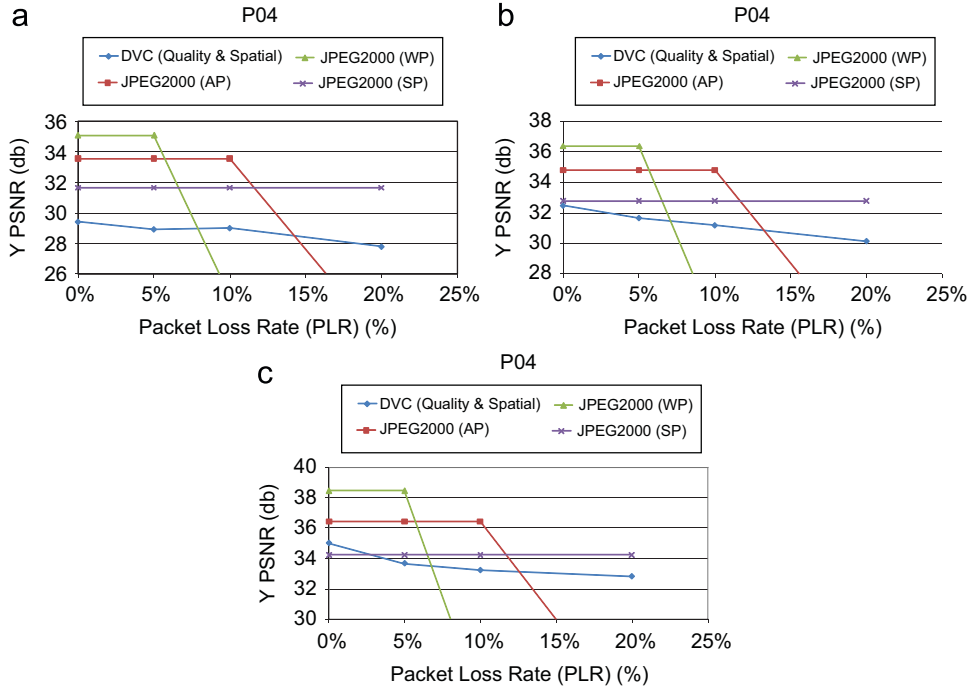
Fig. 20. Comparison of the DVC spatial and quality image coding scheme with RS-protected JPEG2000 for different PLRs at different rates for *P*04. WP, AP and SP stand for weak, average and strong RS protection, respectively. (a) P04 after first quality layer. (b) P04 after second quality layer. (c) P04 after third quality layer.
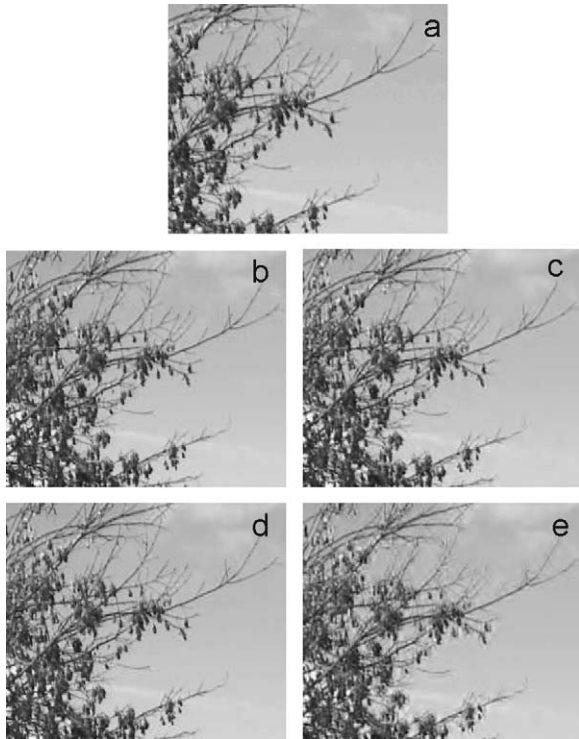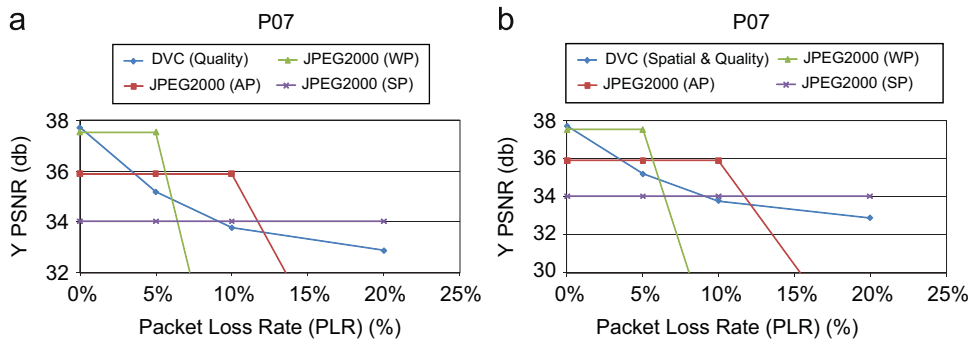

Fig. 21. Comparison of the visual quality of the decoded image for RS-protected scalable JPEG2000 and the scalable DVC schemes at 10% PLR for *P*04. (a) Original. (b) Quality DVC. (c) Scalable JPEG2000 with 3 quality layers. (d) Spatial and quality DVC. (e) Scalable JPEG2000 with 3 quality and 2 spatial layers.

