MULTI-LOOP SCALABLE MPEG-2 VIDEO CODERS

Sławomir Maćkowiak

Poznań University of Technology,
Institute of Electronics and Telecommunication,
Piotrowo 3A, 60-965 Poznań, Poland
Phone: +48 61 66 52 171 Fax: +48 61 66 52 572
smack@et.put.poznan.pl

Abstract. In this paper, a structure of a hybrid highly scalable video encoder with fine granularity scalability in a scheme based on a modified version of the classic MPEG-2 scalable encoder is proposed. The assumption is that high level of compatibility with the MPEG video coding standards would be ensured. In particular, it is assumed that the low-resolution base layer bitstream is fully compatible with the MPEG-2 standard. The proposed scalable coder exhibits good satisfactory coding performance. These results are much better than those usually obtained for standard scalable profiles of MPEG-2.

Keywords: MPEG-2, scalable video coding, spatio-temporal scalability, fine granularity scalability

1. Introduction

Video and multimedia communication services have been developing rapidly in the last years. Their availability depends strongly on communication network infrastructure. High bitrates needed for video transmission impose severe requirements on communication networks. The existing networks are very inhomogeneous. Links are characterized by different bitrates, by different error rates, by different levels of Quality of Service (QoS). Different levels of Quality of Service are often related to different available transmission bitrates.

On the other hand, service providers demand the data to be broadcasted only once to a group of users accessed via heterogeneous links. Different networks are indeed different groups of users who have varying expectations. The same content may reach different group of users. For this purpose, the transmitted bitstream should be partitioned into some layers.

Scalability of video means the ability to achieve a video of more than one resolution or quality simultaneously. Scalable video coding involves generating a coded representation (bitstream) in a manner that facilitates the derivation of video of more than one resolution or quality from this bitstream.

For a given overall decoded video quality, scalable coding is not acceptable in common applications, if the bitrate is significantly greater than the bitrate achieved in single-layer coding.

The existing video compression standard MPEG-2 [1,2] defines scalable profiles, which exploit classic Discrete Cosine Transform-(DCT)-based schemes with motion compensation. Unfortunately, spatial scalability as proposed by the MPEG-2 coding standard is inefficient because the bitrate overhead is too large. Additionally, the solutions defined in MPEG-2 do not allow flexible allocation of the bitrate. There exists a great demand for flexible bit allocation to individual layers, i.e. for fine granularity scalability (FGS) [3], which is also already proposed for MPEG-4, where the fine granular enhancement layers are intraframe encoded.

The scalability is expected to find many applications. The goal of scalable coding is to provide interoperability between different services and to flexibly support receivers characterized by different display capabilities. For example, flexible support of multiple resolutions is of particular importance in interworking between High Definition Television (HDTV) and Standard Definition Television (SDTV), in which case it is important for a HDTV receiver to be compatible with a SDTV application. Compatibility of the receivers can be achieved by means of scalable coding of the HDTV source. Moreover, the transmission of two independent bitstreams to the HDTV and SDTV receivers can be avoided.

The importance of scalability is being more and more recognized as more attention is paid to video transmission in error-prone environments, such as wireless video transmission systems [4]. A video bitstream is error-sensitive due to extensive employment of variable-length coding. A single transmission error may result in a long, undecodable string of bits. It has been shown that an efficient method of improving transmission error resilience is to split the coded video bitstream into a number of separate bitstreams (layers) transmitted via channels with different degrees of error protection. The base layer is better protected, while the enhancement layers exhibit a lower level of protection. A receiver is able to reproduce at least low-resolution pictures if Quality of Service decreases.

In this paper, a structure of a hybrid highly scalable video encoder with fine granularity scalability in a scheme based on a modified version of the classic MPEG-2 scalable encoder coder is proposed. The goal is to achieve total bitrate of all layers of scalable coding possibly close to the bitrate of single-layer coding. The assumption is that high level of compatibility with the MPEG video coding standards would be ensured. In particular, it is assumed that the low-resolution base layer bitstream is fully compatible with the MPEG-2 standard.

2. Spatio-Temporal Scalability

Most scalability proposals are based on one type of scalability. The universal scalable coding has to include different types of scalability [5]. A single scalable video technique cannot serve a broad range of bitrates in networks (e.g. from a few kbps to several Mbps) or a wide selection of terminals with different characteristics.

