Towards the Future Internet
G. Tselentis et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-539-6-205

Scalable Video Coding: Source for Future Media Internet

Naeem RAMZAN, Toni ZGALJIC, and Ebroul IZQUIERDO
School of Electronic Engineering and Computer Science,
Queen Mary University of London

Abstract. A flexible wavelet-based scalable video coding framework (W-SVC) is proposed to support future media internet, specifically content delivery to different display terminals through heterogeneous networks as the Future Internet. Scalable video bit-stream can easily be adapted to required spatio-temporal resolution and quality, according to the transmission and user context requirements. This enables content adaptation and interoperability in Internet networking environment. Adaptation of the bit-stream is performed in the compressed domain, by discarding the bit-stream portions that represent higher spatio-temporal resolution and/or quality than the desired. Thus, the adaptation is of very low complexity. Furthermore, the embedded structure of a scalable bit-stream provides a natural solution for protection of the video against transmission errors inherent to content transmission over Internet. The practical capabilities of the W-SVC are demonstrated by using the error resilient transmission and surveillance applications. The experimental result shows that the W-SVC framework provides effusive flexible architecture with respect to different application in future media internet.

Keywords. Scalable video coding, wavelet-based video coding, transcoding, unequal error protection, turbo codes, surveillance centric coding.

Introduction

In recent years, there have been increasing developments in technologies for transmission of and access to multimedia content over Internet. When video is delivered to the user it usually needs to traverse network paths with very different traffic capacities: from very high bandwidth on dedicated glass fibre connections to very low bit-rate connectivity for wireless transmissions. Furthermore, the same content needs to be accessible from a variety of devices at the user side, which have different displaying and computational capabilities [1]. To tackle this challenge, the compression technology used in the transmission system should ideally provide two important features; first, it has to be highly efficient in terms of compression and second, it has to provide flexible, low-complexity, real-time content adaptation to the network and user's device properties.

Conventional compression technologies can provide solutions for video adaptation based either on transcoding [2] or storing and transmitting different instances of the same content. In the first approach the content is adapted by decoding and re-encoding using compression settings tailored to the user context parameters, such as available bandwidth or user-device capability (due to Internet diversity). This process is, however, computationally expensive and therefore may not meet the requirement for low-complexity adaptation. The second approach is based on storing and transmitting multiple streams encoded with different parameters, where each stream represents an instance of the same content. Then the stream that most closely matches the user context parameters is decoded at the user's side. Here, low computational complexity of adaptation is ensured, but this approach generally provides inefficient solution in terms of trade-off between space / bandwidth required to store / transmit multiple streams and the quality of decoded content [3].

Scalable Video Coding (SVC) [1]-[5] is a relatively recent approach to video coding [6]-[9], which enables encoding of video content to produce a single scalable bit-stream that can be seamlessly adapted to network or terminal properties while providing high compression efficiency. A bit-stream is scalable if it is composed of a hierarchically embedded family of content instances at different resolution levels. The underlying multiresolution family of content instances should be embedded in the sense that when looking at two different instances of the same content at two different resolution levels, the two representations must match exactly over the extent of the lowest one. Any instance of the content embraces all representations for lower resolutions. The representation gets more accurate as the resolution increases. This allows very low complexity of adaptation as lower resolutions can be extracted from a higher resolution directly in the compressed domain, thus without performing computationally expensive transcoding. In addition, for a given resolution the representation should be good if not optimal. For instance, just attaching a whole high-resolution version to a low-resolution one does not qualify as hierarchically embedded content.

SVC also provides a natural solution with a truncateable bit-stream for error-prone transmissions inherent to Internet. In addition, channel coding methods can be adaptively used to attach different degrees of protection to different bit-layers according to their relevance in terms of decoded video quality. Usually, Joint Source Channel Coding (JSCC) [4] applies different degrees of protection to different portions of the bit-stream. This means that Unequal Error Protection (UEP) is used according to the importance of a given portion of the bit-stream. In this context, scalable coding emerges as the natural choice for highly efficient JSCC with UEP.