Nevertheless, the DVC-based image has sharper edges and more contrast than the JPEG2000 image.

Figs. 22 and 23 illustrate, respectively, the objective quality and decoded image samples for *P*07 at different PLRs comparing the DVC schemes and RS-protected JPEG2000. The image produced by the JPEG2000 decoder is highly blurred while the one by DVC has more contrast but with damaged edges.

6. Conclusion

In this research, SDVC schemes for image and video coding are introduced. They use conventional coding to generate the base layer whereas the enhancement layers are represented in the WZ bits. The simulation results show that the proposed schemes are codec-independent. The latter means that base and enhancement layer are completely independent. This offers high flexibility in the way the stream is distributed over the network and helps reducing the traffic at the same time in certain scenarios. In the presence of transmission errors due to packet loss, the DVC schemes show more resistance to error rate increase when compared to conventional coding. More specifically, SDVC image coding shows better resistance to the increase in PLR as its performance decreases steadily. On the other, scalable JPEG2000 protected with RS codes for image coding exhibits a cliff effect as its performance drops significantly after a certain limit. Nevertheless, an ideal encoder rate control mechanism is assumed in this work. In other words, the availability of a rate encoder

Fig. 22. Comparison of the DVC scalable coding schemes with RS-protected JPEG2000 for different PLRs at different rates for *P*07. WP, AP and SP stand for weak, average and strong RS protection, respectively. (a) P07 after third quality layer. (b) P07 after third quality layer.
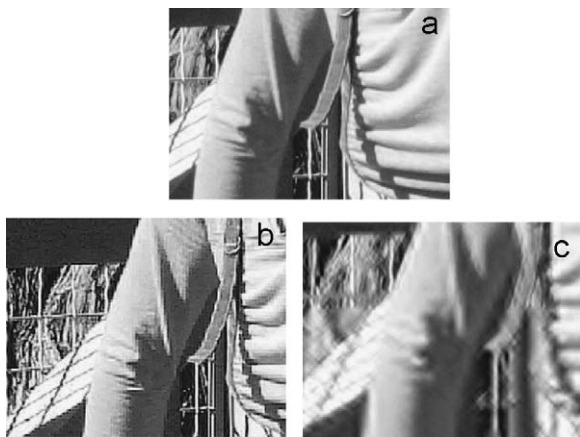
Fig. 23. Comparison of the visual quality of the decoded image for RS-protected scalable JPEG2000 and the scalable DVC schemes at 10% PLR for *P*07. (a) Original. (b) Quality DVC. (c) Scalable JPEG2000 with 3 quality layers.

mechanism as efficient as the one with a feedback channel is assumed.

Future work can be pursued into two directions. First, evaluating the same schemes when maintaining the feedback channel for the DVC-based schemes and compareing it to the case where it is switched off. Second, incorporate a rate control mechanism at the encoder and perform similar simulations. A poorer performance is expected for the DVC schemes in the error-free case. On the other hand, we would expect a better performance in the error-prone case since the rate will be overestimated when compared to the feedback channel case. In other words, more parity bits are sent toward the decoder. Finally, no a priori knowledge of the network's PLR is assumed in this work. Thus, future work can consider scenarios, where this information is available and exploited to adjust the RS protection for JPEG2000.

Acknowledgment

References

[1] H. Schwarz, D. Marpe, T. Wiegand, Overview of the scalable video coding extension of the H.264/AVC standard, IEEE Transactions of Circuits and Systems for Video Technology 17 (9) (2007).

[2] Scalable video coding ⟨http://www.chiariglione.org/mpeg/technologies/mp04-svc⟩.

[3] T. Wiegand, G.J. Sullivan, G. Bjøntegaard, A. Luthra, Overview of the H.264/AVC video coding standard, IEEE Transactions on Circuits and Systems for Video Technology 13 (7) (2003) 560–576.

[4] A. Skodras, C. Christopoulos, T. Ebrahimi, The JPEG2000 still image compression standard, IEEE Signal Processing Magazine 18 (5) (2001) 36–58.

[5] B. Girod, A. Aaron, S. Rane, D. Rebollo-Monedero, Distributed video coding, in: Proceedings of the IEEE, Special Issue on Video Coding and Delivery, vol. 93, January 2005, pp. 71–83.