Among various possibilities, the combination of spatial and temporal scalability called spatio-temporal scalability seems very promising [13]. Spatio-temporal decomposition allows to encode the base layer with a smaller number of bits because the base layer corresponds to reduced information. Here, the term of spatio-temporal scalability describes a functionality of video compression systems where the base layer (low resolution layer) corresponds to frames with reduced spatial and temporal resolution. An enhancement layer (high resolution layer) is used to transmit the information needed for restoration of full spatial and temporal resolution. Fig. 1 shows an example of a video sequence structure obtained after spatio-temporal decomposition, for three layers. For the sake of simplicity, spatio-temporal scalability will be reviewed for the simplest case of a two-layer encoder, i.e. a system that produces one low resolution bitstream and one high resolution bitstream.

Spatio-temporal scalability has been proposed in several versions, in particular:

- with 3-D spatio-temporal subband decomposition [6-8],
- with 2-D spatial subband decomposition and partitioning of B-frames data [8,9],
- exploiting as reference frames the interpolated low resolution images from the base layer [10,11].

The following basic video sequences are processed in the presented encoder:

- The low resolution sequence with reduced picture frequency and reduced horizontal and vertical resolution.
- The high resolution sequence with original resolutions in time and space.

The advantage of this solution is that the low resolution layer is an independently coded layer and does not use any information from the other layer.

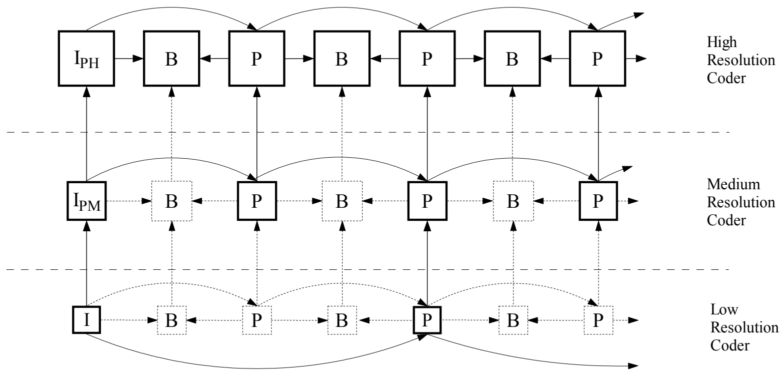

Fig. 1. Exemplary structure of the video sequence.

Temporal scalability is achieved using bi-directionally predicted frames, or B-frames. B-frames are disposable, since they are not used as reference frames for the prediction of any other frames. This property allows B-frames to be discarded without destroying the ability of the decoder to decode the sequence and without adversely affecting the quality of any subsequent frames, thus providing temporal scalability.

In this paper, temporal resolution reduction is achieved by partitioning the stream of B-frames: every second frame is skipped in the low resolution layer.

The choice of the spatial decimator and interpolator has substantial impact on the overall coding efficiency. In the experiments, for decimation, an FIR lowpass zero-phase 7-tap filter has been applied. Of course, these are only exemplary filter parameters that have been applied for the experiments described in this document.

One can use other filters providing some trade-off between aliasing attenuation and spatial response length.

3. Coder Structure

The proposed encoder consists of a low resolution encoder and a high resolution encoder (Fig. 1). The low resolution encoder is implemented as a motion-compensated hybrid MPEG-2 encoder of the Main Profile@Main Level (MPEG-2 MP@ML).

The high resolution encoder is a modification of the MPEG-2 encoder. The motion-compensated predictor employed in the high resolution layer uses a modified prediction proposed by Łuczak for B-frames [10,11]. As an extension to the MPEG-2 compression technique, in the modified prediction those B-frames which correspond to B-frames from the base layer can be used as reference frames for predicting other B-frames in the enhancement layer.

