Low-Complexity Standard-Compatible Robust and Scalable Video Streaming over Lossy/Variable Bandwidth Networks

Andrea L. Vitali, Fabrizio S. Rovati, Roberto Rinaldo, Riccardo Bernardini, Marco Durigon

Abstract — We propose a robust and scalable transmission scheme based on Multiple Description (MD) coding. Each description is encoded using a standard video codec such as MPEG-2 or H.264. This scheme provides equal or better robustness with respect to other solutions such as Forward Error Correction (FEC), at lower system complexity; also, it provides scalability by selectively dropping descriptions.1

Index Terms — multiple description coding, video streaming, mpeg-2, h.264.

I. INTRODUCTION

Several techniques have been proposed to enhance error-resiliency of video streams sent through lossy channels [1]. Among these, there are techniques like forward error detection/correction codes (FEC) or automatic repetition requests (ARQ). ARQ is very effective but it requires a feedback channel and it can be used only in point-to-point communications, not for broadcast. On the opposite, FEC does not require a feedback channel and it is suitable for broadcast. However, FEC usually needs to be complex in order to be effective and it has an all-or-nothing performance: if the correction capability is exceeded, almost nothing is delivered to the receiver.

It is often useful to jointly optimize the parameters of the source and channel encoders. In the case of video communications this means to exploit the error resiliency already present in every compressed video bitstream rather than using complex FEC.

As an example, in MPEG-x/H.26x encoders it is possible to increase error resiliency by using one or more of the following techniques: more frequent Intra pictures (to reset the motion prediction loop) or using a suitable intra macroblock update policy (random Intra MB refresh); more slices per picture (to reset differential motion vector and DC coefficient coding) or using flexible macroblock order (FMO eases concealment, with H.264); encoding concealment motion vectors or using interleaved multiframe prediction policies; using reversible variable length coding (as RVLC in MPEG-4) or an error resilient entropy coding scheme (EREC) or more sync marks; etc...

Another way of enhancing error-resiliency without using complex FEC codes is by means of Multiple Description (MD) video coding schemes [2, 3]. The goal of the proposed Multiple Description Coding scheme is to create several independent bitstreams using an existing video codec. In the proposed approach the video codec need not to be MD aware but can be used as a black box. Bitstreams can be decoded independently or jointly.

MD does not require a feedback channel. It has a graceful degradation in performance because the more bitstreams decoded, the larger the output video quality. It improves error resiliency because it is not likely to have the same portion of the same frame corrupted in every description. MD can exploit path diversity. Variable bandwidth/throughput can be easily managed by transmitting a suitable number of descriptions (this is analogous to scalable coding). However, coding efficiency is somewhat reduced depending on the amount of redundancy left among subsequences, but this is also true when one of the aforementioned solution is used2.

Example applications are in-home wireless robust distribution or scalable multicast through internet or cellular networks. Receivers can decode as many descriptions as needed based on their memory, power, display capabilities.

II. MULTIPLE DESCRIPTION VIDEO CODING

MD is analogous to scalable coding. The difference lies in the dependency among bitstreams. The simplest case is when two bitstreams are created. In the case of scalable coding they are referred to as base layer and enhancement layer, the latter depends on the former and cannot be decoded independently. On the other hand, in the case of multiple description coding, each description can be individually decoded to get a base quality video. As for layers in scalable coding, descriptions can contribute to one or more of the following: spatial resolution, temporal resolution or SNR quality.

There are many techniques to create multiple descriptions: MD quantization, correlating transforms and filters, quantized frames or redundant bases, FEC combined with layered coding, spatial or temporal polyphase downsampling [4, 5].

Many of these schemes can be adapted to existing video codecs which are based on prediction, transform, quantization and entropy coding: it is possible to create descriptions in the pixel domain, in the error-prediction domain or in the transform domain.