The JSCC approach proposed in this chapter performs a joint optimization of a wavelet-based SVC (W-SVC) and a Forward Error Correction method (FEC) based on Turbo Codes (TC) [10]-[11] to provide a smooth delivery of video over Internet. The proposed JSCC [12]-[14] scheme minimizes the reconstructed video distortion at the decoder subject to a constraint on the overall transmission bit-rate budget. The minimization is achieved by exploiting the source Rate- Distortion (RD) characteristics and the statistics of the available codes. Here, the critical problem of estimating the Bit Error Rate (BER) probability in error prone applications is also discussed. Regarding the error rate statistics, not only the channel coding rate, but also the interleaver and packet size for TCs are considered in the proposed approach. The aim is to improve the overall performance of the underlying JSCC.

In addition to transmission over Internet, SVC is apt for an event driven adaptation as a potential application for Future Internet. For instance, in surveillance applications, the scene remains essentially static for seconds and even minutes in some cases. During these periods of time nothing interesting happens from the surveillance standpoint, and the video resembles a still picture for long periods of time with no other activity than random environmental motion. An approach to reduce the bit-rate of the encoded video segments that are irrelevant from the surveillance standpoint is discussed that enables increased protection of bit-stream for seamless delivery of video over Internet. This approach combines background subtraction and W-SVC. This produces a single scalable bit-stream that contains segments of video encoded at different qualities and / or spatio temporal resolutions. The irrelevant segments are encoded using low resolution / quality while the relevant segments are encoded at high resolution / quality.

The performance evidence of W-SVC in different scenarios i.e. event driven coding and transmission is demonstrated in this chapter. This chapter starts with the introduction of W-SVC and its bit-stream organization in section 1. Section 2 explains the event driven scalable video coding. The joint scalable source and channel coding is explained in section 3. The conclusion is explained in the last section where it is envisaged that SVC is the potential candidate for the future media internet.

1. Scalable Video Coding

During the last two decades a significant amount of research has been dedicated to scalable video coding with the aim of developing the technology that would provide a low-complexity video adaptation, but retain the comparable compression efficiency and decoding complexity to those of conventional (non-scalable) video coding systems. This research evolved from two main branches of conventional video coding: 3D wavelet [5] and hybrid video coding [6] techniques. Although some of the earlier video standards, such as H.262 / MPEG-2 [7], H.263+ and MPEG-4 Part 2 [8] included limited support for scalability, the use of scalability in these solutions came at the significant increase in the decoder complexity and / or loss in coding efficiency. The latest video coding standard, H.264 / MPEG-4 AVC [6] provides a fully scalable extension, SVC, which achieves significant compression gain and complexity reduction when scalability is sought, compared to the previous video coding standards.

Although a hybrid based technology was chosen for standardisation within MPEG, a great amount of research continued also on Wavelet-based Scalable Video Coding (W-SVC). Several recent W-SVC systems (e.g. [5]) have shown a very good performance in different types of application scenarios, especially when fine granular quality scalability is required. In the proposed approach a W-SVC [5] framework is used for source coding and its brief overview is given in the remainder of this section.

The scalability is usually required in three different directions (and their combinations). We define these directions of scalability as follows:

- Temporal scalability refers to the possibility of reducing the temporal resolution of encoded video directly from the compressed bit-stream, i.e. number of frames contained in one second of the video.
- Spatial scalability refers to the possibility of reducing the spatial resolution of the encoded video directly from the compressed bit-stream, i.e. number of pixels per spatial region in a video frame.
- Quality scalability, or commonly called SNR (Signal-to-Noise-Ratio) scalability, or fidelity scalability, refers to the possibility of reducing the quality of the encoded video. This is achieved by extraction and decoding of coarsely quantised pixels from the compressed bit-stream.