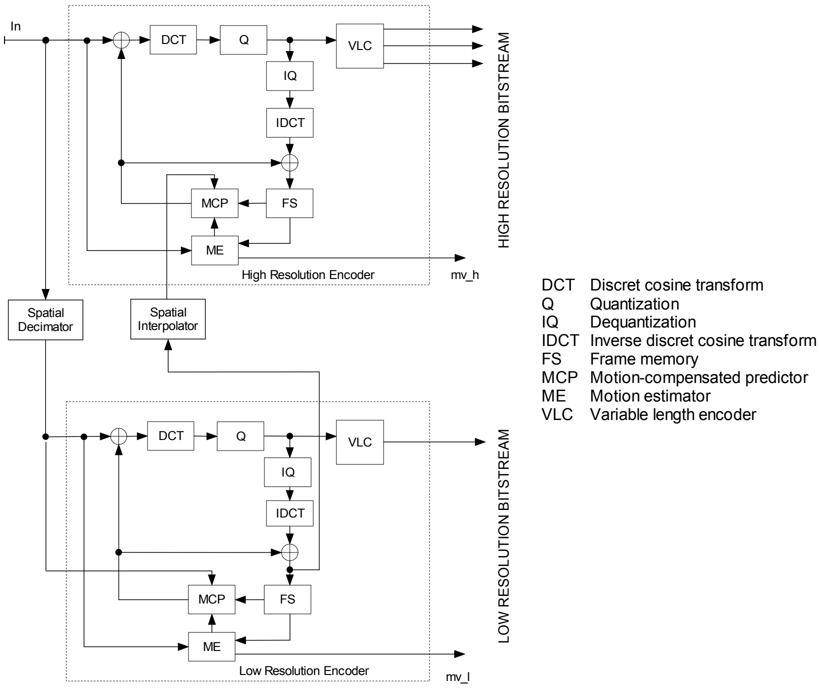

Fig. 2. General coder structure.

High efficiency of scalable encoding requires using some information from the base layer in the high resolution encoder. The high resolution layer encoder uses the interpolated decoded frame from the base layer in the prediction of the full resolution frame. It is assumed that the base layer represents a video signal with half the spatial resolution. Therefore one macroblock in the base layer corresponds to four macroblocks in the enhancement layer

The proposed encoder applies independent motion compensation loops in all layers [13]. The motion vectors mv_b for the low resolution frames are estimated independently from the mv_e, which are estimated for the high resolution images

Since every second frame is skipped in the low resolution encoder, motion estimation and compensation processes in the low and high resolution layers are performed for the frames from different time moments.

Two independent motion estimation and compensation processes yield the best results because due to the estimation and compensation in the low resolution layer, there is coarse motion compensation for slowly moving objects in the high resolution layer. The second motion estimation and compensation give more precise prediction.

4. Fine Granularity Scalability with Motion Compensation

Fine granularity is obtained by partitioning transform coefficients between layers. Except from headers and motion vectors, the bitstreams can be arbitrarily split into layers and multi-layer fine granularity can be achieved [13]. All header data and the enhancement motion vectors mv_h may be treated as basic granules [11]. The next granules are constituted by DCT coefficients that are encoded as (run, level) pairs, as described in the MPEG-2 standard. The lower layer contains N_m first (run, level) pairs for individual blocks. The control parameter N_m influences bit allocation to layers. The bitrates in subsequent layers can be controlled individually. To some extent, nevertheless, each additional layer increases the bitrate overhead because at least slice headers should be transmitted in all layers in order to guarantee resynchronization after an uncorrected transmission error. The total bitstream increases by about 3% per each layer obtained using data partitioning.

The drawback of this strategy is accumulation of drift. Drift is generated by partitioning the high resolution bitstream. Moreover, when the enhancement layer bitstream is corrupted by errors during transmission, the enhancement layer DCT coefficients cannot be properly reconstructed due to the loss of DCT information. This causes drift between the local decoder and remote decoder. It means that the decoding process exploits only the base layer bitstream.

In some applications, drift is not a significant problem. In particular, the MPEG-2-related encoders mostly use relatively short independently coded Groups of Pictures (GOPs), thus preventing drift from significant accumulation. In the author’s solution, drift accumulation is also reduced because the total bitstream is divided into two drift-free parts, i.e. the low and the high resolution bitstream. Drift propagates within one part only when fine granularity is applied to a given bitstream. Furthermore, drift in the high resolution part may be reduced by more extensive use of the low resolution images as reference.

5. Experimental Results

In order to evaluate compression efficiency, a verification model has been written in the C++ language and is currently available for progressive 4CIF (704 x 576), 50 Hz, 4:2:0 video test sequences. This software also provides an implementation of the MPEG-2 encoder, which has been cross-checked with the MPEG-2 verification model [12].