1 Andrea L. Vitali and Fabrizio S. Rovati are in STMicroelectronics, Advanced System Technology (AST), via Olivetti 2, 20041 Agrate Brianza, Italy; e-mail: andrea.vitali@st.com, fabrizio.rovati@st.com.

Prof. Roberto Rinaldo, prof. Riccardo Bernardini and Marco Durigon are in University of Udine, Dipartimento di Ingegneria Elettrica Gestionale e Meccanica (DIEGM), via delle Scienze 208, 33100 Udine, Italy; e-mail: rinaldo@uniud.it, bernardini@uniud.it.

2 There is a notable exception: ARQ or hybrid FEC schemes in which data or FEC is sent only if needed. However, these techniques are orthogonal to MD and in fact they can be applied to compressed descriptions.
A. MD in the error-prediction or in the transform domain

Working in the error-prediction domain or in the transform domain yields very efficient but very complex and cumbersome schemes. If not all descriptions are received correctly, the prediction drifts because the frame memory in the decoder will not be the same as the one used in the encoder. To solve this problem, prediction may be removed, but this greatly reduces the video compression capability of the codec. A smarter solution is to send a drift compensation term together with each description, but if there are more than two descriptions, the number of drift compensation terms increases dramatically, again reducing efficiency.

Because of this, we restricted our attention to schemes that work in the pixel domain.

B. MD in the pixel domain

Working in the pixel domain has the advantage that MD can be completely decoupled from the underlying video codec; descriptions can be created in a pre-processing stage before compression; successfully received descriptions can be merged in a post-processing stage after decompression. Spatial and temporal descriptions can be created by using Polyphase Downsampling (PDMD), programmable lowpass filters controls redundancy; SNR descriptions can be created by means of MD quantizers (either scalar or vector), the structure of quantizers controls redundancy.

Replicated headers/syntax and replicated motion vectors among bitstreams greatly impede coding efficiency in SNR MD. Replicated headers/syntax also hinder temporal MD, and motion compensation is less effective because of the increased temporal distance between frames. Spatial MD is hindered by headers/syntax as well but unlike with temporal MD, motion compensation is not affected, particularly when 8x8 blocks are split into smaller blocks, as in the latest H.264 codec.

C. Proposed scheme: polyphase downsampling MD

The proposed scheme (fig. 1) creates multiple descriptions via spatial Polyphase Downsampling (PDMD). Descriptions are all equal in importance, and each one is fully representative of the original image. Any video codec can be used to encode and decode descriptions.

We tested the scheme with MPEG-2 and the recent H.264. If all descriptions are received, we can reconstruct exactly the original information, except for video codec quantization noise. This method provides same or better robustness that other state-of-the-art schemes and can be implemented at a lower cost. In fact, complex FEC schemes may be avoided in the system, greatly simplifying the overall transmission chain and the decoding terminals. We are not comparing against ARQ, as this technique is not usable in broadcast/multicast, and anyway it is complementary to the use of our approach.

Error concealment algorithms and decoder architectures are also greatly simplified, as we use simple spatial interpolation filters for data loss concealment, instead of motion-estimation and compensation for macroblock replacement, commonplace in state-of-the-art error concealment techniques.

The specific MD error concealment can be very effective even if it is simple (fig. 2). On the opposite, state-of-the-art concealment for SD can be complex and still be unsatisfactory. PNSR statistics reflect this fact: the variance of the PSNR is lower when MD is used instead of SD.

Actually experiments - not discussed here - have confirmed that spatial PDMD is preferable over temporal PDMD, but the reason is slightly different: at very low bitrates, when temporal PDMD is used, there is an annoying flashing in the decoded sequence due to independent compression of descriptions; instead, when spatial PDMD is used, there is a kind of dithering due to the same reason. Dithering is not annoying, can be easily detected and eliminated, and it can even improve the perceived quality.
III. EXPERIMENTAL SETUP

It is not fair to compare MD against plain SD because SD is not designed for errors or losses, and because MD is designed to trade compression efficiency for robustness.