An example of basic scalabilities is illustrated in Figure 1, which shows a typical SVC encoding, extraction and decoding chain. The video is encoded at the highest spatio-temporal resolution and quality. After encoding, the video is organised into a scalable bit-stream and the associated bit-stream description is created. This description indicates positions of bit-stream portions that represent various spatio-temporal

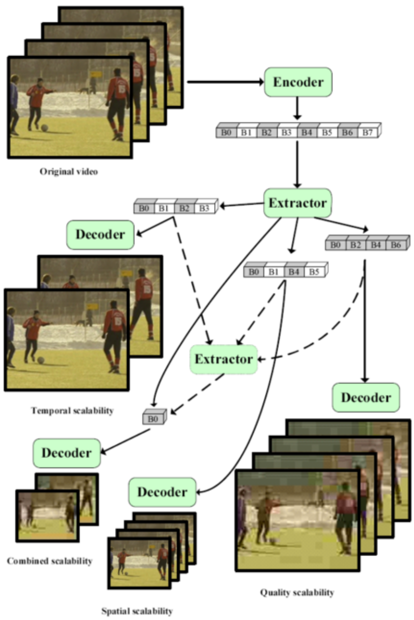


Figure 1: A typical scalable video coding chain and types of scalabilities by going to lower-rate decoding.

resolutions and qualities. The encoder is the most complex between the three modules. The compressed video is adapted to a lower spatio-temporal resolution and / or quality by the extractor. The extractor simply parses the bit-stream and decides which portions of the bit-stream to keep and which to discard, according to the input adaptation parameters. An adapted bit-stream is also scalable and thus it can be fed into the extractor again, if further adaptation is required. The extractor represents the least complex part of the chain, as its only role is to provide low-complexity content adaptation without transcoding. Finally, an adapted bit-stream is sent to the decoder, which is capable of decoding any adapted scalable video bit-stream.

1.1. High level architecture of the W-SVC framework

A typical W-SVC encoder is shown in Figure 2. First, the input video is subjected to a Spatio-Temporal (ST) decomposition, which is based on wavelet transform. The purpose of the decomposition is to decorrelate the input video content and provide the basis for spatial and temporal scalability. The ST decomposition results in two distinctive types of data: wavelet coefficients representing the texture information remaining after the wavelet transform, and motion information (obtained from Motion Estimation (ME)), which describe spatial displacements between blocks in neighbouring frames. Although generally, the wavelet transform performs very well in the task of video content decorrelation, some amount of redundancies still remains between the wavelet coefficients after the decomposition. Moreover a strong correlation also exists between motion vectors. For these reasons, further compression of the texture and motion vectors is performed.

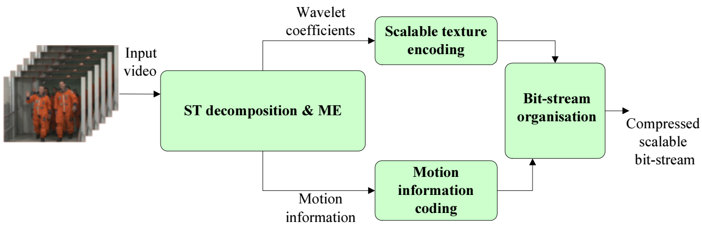


Figure 2: A high-level structure of a W-SVC encoder.

Texture coding is performed in conjunction with so-called embedded quantisation (bit-plane coding) in order to provide the basis for quality scalability. Finally, the resulting data are mapped into the scalable stream in bit-stream organisation module, which creates a layered representation of the compressed data that is explained in next subsection in detail. This layered representation provides the basis for low-complexity adaptation of the compressed bit-steam, as discussed in the introductory section of this chapter.