[6] J. Pedro, L. Soares, C. Brites, J. Ascenso, F. Pereira, C. Bandeirinha, S. Ye, F. Dufaux, T. Ebrahimi, Studying error resilience performance for a feedback channel based transform domain Wyner–Ziv video codec, in: Picture Coding Symposium (PCS), Lisbon, Protugal, 2007.

[7] S. Ye, M. Ouaret, F. Dufaux, T. Ebrahimi, Hybrid spatial and temporal error concealment for distributed video coding, in: IEEE (Ed.), International Conference on Multimedia and Expo (ICME), Hannover, Germany, 2008.

[8] M. Tagliasacchi, A. Majumdar, K. Ramchandran, A distributed source coding based spatio-temporal scalable video codec, in: Picture Coding Symposium (PCS), San Francisco, USA, 2004.

[9] R. Puri, K. Ramchandran, PRISM: a new robust video coding architecture based on distributed compression principles, in: Allerton Conference on Communication, Control and Computing, 2002.

[10] A. Sehgal, A. Jagmohan, N. Ahuja, Scalable video coding using Wyner–Ziv codes, in: Picture Coding Symposium (PCS), San Francisco, USA, 2004.

[11] H. Wang, N.-M. Cheung, A. Ortega, A framework for adaptive scalable video coding using Wyner–Ziv techniques, EURASIP Journal on Applied Signal Processing 2006 (2008) 267.

[12] T. Ebrahimi, F. Pereira, The MPEG-4 Book, Prentice-Hall, New York, 2002.

[13] F. Dufaux, G. Baruffa, F. Frescura, D. Nicholson., JPWL—an extension of JPEG 2000 for wireless imaging, in: International Symposium on Circuits and Systems, Island of Kos, Greece, 2006.

[14] I.S.R.G. Solomon, Polynomial codes over certain finite fields, Journal of the Society for Industrial and Applied Mathematics 8 (2) (1960) 300–304.

[15] J. Slepian, J. Wolf, Noiseless coding of correlated information sources, IEEE Transactions on Information Theory 19 (4) (1973) 471–480.

[16] A. Wyner, J. Ziv, The rate-distortion function for source coding with side information at the decoder, IEEE Transactions on Information Theory 22 (1) (1976) 1–10.

[17] C. Brites, J. Ascenso, F. Pereira, Improving transform domain Wyner–Ziv video coding performance, in: International Conference on Acoustics, Speech and Signal Processing, Toulouse, France, 2006.

[18] H. Malvar, A. Hallapuro, M. Karczewicz, L. Kerofsky, Low-complexity transform and quantisation in H.264/AVC, IEEE Transactions on Circuits and Systems for Video Technology 12 (7) (2003) 598–603.

[19] C. Berrou, A. Glavieux, P. Thitimajshima, Near Shannon limit error-correcting coding and decoding: turbo codes, in: International Conference on Communications, Geneva, Switzerland, 1993.

[20] W.W. Peterson, D.T. Brown, Cyclic codes for error detection, in: Proceedings of the IRE, vol. 49, 1961, pp. 228–235.

[21] J. Ascenso, C. Brites, F. Pereira, Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding, in: EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services, Slovak, July 2005.

[22] A. Aaron, R. Zhang, B. Girod, Wyner–Ziv coding for motion video, in: Asilomar Conference on Signals, Systems and Computers, Pacific Grove, USA, 2002.

[23] M. Ouaret, F. Dufaux, T. Ebrahimi, Codec-independent scalable distributed video coding, in: International Conference on Image Processing (ICIP), San Antonio, USA, 2007.

[24] C. Tonoli, M. Dalai, P. Migliorati, R. Leonardi, Error resilience performance evaluation of a distributed video codec, in: Picture Coding Symposium (PCS), Lisbon, Portugal, 2007.

[25] I. Moccagatta, S. Soudagar, J. Liang, H. Chen, Error-resilient coding in JPEG-2000 and MPEG-4, IEEE Journal on Selected Areas in Communications 18 (6) (2000) 899–914.

[26] D. Taubman, M. Marcellin, JPEG 2000: Image Compression Fundamentals, Standards and Practice, Kluwer Academic Publishers, 2002.

[27] D. Santa-Cruz, R. Grosbois, T. Ebrahimi, JPEG2000 performance evaluation and assessment, Signal Processing: Image Communication 17 (1) (2002) 113–130.

[28] A. Bilgin, Z. Wu, M. Marcellin, Decompression of corrupt JPEG2000 codestreams, in: Data Compression Conference, Snowbird, Utah, USA, 2003.

[29] AVC/H.264 Software ⟨http://iphome.hhi.de/suehring/tml/⟩.

[30] Joint scalable verification model (JSVM) ⟨http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.htm⟩.

[31] JPEG2000 Kakadu Software ⟨http://www.kakadusoftware.com/⟩.

[32] S. Wenger, Proposed error patterns for Internet experiments, Technical Report Doc. VCEG Q15-I-16R1, VCEG Meeting, Red Bank, NJ, USA, October 1999.