When there are no errors or losses, MD is penalized. For a given bitrate budget, descriptions must be compressed slightly more to compensate for headers/syntax overhead: there are multiple independent bitstreams.

Because of this we decided to compare MD against error resilient SD and SD protected by FEC. The data or FEC redundancy is increased until the decoded quality is the same, for a given bitrate budget and when there are no errors. Then error rate is increased: the winner is the scheme that delivers the highest PSNR with the highest probability (which means: lowest variance). Fig. 3, 4 and 5 and table 1 show results for such tests.

A. Multiple Descriptions: 2MD, 4MD and 3MD

The proposed Multiple Description scheme has been implemented in three different ways:

- 2MD: two descriptions generated by separating odd and even rows, which corresponds to polyphase downsampling of 1x2 pixel blocks.
- 4MD: four descriptions generated by separating odd and even rows and columns, which correspond to polyphase downsampling of 2x2 pixel blocks.
- 3MD: three descriptions, two generated as specified for the 2MD system, the third is obtained by filtering and applying a 2:1 downsampling along the vertical direction; this is know as Frames Expansion.

Concerning 2MD and 4MD: more descriptions means more robustness because it is less and less unlikely that the same portion of the same frame is lost in all descriptions.

Concerning 3MD: frame expansion [6] is a way to expand the original data so that some controlled redundancy is added. As an example: 2 descriptions can be generated by separating odd and even rows; the 3rd description can be simply the average of odd and even rows. It is clear that perfect reconstruction is achieved if any 2 descriptions out of 3 are correctly received. In practice, longer vertical filters are used. The 3MD system can be seen as equivalent to a FEC code with rate 2/3: one single erasure can be recovered.

B. Single Description: SD with 25% Intra MB Refresh

The proposed Multiple Description (MD) scheme has been compared to error resilient Single Description (SD) encoding. We set the percentage of Intra macroblocks to 25%, so that the decoded quality is approximatively equal to that obtained by 2MD at the same bitrate (see fig. 3).

C. Forward Error Correction at application level

The proposed scheme has been compared also to application
level Forward Error Correction (FEC) as specified in the IETF draft RFC2733; groups of data packets are xored so that a loss can be recovered at the receiver side. Redundancy is proportional to the number of xor packets. As an example, couples of packets can be xored together to add 50% FEC redundancy: a xor(b,c), b xor(a,c). At the receiver it is possible to recover one loss every three packets: a xor(b,c), b xor(a,c).

We added to SD as many xor packets as needed to get the same bitstream size as for 4MD at the same decoded quality. Table 1 shows results for such tests.

<table>
<thead>
<tr>
<th>SDTV sequence, 25 Hz, H.264 codec</th>
<th>Scheme</th>
<th>Loss</th>
<th>Bitrate</th>
<th>PSNR Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD</td>
<td>0%</td>
<td>4.1 Mbit/s</td>
<td>45 dB</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>10%</td>
<td>4.1 Mbit/s</td>
<td>31 dB</td>
<td></td>
</tr>
<tr>
<td>4MD</td>
<td>10%</td>
<td>6.5 Mbit/s</td>
<td>40 dB</td>
<td></td>
</tr>
<tr>
<td>SD+FEC</td>
<td>10%</td>
<td>6.5 Mbit/s</td>
<td>39 dB</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. SDTV sequence (Moulin Rouge taken from DVD source), PAL resolution, 25Hz, encoded with H.264 (fixed number of macroblocks/packet, fixed QP). FEC has been added as specified by RFC2733.

### IV. CONCLUSION

MD improves error resiliency because each bitstream can be decoded independently and it is not likely to have the same portion of the same frame corrupted in every description. Variable bandwidth can be easily managed by transmitting a suitable number of descriptions. MD can also exploit path diversity, key for example on base-stations handover. Each description has the same hierarchical importance, so MD does not need prioritization on the channel.