1.1.1. Bit-stream Organisation

In order to achieve efficient layered extraction, scalable video bit-stream consists of packets of data called atoms. An atom represents the smallest entity that can be added or removed from the bit-stream. Following such an organisation the extractor simply discards atoms from the bit-stream that are not required to obtain the video of the desired spatial resolution, temporal resolution and / or quality. Each atom can be represented by its coordinates in 3D temporal-spatial-quality space, denoted as (T, S, Q). The maximum coordinates are denoted as (T_M, S_M, Q_M), where T_M, S_M, Q_M represent the number of refinement layers in temporal, spatial and quality directions, respectively. Except for refinement layers, a basic layer in each direction exists, which is denoted as zeroth layer and cannot be removed from the bit-stream. If (m, n, i), $m \in \{0, 1, \ldots, T_M\}$, $n \in \{0, 1, \ldots, S_M\}$ and $i \in \{0, 1, \ldots, Q_M\}$, represents the desired temporal resolution, spatial resolution and quality (bit-rate), respectively, then the atoms with the coordinates $T > m$, $S > n$, $Q > i$ are discarded from the bit-stream during the extraction process. For simplicity we have assumed an equal number of atoms corresponding to each spatio-temporal resolution. Generally, this is not the case as the number of atoms corresponding to different spatio-temporal resolutions may be different. Moreover, the atoms corresponding to different qualities can be truncated at any byte, in order to achieve fine granular scalability. However, the principle of extraction remains the same.

2. Event-based Scalable Coding

The basic principle behind the event-based scalable coding is to use different compression settings for time segments representing events of different importance in a video sequence. For this purpose the temporal segments of the video sequence are classified into two types:

- temporal segments representing an essentially static scene (e.g. only random environmental motion is present – swaying trees, flags moving on the wind, etc.)

- temporal segments containing non-randomised motion activity (e.g. a vehicle is moving in a forbidden area).

To enable the above classification, background subtraction and tracking module from [15] is used as Video Content Analysis (VCA). It uses a mixture of Gaussians to separate the foreground from the background. Each pixel of a sequence is matched with each weighted Gaussian of the mixture. If the pixel value is not within 2.5 standard deviations of any Gaussians representing the background, then the pixel is declared as the foreground. Since the mixture of Gaussians is adaptive and more than one Gaussians are allowed to represent the background, this module is able to deal robustly with light changes, bimodal background like swaying trees and introduction or removal of objects from the scene. The output of the module defines parameters of compressed video, which is encoded with the W-SVC framework.

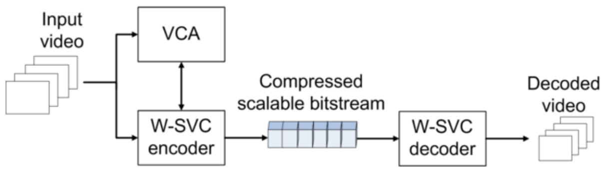
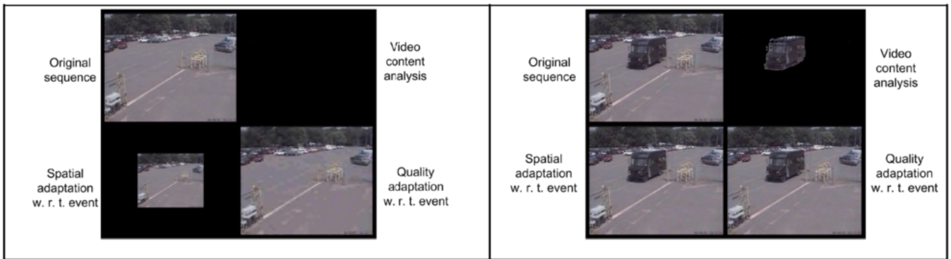


Figure 3: Event-based scalable video encoding framework.

