ERROR-RESILIENT TRANSMISSION OF H.264 SVC STREAMS OVER DVB-T/H AND WIMAX CHANNELS WITH MULTIPLE DESCRIPTION CODING TECHNIQUES

Peter Schelkens[1], Augustin Gavrilescu[1], Adrian Munteanu[1], Yves D'Hondt[2], Kenneth Vermeirsch[2], Rik Van de Walle[2], Olivier Guye[3], Stewart Worrall[4], Antonio Navarro[5]*

[1]IBBT, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium
[2]IBBT, Universiteit Gent, Gaston Crommenlaan 8, 9050 Gent, Belgium
[3]VITEC Multimedia, rue Pierre Semard 99, 92324 CHATILLON Cedex, France
[4]Centre for Communications Systems Research, University of Surrey, GU2 7XH, Guildford, UK
[5]Telecomunications Institute, University of Aveiro, 3810 Aveiro, Portugal
**Email: peter.schelkens@ibbt.be*

ABSTRACT

The European STREP Project SUIT involves the development of a multiple description scalable video codec for the transmission of video content over DVB-T/H and WiMAX channels. In this paper, we elaborate on the architectural design choices for such a multiple description transmission system, which utilizes H.264 Scalable Video Coding (SVC). We follow a phased approach focusing on the following unbalanced multiple description coding approaches: fully redundant unbalanced multiple description coding, multiple description coding based on redundant slices and multiple description coding by means of embedded multiple description quantisation.

1. INTRODUCTION

The ever increasing demand for efficient transmission of multimedia content over best effort networks and error-prone channels (e.g. packet networks, low-power wireless links) has fuelled intensive research in the area of robust communication techniques. In this context, multiple description coding (MDC) is a competitive solution to overcome the channel impairments by exploiting diversity [1][2] [3][4]. This coding paradigm relies on generating more than one description of the source such that (1) each description independently describes the source with certain fidelity and (2) when more than one description is available at the decoder, one can combine them to enhance the decoded quality. This has the inherent advantage that the quality of the reconstructed data gracefully degrades with increasing probability of failure on the transmission channel. The produced descriptions are independently decodable, allowing the decoder to reconstruct the source with certain fidelity from a subset of initial number of descriptions.

Within the scope of the European FP6 STREP *"Scalable, Ultra-fast and Interoperable Interactive Television (SUIT)"* project [5], two descriptions will be provided by an MDC system for transmission over two different networks i.e. DVB-T/H and WiMAX, respectively. It is important to note that, within the SUIT project, each of the DVB and WiMAX networks containing a (different) description are able to operate independently.

In addition to error-resilience, scalability is a desired feature in the context of efficient adaptation to data transmission conditions. In the proposed architecture, we deploy H.264 SVC (Annex F) [6] that is currently configured to allow for quality scalability.

Consequently, the design of a Multiple Description Scalable Video Coding (MD-SVC) system is of paramount importance. MD-SVC will be capable to deliver video content over best-effort error-prone packet networks, and, due to its scalable erasure-resilient compression capabilities, it is able to (1) meet the users' requirements in terms of quality and resolution; (2) dynamically adapt the rate to the available channel capacity, and (3) provide robustness to data losses as retransmission is often impractical.

It is important to note that in a video streaming scenario, selective retransmission of lost packets is often not desired because of the timing requirements and low delay that are expected from the system. This is especially true for broadcast and multicast scenarios where the streaming server would be burdened by a potentially very large amount of retransmission requests. An alternative approach taken in such a scenario is to deploy additional Forward Error Correction (FEC) at the video layer. This allows receivers to autonomously correct bit errors or packet erasures caused by the lower layers of the network, without the need for retransmission of information. Some examples of FEC codes are block codes, convolutional codes, and LDPC codes.

The basic protection schemes incorporating FEC codes are designed from a worst case point of view, and the amount of added redundancy is fixed even under ideal network conditions. In this case, FEC codes suffer from what is known as the cliff effect: by design, they show a constant excellent performance up to a well-defined number of erasures; once the number of actual erasures exceeds this figure, performance drops very sharply to a very low level. MD-SVC provides a much more gradual performance vs. erasures curve and wastes less bandwidth resources under good network conditions [7].