Coding efficiency is reduced compared to sending one single stream. However, this is also true when source coding is used to have some error resiliency (e.g. FMO, increased intra refresh) or when FEC is used. Robustness vs. efficiency loss can be traded-off by varying the number of descriptions in which original sequence is split.

### REFERENCES

6. R. Bernardini, R. Rinaldo, A. Tonello, A. Vitali “Frame based Multiple Description for Multimedia Transmission over Wireless Networks”

4 More sophisticated schemes are possible, as proposed by Siemens at 3GPP or in the standard draft IETF UX. Packets are put in a matrix column-wise or row-wise, redundancy rows are added using a sophisticated block code (such as Reed-Solomon) column by column, then data and redundancy are extracted row-wise, multiplexed and transmitted.

---


**Andrea Lorenzo Vitali** was born in Bergamo, Italy in 1971. He received a degree in Electronic Engineering in 1998 from Politecnico di Milano, Italy. Since 1998, he joined ST-Microelectronics’ research group on Advanced System Technologies. He worked on digital receivers and on digital multistandard decoders for analog TV. In 2001 he switched to advanced video algorithms for digital TV, mainly aimed to non-standard video compression methods. In the meanwhile he gave several lectures on Digital Electronics (Politecnico di Pavia, Italy). Since 2002 he’s working on joint source-channel coding, mainly aimed to internet/wireless video streaming. He authored or co-authored several patents and publications. His interests are pure and applied mathematics.

**Fabrizio Simone Rovati** was born in Monza, Italy in 1971. He received electronic engineering degree at the Milan Polytechnic University, Italy, in 1996. He joined STMicroelectronic Ltd., Bristol, UK (formerly INMOS Ltd.) where he contributed to the development of an MPEG-2 transport demultiplexer co-processor. He then joined STMicroelectronics’ Advanced System Technologies in 1998 where he worked on an MPEG-2 motion estimation co-processor and on MPEG video encoders system architectures. He authored or co-authored 15 granted patents and 6 publications. He has been contract professor at Pavia Polytechnic University during 2001/2002. His current interests are in digital video signal processing and robust delivery through IP-based networks, and related system-level and processors architectures.

**Roberto Rinaldo** obtained the “Laurea in Ingegneria Elettronica” degree in 1987 from the University of Padova, Padova, Italy. From 1990 to 1992, he was with the University of California at Berkeley, where he received the MS degree in 1992. He received the Doctorate degree in “Ingegneria Elettronica e dell’Informazione” from the University of Padova in 1992. In 1992 he joined the Dipartimento di Elettronica e Informatica of the University of Padova. Starting from November 1st 1998, he was associate professor in Communications and Signal Processing in the same Department. Since November 2001, he has been an associate professor in the “Dipartimento di Ingegneria Elettrica, Gestionale e Meccanica” of the University of Udine. Starting December 2003, he is now a professor in the same department. His interests are in the field of multidimensional signal processing, video signal coding, fractal theory and image coding.

**Riccardo Bernardini** was born in Genova (Italy) in 1964. He received the “Laurea in Ingegneria Elettronica” degree from the University of Padova in 1990. Since then he has been with the Dipartimento di Elettronica e Informatica of the University of Padova with a scholarship of the Consorzio Padova Ricerca, and, from November 1992 to November 1995, as a Ph.D. student. He spent the last year of his Ph.D. at formerly AT&T Bell Labs (Murray Hill). From April 1996 to April 1997 he was in EPFL (Lausanne) as a Postdoctoral Fellow. Now he is working as researcher in the Dipartimento di Ingegneria Elettrica, Gestionale e Meccanica of the University of Udine. His main interests are in the area of multidimensional signal processing, wavelets, filter banks and robust transmission.

**Marco Durigon** was born in 1976. He graduated in Telecommunication Engineering from the University of Padova, Italy, in 2002. He is currently a Ph.D. student at the University of Udine, Italy. His research interests include image and video lossless compression, video streaming in lossy environments and multiple description coding.