At each time instance the W-SVC encoder communicates with the VCA module, as shown in Figure 4. When the input video is essentially static the output of the background subtraction does not contain any foreground pixels. This signal to the W-SVC encoder to adapt captured video at low spatio-temporal resolution and quality. This allows, for instance, storing and/or transmitting the portions of the video containing long, boring, static scenes using low quality frame-rate and spatial resolution. On the other hand, when some activity in the captured video is detected, the VCA module notifies the W-SVC encoder to automatically switch its output to a desired much higher spatio-temporal resolution and quality video. Therefore, decoding and use of the video at different spatio-temporal resolutions and qualities corresponding to different events is achieved from a single bit-stream, without multicasting or complex transcoding. Moreover, additional optional adaptation to lower bit-rate is also possible without re-encoding the video. This is very useful in cases where video has to be delivered to a device with a low display capability. Using this approach, the bit-rate of video portions that are of low interest is kept low while the bit-rate of important parts is kept high. Since in many realistic applications it can be expected that large portions of the captured video have no events of interest, the proposed model leads to significant reduction of resources without jeopardizing the quality of any off-line event detection module that may be present at the decoder.

Subjective results of the event-based scalable encoding module are presented in Figure 4. The top-left figure in Fig 4a and Fig 4b shows original frames. The top-right figure represents the foreground. The bottom row of Fig. 4a shows the reconstructed sequence whose essentially static segments (no event) were encoded at lower spatial resolution (bottom-left) or at lower quality (bottom-right). The bottom row of Fig 4b (event occurs) shows the reconstructed sequence, which is encoded at the original spatial resolution and higher quality.

Event-based scalable encoding can be used in Future Internet to perform rate-optimization for transmission and storage according to the event significance in a video sequence.



a. there is no event in the video sequence. b. the event occurs in the video sequence.

Figure 4: Subjective result of event-based scalable encoding.

3. Bit-stream protection for transmission

SVC provides different bit-layers of different importance with respect to decoded video resolution or quality as explained in section 1.1.1. Thus, the JSCC can apply UEP according to the importance of a given portion of the bit-stream. In this context, scalable coding emerges as the natural choice for highly efficient JSCC with UEP.

The JSCC approach proposed in this chapter exploits the joint optimization of the wavelet-based SVC and a Forward Error Correction method (FEC) based on Turbo Codes (TC) [9]. Regarding channel coding, TC are one of the most prominent FEC techniques having received great attention since their introduction in 1993. Its popularity is mainly due to its excellent performance at low bit error rates, reasonable complexity, and versatility for encoding packets with various sizes and rates. In this research work Double Binary TC (DBTC) [10] is used for FEC rather than conventional binary TC, as DBTC usually performs better than classical TC in terms of better convergence for iterative decoding, a large minimum distance and low computational cost [10].

3.1. System Overview

The proposed framework consists of two main modules as shown in Figure 5: scalable video encoding and UEP encoding. At the transmitter side, the input video is coded using the wavelet-based scalable coder. The resulting bit-stream is adapted according to channel capacities. The adaptation can also be driven by terminal or user requirements when this information is available. The adapted video stream is then passed to the UEP encoding module where it is protected against channel errors. Three main sub-modules make up the UEP encoding part. The first one performs paketisation, interleaver design and CRC. The second one estimate and allocate bit rates using a rate-distortion optimization. The last UEP encoding sub-module is the actual DBTC.

After Quadrature Phase Shift Keying (QPSK) modulation, the video signal is transmitted over a lossy channel. At the receiver side, the inverse process is carried out. The main processing steps of the decoding are outlined in Figure 5. In this chapter Additive White Gaussian Noise (AWGN) and Rayleigh fading channels are considered. However, the proposed method can be equally applied to other lossy channels.

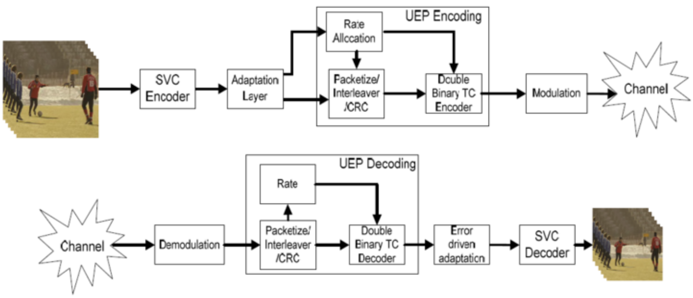


Figure 5: Communication chain for video transmission.