In the next section, we will first introduce three different approaches for multiple description coding of scalable video: fully redundant unbalanced MDC, MDC based on redundant slices and MD based on embedded multiple description quantisation. Thereafter, two potential scenarios: the gateway and the terminal scenario are discussed as well as their impact on the SUIT system architecture. Finally, the last section summarizes this paper.

2. MULTIPLE DESCRIPTION, SCALABLE VIDEO CODING

2.1 Fully Redundant Unbalanced MDC

A Balanced Multiple Description (BMD) system, in which all descriptions are of equal rate and equally important, suffers from the inherent drawback that bandwidth utilization is often not optimal since all descriptions are of equal rate: the rate must be kept below the minimum available bandwidth among all channels.

Hence, in SUIT's base-line two-channel scenario, Unbalanced Multiple Description (UMD) is deployed where the UMD coder generates a coded video stream at full quality, along with a second version at reduced quality. Both descriptions are transmitted over a different network. The low quality version is only used to support error concealment in the high quality version in case the latter gets distorted during transmission. The descriptions are unbalanced, since the low quality version will typically be of much lower rate than the full quality description. UMD enables improved allocation of available bandwidth resources in the underlying networks.

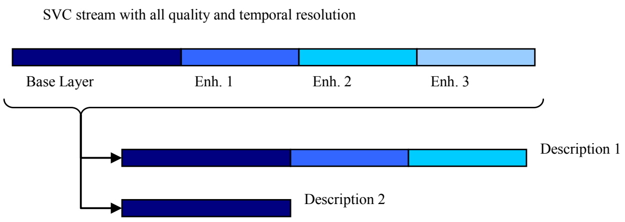

Figure 1- Two descriptions Unbalanced MDC generated from a full quality SVC stream

In an SVC context, the low fidelity version potentially differs from the high fidelity version along all three axes of scalability: temporal, spatial, and/or quality (SNR). However, to maintain acceptable video quality when every description is lost, the use of SVC's inter-layer prediction within both descriptions must be avoided. In addition it has to be noted that the degree of redundancy across both descriptions will become very large when both networks offer comparable bandwidth resources. In that case alternative solutions prove to be more efficient.

2.2 MDC based on redundant slices

Using redundant slices, multiple descriptions of a video sequence can be generated at a high level in the video coding process, namely the Network Abstraction Layer (NAL) unit level. The fundamental idea is that (depending on the amount of redundancy that is desired) some or all coded slices are sent in both descriptions. At the receiving side, it is sufficient that only one copy of every slice is received in order to fully reconstruct the video sequence.

In the proposed architecture, the provision in H.264 to signal in the slice header whether the slice is redundant or not, is deployed. In this way, both descriptions can be merged at the receiver side and sent to the decoder in a single, compliant bitstream. This simplifies the task of the MDC combiner, but requires a decoder that supports the redundant slice syntax. It also means that more traffic will be sent over the last-mile network in a gateway scenario (see section 3). Alternatively, the MDC combiner can combine the MD video streams to a video stream that is compliant with an AVC/SVC decoder not supporting the redundant slice syntax at the expense of having to parse and modify the slice headers.

It should be stressed that, since the unit of granularity in these techniques is a slice, the proposed methods can be applied to any layer in a scalable bitstream. Which layers will be protected by redundant slices is purely a matter of policy.

2.2.1 Duplication of slices with H.264/AVC redundant slice signalling

A first solution transmits a primary slice and its redundant copy, each containing identical data while signalling the nature of the slices (original or redundant) in the bitstream using the H.264 redundant slice syntax. More precisely, the *redundant_pic_cnt* syntax element shall be zero in the primary slice and nonzero in redundant slices. From the perspective of the terminal/gateway, the use of this scheme implies that we no longer need to eliminate duplicate NAL units 'by hand'. However, we do need a decoder that supports redundant slices. Also, in the gateway scenario, all NAL units from both descriptions need to be sent over the last-mile network, thus greatly increasing its bandwidth requirements.