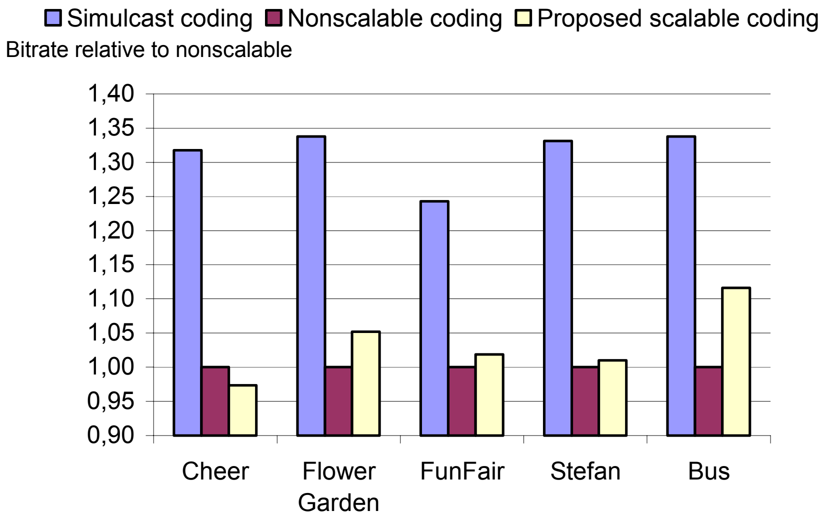

Fig. 3. Approximate bitrate comparison for scalable, nonscalable (single-layer) and simulcast coding at 3 Mbps for non–scalable MPEG–2 coding of SDTV signal.

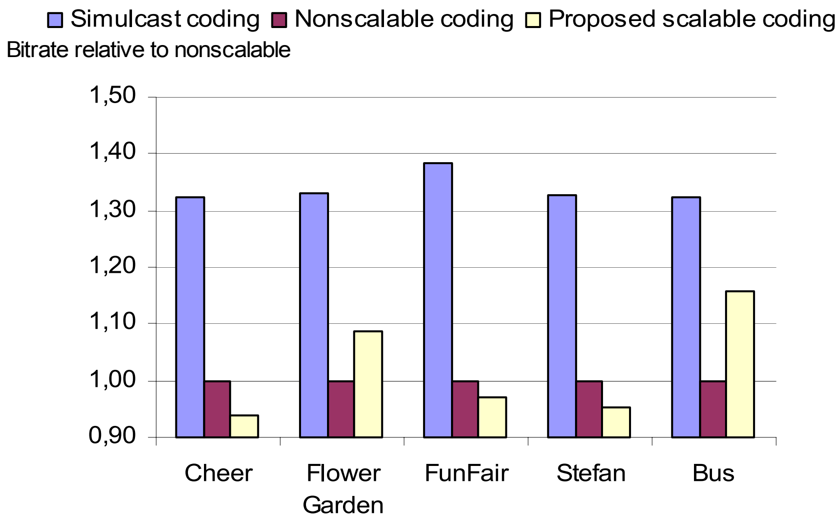

Fig. 4. Approximate bitrate comparison for scalable, nonscalable (single-layer) and simulcast coding at 4 Mbps for non–scalable MPEG–2 coding of SDTV signal.

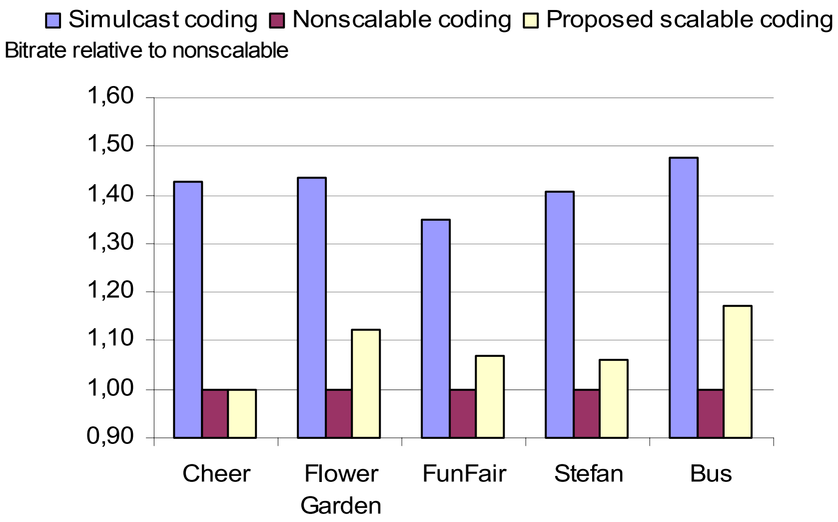

Fig. 5. Approximate bitrate comparison for scalable, nonscalable (single-layer) and simulcast coding at 5 Mbps for non–scalable MPEG–2 coding of SDTV signal.