In the proposed JSCC framework DBTC encoding is used for FEC before BPSK/QPSK modulation. CRC bits are added in the packetisation of DBTC, in order to check the error status during channel decoding at the receiver side. Effective selection of the channel coding parameters leads to a minimum overall end to end distortion, i.e., maximum system PSNR, at a given channel bit rate. The underlying problem can be formulated as:

$$\min D_{s+c} \text{ subject to } R_{s+c} \le R_{\max} \tag{1}$$

$$\max\left(PSNR\right)_{s+c} \text{ subject to } R_{s+c} \le R_{\max} \tag{2}$$

where $R_{s+c} = R_{SVC} / R_{TC}$, D_{s+c} is the expected distortion at decoder, R_{s+c} is the overall system rate, R_{SVC} is the rate of the SVC encoded video, R_{TC} is the channel coder rate and $R_{\max}$ is the given channel capacity. Here the index notation $s+c$ stands for combined source-channel information.

The constrained optimization problem (1)-(2) can be solved by applying unconstrained Lagrangian optimization. Accordingly, JSCC aims at minimizing the following Lagrangian cost function J_{s+c}:

$$J_{s+c} = D_{s+c} + \lambda \cdot R_{s+c}, \tag{3}$$

with λ as the Lagrangian parameter. In the proposed framework the value of λ is computed using the method proposed in [11]. Since quality scalability is considered, R_{s+c} can be defined as the total bit rate over all quality layers:

$$R_{s+c} = \sum_{i=0}^{Q_M} R_{s+c,i}\,. \tag{4}$$

To estimate D_{s+c} in (3), let $D_{s,i}$ be the source coding distortion for layer i at the encoder. Since the wavelet transform is unitary, the energy is supposed to be unaltered after wavelet transform. Therefore the source coding distortion can be easily obtained in wavelet domain. Assuming that the enhancement quality layer i is correctly received, the source channel distortion at the decoder side becomes $D_{s+c,i} = D_{s,i}$. On the other hand, if any error happens in layer i, the bits in this layer and in the higher layers will be discarded. Therefore, assuming that all layers k, for $k < i$ are correctly received and the first corrupted layer is $k = i$, the jointly source-channel distortion at any layer $k = i, i+1, \ldots, Q_M$ at the receiver side becomes $D_{s+c,k} = D_{s,i-1}$. Then, the overall distortion is given by

$$D_{s+c} = \sum_{i=0}^{Q_M} p_i \cdot D_{s,i} \ , \tag{5}$$

where p_i is the probability that the *i-th* quality layer is corrupted or lost while the *j-th* layers are all correctly received for $j = 0,1,2,...,i-1$. Finally, p_i can be formulated as:

$$p_i = \left(\prod_{j=0}^{i-1} \left(1 - pl_j\right) \right) \cdot pl_i \, , \tag{6}$$

where pl_i is the probability of the *i-th* quality layer being corrupted or lost. pl_i can be regarded as the layer loss rate [11].

According to (6) the performance of the system depends on the layer loss rate, which in turn depends on the DBTC rate, the packet size and the interleaver. Once the channel condition and the channel rate are determined, the corresponding loss rate pl_i can be estimated by applying an iterative algorithm to estimate minimum distance between the zero code word and any other codeword $d_{\min}$ in the DBTC. Assuming that $d_{\min}$ is available, pl_i can be estimated as:

$$pl_i \propto \left(1/d \min\right) \tag{7}$$

Using (7), p_i can be evaluated from (6). As a consequence the problem of finding p_i boils down to find $d_{\min}$ that is explained in [11].