At the playout side, the use of this technique allows two approaches:

1. An encoder is needed that is capable of producing the redundant slices itself, and outputs two NAL unit streams corresponding to both descriptions.
2. The encoder has no support for redundant slices; as in the previous technique, a separate module generates the descriptions from the single NAL unit stream produced by the encoder. However, now this module not only duplicates the NAL units, but also modifies the slice headers, making sure the result complies to H.264/SVC. These modifications require the parsing of slice headers, picture parameter sets and sequence parameter sets, so it must be stressed that this is more complex than the previous approach.

2.2.2 Redundant slices with lower texture quality

When using redundant pictures, no hard requirement exists that forces a redundant slice to contain data identical to its corresponding primary slice. In other words, redundant slices can be of lower quality than their original. This is a useful feature in case one description is sent over a secondary network with a considerably lower available bandwidth (Figure 2).

However, when image data from a slice is used as a prediction signal for future coded slices, but the slice is replaced by a version with lower texture quality, drift errors may occur in: (1) other slices of the same picture, due to H.264 intra prediction; (2) future pictures (in coding order), due to inter-frame prediction and (3) higher layers of a scalable bitstream, due to inter-layer prediction when the current layer is used as a reference for motion or texture information in higher layers.

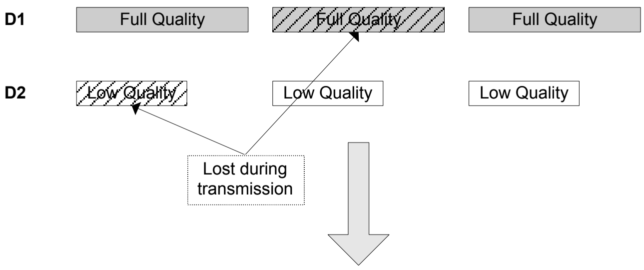

Figure 2 - MDC combining of two descriptions of unequal quality.

In other words, a trade-off exists between the amount of redundancy and the attainable reconstruction quality at the decoder given a specific network condition.

2.2.3 Concluding remarks on redundant slice coding

The list of redundant slices techniques presented here is in certainly not exhaustive. Because of the freedom that exists in exploiting redundant slices, many more schemes are imaginable; for example, the choice of the location of redundant slices within a picture could be driven by Region-Of-Interest (ROI). It is important to remember that the more advanced techniques are only viable when encoder and decoder support for redundant slices is present.

2.3 MDC based on EMDSQ

2.3.1 Embedded Multiple Description Scalar Quantisation – EMDSQ

The first practical MD coding system based on scalar quantization was proposed by Vaishampayan in [1] where the concept of MD scalar quantisation (MDSQ) was introduced. MDSQ can be seen conceptually as a set of independent scalar quantizers that output several descriptions of the same input source sample. The MDSQ approach can be divided in two steps. First, the central quantizer, leading to an optimal partitioning of the contained cells, has to be determined. Secondly, given the partitioning of the central quantizer, the design problem of an index assignment scheme, which efficiently allocates the indices of the individual side quantizers has to be tackled. In brief, MDSQ consists of two main components: (a) a scalar quantizer (b) an index assignment (IA). Figure 3 illustrates the structure of an MDSQ with the two main components as described above.

In order to shed light on the technique presented above, let us consider a MDSQ based on a uniform scalar quantizer and a staggered IA matrix (two diagonals). We first define the so-called central quantizer by partitioning the input sample range into a number of cells (Figure 4). The central quantizer reflects the video quality we would like to attain when all descriptions arrive at the receiver side. However, the centrally-quantized coefficients are not transmitted to the receiver; instead, neighbouring central quantizer cells are grouped together in two distinct ways, resulting in two distinct side quantizers, each belonging to a description. This grouping is determined by the Index Assignment (IA) matrix. For each side quantizer, the resulting quantized coefficient is transmitted through the corresponding description.