Simulations have been carried out for constant quality coding, for three bitrates, i.e. 3 Mbps, 4 Mbps and 5 Mbps, for non–scalable MPEG–2 coding of SDTV signals.

In simulcast coding, each bitstream of video is associated with a certain resolution or quality and is encoded independently. Thus, any bitstream can be decoded by a single-layer decoder. The total bitrate required for transmission of encoded streams is the sum of bitrates of these streams.

The results from Figs 3,4 and 5 prove high efficiency of the two-layer encoder. With the same bitrate as in the MPEG-2 non-scalable profile, the proposed scalable encoder ensures almost the same quality. Bitrate overhead due to scalability is about 0% - 18%.

6. Conclusion

Described is a modified MPEG-2 scalable codec with fine granularity scalability. The basic features of the two-loop coder structure are:

- mixed spatio-temporal scalability with fine granularity scalability,
- independent motion estimation for each motion-compensation loop, i.e. for each spatio-temporal resolution layer,
- BR/BE-frame structure.

The scalable coder exhibits good satisfactory coding performance. These results are much better than those usually obtained for standard scalable profiles of MPEG-2. Scalable coder complexity is similar to that of the simulcast structure.

References

1. ISO/IEC IS 13818-2 / ITU-T Rec. H.262, "Generic coding of moving pictures and associated audio, part 2: video", November 1994.
2. Haskell B.G., Puri A., Netravali A.N., Digital video: an introduction to MPEG-2, New York, Chapman & Hall, September 1996.
3. Li W. „Overview of Fine Granularity Scalability in MPEG-4 Video Standard", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11 Issue: 3, pp. 301 -317, March 2001.
4. Girod B., Färber N., "Wireless Video," in Reibman A., Sun M.-T. (eds.), Compressed Video over Networks, Marcel Dekker, 2000.
5. Ohm J.-R., Li W., Ling F., Chun K.-W., Li S., Wu F., Zhang Y.-Q., van der Schaar M., Jiang H., Chen X., Vetro A., Sun H., Wollborn M., "Summary of Discussions on Advanced Scalable Video Coding", ISO/IEC JTC1/SC29/WG11 M7016, Singapore, March 2001.
6. Taubman D. and Zakhor A., "Multirate 3-D subband coding of video," IEEE Trans. Circ. Syst. Video Techn., vol. 3, pp. 572-588, September 1994.
7. Kim B.-J., Xiong Z., Pearlman W.A., "Low Bit-Rate Scalable Video Coding with 3-D Set Partitioning in Hierarchical Trees (3-D SPIHT)", IEEE Transactions on Circuits and Systems for Video Technology, Volume: 10, No. 8, December 2000.
8. Domański M., Łuczak A., Maćkowiak S., Świerczyński R., "Hybrid coding of video with spatio-temporal scalability using subband decomposition", Signal Processing IX: Theories and Applications, pp. 53-56, Typorama, 1998.
9. Domański M., Łuczak A., Maćkowiak S., „Spatio-Temporal Scalability for MPEG" IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10, No. 7, pp. 1088-1093, October 2000.
10. Domański M., Łuczak A., Maćkowiak S., "On improving MPEG Spatial Scalability", Proceedings of the IEEE International Conference on Image Processing ICIP'2000, Vancouver, pp. II-848 - II-851, 2000.
11. Domański M., Maćkowiak S., "Modified MPEG-2 Video Coders with Efficient Multi-Layer Scalability", Proceedings of the IEEE International Conference on Image Processing ICIP'2001, vol. II, pp. 1033-1036, Thessaloniki, 2001.
12. MPEG Software Simulation Group, MPEG-2 Encoder Decoder version 1.2, July 19, 1996, ftp://ftp.mpeg.org/pub/mpeg/mssg.
13. Maćkowiak S., "Scalable Coding of Digital Video", Doctoral dissertation, Poznań University of Technology, Poznań 2002.