3.2. Performance Analysis

The performance of the proposed JSCC [11] framework has been extensively evaluated using the W-SVC codec. For the proposed JSCC UEP optimal channel rate, packet size and interleaver for DBTC were estimated and used. as described in this section. The proposed technique is denoted as “ODBTC”. Two other advanced JSCC techniques were integrated into the same SVC codec for comparison. The first technique used serial concatenated convolutional codes of fixed packet size of 768 bytes and pseudo random interleaver [12]. It is denoted as “SCTC”. The technique using product code proposed in [13] was used for the second comparison. This product code used Reed Solomon (RS) code as outer code and Turbo codes as inner code, so it is denoted by “RS+TC”.

After QPSK modulation, the protected bit-streams were transmitted over error-prone channels. Both AWGN and Rayleigh fading channels were used in the experimental evaluation. For each channel emulator, 50 simulation runs were performed, each using a different error pattern. The decoding bit rates and sequences for SNR scalability defined in [5] were used in the experimental setting. For the sake of conciseness the results reported in this section include only certain decoding bit rates and test sequences: Soccer at CIF (352 × 288) resolution and several frame rates. Without loss of generality, the t + 2D scenario for W-SVC was used in all reported experiments. The average PSNR of the decoded video at various BER was taken as objective distortion measure. The PSNR values were averaged over all decoded frames. The overall PSNR for a single frame is averaged over all colour components.

A summary of PSNR results is shown in Figure 6 and Figure 7. These results show that the proposed ODBTC consistently outperforms SCTC, achieving PSNR gains at all signal to noise ratios (E_b / N_o) for the AWGN channel. We have observed similar pattern of PSNR also for the Rayleigh fading channel. For the Sequence soccer up to 3 dB can be gained over SCTC when low E_b / N_o, or high channel errors, are considered for both AWGN channel and Rayleigh fading channel. It can be observed that the proposed scheme achieves the best performance for different channel conditions. As the channel errors increase or E_b / N_o decrease, a gap between the proposed scheme and SCTC becomes larger. The performance of RS+TC is almost comparable to ODBTC, with a slight PSNR degradation in most of the cases. However, it should be noted that RS+TC uses product code, where a much larger complexity is introduced by encoding and decoding of the RS codes and TC together.

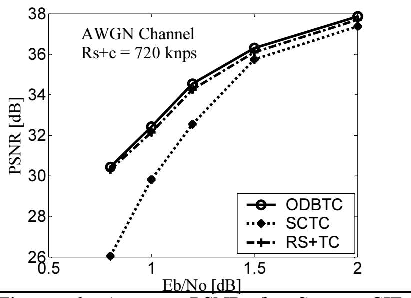


Figure 6: Average PSNR for Soccer CIF sequence at 30 Hz at different signal to noise ratio for AWGN channel.

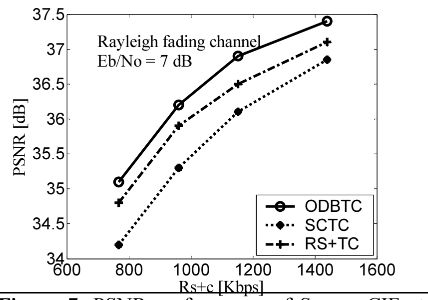


Figure 7: PSNR performance of Soccer CIF at 30 Hz at different bit rates.

These results demonstrate that for the considered channel conditions, the proposed ODBTC consistently outperforms SCTC, achieving PSNR gains at all tested bit-rates. Specifically, for the Sequence Soccer up to 1 dB can be gained for Rayleigh fading channel at 7 dB. RS+TC performs better than SCTC: it is comparable to ODBTC. However, at high SNR, the gap widens up to 0.4 dB. As the error rate probability of Gilbert-Elliot (GE) is very similar to Rayleigh fading channel as explained in [14]. Hence similar behaviour of the proposed technique can be observed for other channels like GE. In this chapter, we presented the result of low SNR as this is more critical in terms of error rate, however quite stable results can be observed at high SNR with the proposed technique.