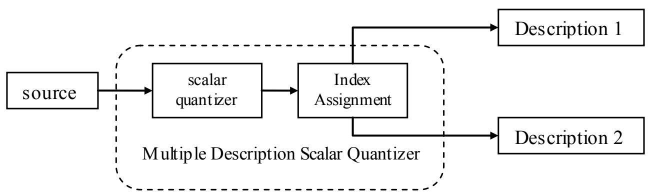

Figure 3 - The MD coding system based on multiple descriptions scalar quantizers

The grouping is done in such a way that the receiver is able to inverse quantize the coefficient according to the central quantizer when all descriptions are received. In case one or more descriptions are missing, the receiver will be unable to narrow down to a single central quantizer cell; hence, the coefficient will be reconstructed to a lesser degree of precision. As a result, the loss of descriptions will result in a graceful progressive loss of picture quality. An example of how the quantizers could be defined in a two-description system is shown in the figure below. Suppose description 1 signals S_0^1, and description 2 signals S_1^2, then the coefficient will be reconstructed as C_1.

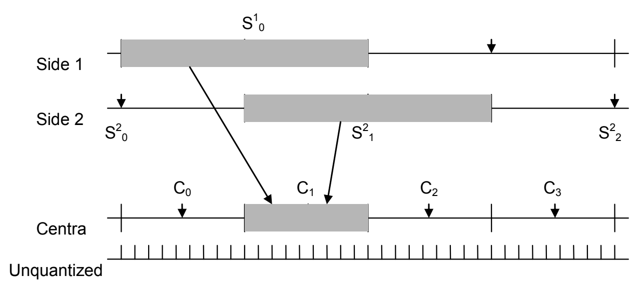

Figure 4 - Example of central and side quantizers in a two-description system (only part of axis shown)

In the case of a staggered IA matrix we will obtain for the side reconstruction the centroid of the following reconstructed side cells:

- Side 1: $S_0 = 2n_1$, $S_1 = 2n_1 - 1$
- Side 2: $S_0 = 2n_2$, $S_2 = 2n_2 + 1$

To provide also in the ability to produce layered descriptions allowing for progressive transmission of each distinct description, we have to deploy an embedded MDSQ (EMDSQ) system. As became apparent in the previous paragraphs, a coding system based on MDSQ has the ability to change the amount of redundancy allocated between the descriptions by changing the corresponding set of central and side quantizers. For EMDSQ we have a corresponding set of side and central quantizers at each distinct level. Therefore, in EMDSQ one can tune not only the overall redundancy, but the redundancy at each distinct quantization level as well. Another property of EMDSQ consists in its ability to yield uniform embedded central quantizers at each quantization level.

In order to produce layered descriptions we have to rely on the so-called embedded IA. The principle behind such an embedded IA consists in designing a recursive matrix decomposition. For more details we refer to [7][8].

2.3.2 SVC based on EMDSQ

EMDSQ is integrated in a SVC system that is generating two descriptions, one to be sent via DVB-T/H and the other via a WiMAX network. The starting point is the standardised SVC video coder. However, in order to obtain an MD-SVC coder, a corresponding MD block has to be included into the video coding scheme. The block scheme of such a video coder is depicted in Figure 5. Basically, instead of directly encoding the quantized indices, they are first processed by an IA matrix supporting EMDSQ, as such resulting in a quality-layered multiple description specification. Thereafter, each separate description is entropy encoded and motion vector information is added. In this way the two resulting descriptions are independently decodable since they contain all the necessarily information. Additionally, the level of redundancy, introduced between the indices contained in the two descriptions corresponding to the same original macroblock, can be tuned. It is important to notice that the motion vector information is completely redundant and as a result, the loss of one description will not affect the motion vectors.