Conclusions

In this chapter, practical implementation of scalable video coding with respect to future media internet was presented. The event-driven scalable video coding is introduced as the potential application for future internet. The approach reduces the bit-rate for those temporal segments of a sequence that do not contain important events. Furthermore an efficient joint scalable source and channel coding scheme elucidated. The rate optimization of the joint SVC and a forward error correction was proposed. UEP is used to minimize the end-to end distortion by considering the channel rate at given channel conditions and limited complexity. The results of simulations show that the proposed techniques provide comprehensive bit-rate optimisation and a more graceful pattern of quality degradation that can be a candidate for the future media internet.

References

[1] H. Schwarz, D. Marpe and T. Wiegand, "Overview of the scalable video coding extension of the H.264 / AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 17, Iss. 9, pp. 1103 - 1120, September 2007.

[2] I. Ahmad, X. Wei, Y. Sun and Y.-Q. Zhang, "Video transcoding: an overview of various techniques and research issues," *IEEE Transactions on Multimedia*, Vol. 7, Iss. 5, pp. 793 – 804, October 2005.
[3] M. Wien, H. Schwarz and T. Oelbaum, "Performance analysis of SVC," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 17, Iss. 9, pp. 1194 – 1203, September 2007.
[4] L. P. Kondi, F. Ishtiaq and A. K. Katsaggelos, "Joint source-channel coding for motion-compensated DCT-based SNR scalable video," *IEEE Trans. Image Process*., vol. 11, no. 9, pp. 1043–1052, Sep. 2002.
[5] 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/ MPEG2006/M13146, 76th MPEG Meeting, Montreux, Switzerland, April 2006.
[6] ITU-T and ISO/IEC JTC 1, *Advanced video coding for generic audiovisual services*, ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC).
[7] ITU-T and ISO/IEC JTC 1, *Generic coding of moving pictures and associated audio information – Part 2: Video*, ITU-T Recommendation H.262 and ISO/IEC 13818-2 (MPEG-2 Video).
[8] ISO/IEC JTC 1, *Coding of audio-visual objects – Part 2: Visual*, ISO/IEC 14492-2 (MPEG-4 Visual)
[9] C. Berrou and A. Glavieux, "Near-optimum error-correction coding and decoding: Turbo codes," IEEE Trans. Commun., vol. 44, no. 10, pp.1261-1271, Oct. 1996.
[10] C. Doulliard and C. Berrou, "Turbo codes with rate-m/(m+1) constituent convolutional codes," IEEE Trans. Commun., vol. 53, no. 10, pp 1630-1638, Oct. 2005.
[11] N. Ramzan, S. Wan and E. Izquierdo, "Joint Source-Channel Coding for Wavelet-Based Scalable Video Transmission Using an Adaptive Turbo Code," *EURASIP Journal on Image and Video Processing*, vol. 2007, Article ID 47517, 12 pages.
[12] B. A. Banister, B. Belzer and T. R. Fischer, "Robust video transmission over binary symmetric channels with packet erasures", in Proc. Data Compression Conference, DCC2002, pp. 162 - 171, Apr. 2002.
[13] N. Thomos, N. V. Boulgouris and M.G Strintzis, "Wireless image transmission using turbo codes and optimal unequal error protection", IEEE Trans. Image Process. Vol. 14, no 11, pp. 1890 – 1901, Nov. 2005.
[14] Hatem Boujemˆaa, Mimouna Ben Said and Mohamed Siala, "Throughput Performance Of Arq And Harq I Schemes Over A Two-States Markov Channel Model", in Proc. of 12 international conference on Electronics, Circuits, and Systems, Dec. 2005.
[15] C. Stauffer, W. E. L Grimson. "Learning patterns of activity using real-time tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, pp. 747-757, (2000).