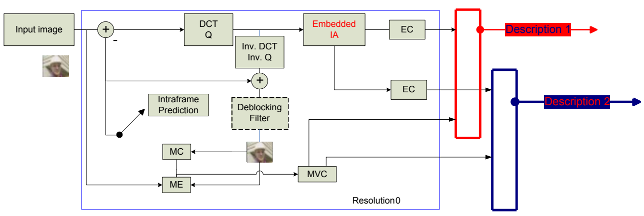

Figure 5 - MD-SVC block scheme

The developed MD-SVC scheme provides two distinct descriptions due to the MD module incorporated into the video coding scheme. Each of the distinct descriptions is independently decodable and has a layered representation allowing for rate adaptation according to the available bandwidth and users' needs. The redundancy between the two descriptions can be independently adjusted for every quality layer providing unequal error protection based on the importance of the information contained in the transmitted stream.

It is though important to remark that in the case of such MD video coding system, it is not possible to adapt on the fly the amount of redundancy to variable network conditions.

3. GATEWAY AND DECODER

This section describes the synchronization between both received descriptions and their combination into a single description video stream. It covers the scenarios of retransmission to the last-mile network (gateway scenario) and without a last-mile network (terminal scenario). The gateway and the terminal scenario follow the same procedure till the MDC combiner (Figure 6).

First, correctly received link-layer packets of both descriptions are reassembled and their UDP and IP headers are removed as they traverse the UDP/IP protocol stack of the corresponding transceiver (WiMAX or DVB). The resulting stream of RTP packets is delivered to a separate RTP depacketization and decapsulation module for each description. This module performs the following tasks: (1) reordering of the received RTP packets based on their RTP sequence number, since the User Datagram Protocol guarantees no in-order arrival of packets; (2) depacketization of RTP Aggregate Packets (if used), i.e. recovering the NAL units one by one; and (3) delivery of the NAL units in decoding order.

However, the resulting streams of NAL units may not be synchronized due to unequal delays in the WiMAX and DVB transmission chains. Furthermore, due to packet loss NAL units may have been lost in both networks. Therefore, synchronization between both NAL unit streams is required before they are delivered to the MDC combiner (Figure 6). To limit the size of the buffer and speed up the synchronization process, delayed NAL units of one of the streams could be marked as lost, while in reality they still have to arrive.

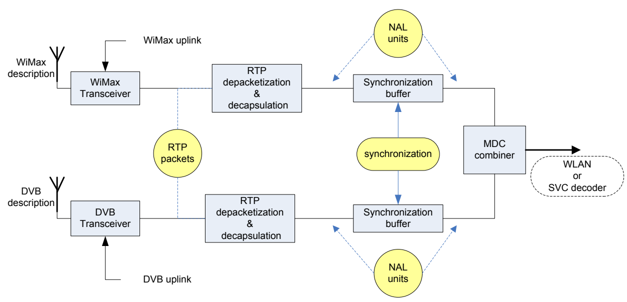

Figure 6 - Schematic view of Gateway and Terminal common blocks (video data path)

For the three, earlier specified MD coding schemes, namely (1) Fully Redundant Unbalanced MD, (2) MDC based on redundant slices and (3) MDC based on EMDSQ, the MDC combiner will operate as specified in the following paragraphs.

In case of *Fully Redundant Unbalanced MDC*, synchronized packets are arriving at the MDC combiner. The role of the MDC combiner limits itself to the simple case of dropping the redundant NAL units and signalling the missing NAL units (in the case of errors in both descriptions). If one of the streams is received correctly, there is no need for synchronization since the other stream does not contain any additional information.

For *MDC based on redundant slices*, combining both descriptions into a single compliant bitstream can be performed in the network abstraction layer (NAL). At the play-out side, the MDC generator will tag each NAL unit with a sequence number, a frame number and/or a timestamp. This will support the synchronisation and combination of both descriptions with each other. The requirements for the MDC combiner in the terminal/gateway depend then upon the approach used.

When using the H.264 support for signalling redundant slices, the combiner has to parse the slice header of all coded slices in order to find out whether or not the slice is redundant, and must then decide whether it can be discarded. Unavoidably, this increases the computational requirements and hence the delay of the combiner. As an alternative, when a decoder with support for H.264 redundant slices is available, the task of removing the redundant slices can be offloaded to this decoder. However, this will increase the bandwidth requirement for the last-mile WLAN network, since both redundant and non-redundant slices need to be transmitted through this network.

When not using the H.264 syntax to mark redundant slices, the MDC combiner will instead investigate the numbers with which each NAL unit was tagged at the play-out side. This allows the combiner to detect which slices were lost during transmission, and hence to correctly replace lost slices with their redundant copy.

In the case of an *MDC based on EMDSQ,* the MDC combiner module for use with the EMDSQ technique requires access to the quantized coefficients of both descriptions; therefore entropy decoding is first performed on both incoming video streams. Then, the actual reconstruction of both side-descriptions into a single description is executed, according to the number of descriptions available. Entropy coding will have to be reapplied in order to obtain an SVC compliant bitstream to be sent over the local WLAN. The EMDSQ combiner is in fact an inverse IA matrix which is mapping pair of coefficients into the reconstructed central descriptions.

4. CONCLUSIONS

This paper presents a system architecture for a quality scalable multiple description video coding system (MD-SVC). This system is capable of providing layered descriptions allowing for rate adaptation among each of the targeted networks, namely WiMAX, DVB-T/H and WLAN. This is feasible since the coder inherits its scalable features from the H.264 SVC coder. In order to provide error resilience an MD module is employed, allowing the exploitation of three different approaches: (1) Unbalanced MD; (2) MD based on redundant slices and (3) MD based on EMDSQ.

All employed MD methods provide several descriptions of the scalable video stream in order to overcome the error-proneness of the transmission channels.

Though the architecture proposed in this paper was defined for a specific network and codec setting, the general principles can be ported to any transmission systems that allows for the exploitation of path or time diversity.

ACKNOWLEDGEMENTS

This work was supported by the European FP6 Project “Scalable, Ultra-fast and Interoperable Interactive Television (SUIT)”, the Belgian Federal Science Policy Office (BFSPO) via the IAP “Mobile” project, the Fund for Scientific Research-Flanders (FWO-Flanders) via the “Metadata-based Multimedia Systems” project, the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Interdisciplinary Institute for Broadband Technology (IBBT) and Ghent University. P. Schelkens and A. Munteanu hold postdoctoral fellowships with the Fund for Scientific Research, Flanders (FWO).

REFERENCES

[1] V. A. Vaishampayan, “Design of multiple description scalar quantizers,” *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 821-834, 1993.

[2] V. K. Goyal, “Transform coding with discrete transforms,” *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 465-473, 2000.

[3] W. Jiang and A. Ortega, “Multiple description coding via polyphase transform and selective quantization,” in *SPIE Int. Conf. Visual Comm. Image Proc. VCIP'99*, San Jose, USA, 1999, pp. 998-1008.

[4] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, “Priority encoding transmission,” *IEEE Transactions on Information Theory*, vol. 42, pp. 1737-1744, 1996.

[5] http://suit.av.it.pt/

[6] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the Scalable H.264/MPEG4-AVC Extension,” to be presented at IEEE International Conference on Image Processing (ICIP 2006), Atlanta, GA, USA, October 8-11, 2006, invited paper.

[7] F. Verdicchio, A. Munteanu, A. Gavrilescu, J. Cornelis, P. Schelkens, “Embedded Multiple Description Coding of Video,” IEEE Transactions on Image Processing, vol. 15, no. 10, pp. 3114 – 3130, October 2006

[8] F. Verdicchio, A. Gavrilescu, A. Munteanu, J. Barbarien, J. Cornelis, P. Schelkens “Multiple Description Coding – State-of-the-art and recent developments,” in “Recent advances in applied signals, systems and image processing,” Springer-Verlag (To be Published)

[6] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the Scalable H.264/MPEG4-AVC Extension,” to be presented at IEEE International Conference on Image Processing (ICIP 2006), Atlanta, GA, USA, October 8-11, 2006, invited paper.