Łukasz Błaszak
Marek Domański
Sławomir Maćkowiak

Institute of Electronics and Telecommunications
Poznań University of Technology
Piotrowo 3A, 60-965 Poznań
e-mail: [lblaszak, domanski, smack] @et.put.poznan.pl

X NATIONAL SYMPOSIUM OF RADIO SCIENCE
Poznań, March 14-15, 2002

SCALABLE VIDEO COMPRESSION FOR WIRELESS SYSTEMS

Abstract: **Scalability became an important functionality of video codecs employed in wireless transmission systems. The paper describes a generic multi-loop coder structure suitable for mixed spatial and temporal scalability combined with fine granular SNR scalability. The encoder exhibits extended capabilities of adaptation to network throughput. The MPEG-2 and H.263 video coding standards are used as a reference but the results are also applicable to the MPEG-4 and H.26L systems with minor modifications.**

1. INTRODUCTION

The well-known classic video coding schemes have been developed and standardized mostly for wireline transmission systems. Recently, the emergence of broadband wireless networks and their expected tremendous growth has changed the requirements for video coding systems. The classic video compression techniques have been well tuned for quasi-error free channels. Now, these techniques have to be adapted to unreliable wireless systems with their fades and transmission errors.

Another important issue is related to the bandwidth fluctuations being typical for wireless systems. Again, the classic but very efficient video coding techniques like H.263 [1] and MPEG-2/4 [2,3] are designed for constant-bitrate transmission systems or for variable bitrate environment where a coder controls the bitrate according to the video content changes. In the wireless systems, the bandwidth fluctuations influence the bitrates available for video streaming [4,5]. Therefore, the video coder functionality of scalability is necessary. Scalability means that a video data bitstream is partitioned into layers in such a way that the base layer is independently decodable into a video sequence with reduced spatial resolution, temporal resolution or signal-to-noise ratio (SNR). Enhancement layers provide additional data necessary for video reproduction with higher spatial resolution, temporal resolution or signal-to-noise ratio. This functionality is called spatial, temporal or SNR scalability, respectively, as defined by video coding standards: MPEG-2 [2] and MPEG-4 [3]. In the case of bandwidth decrease, the receiver decodes only the base part of the bitstream.

Moreover, scalable video codecs are suitable for video transmission over heterogeneous communication networks characterized by various available levels of Quality of Service (Fig.1). The service providers demand that the data are broadcasted once to a group of users accessed via heterogeneous links and this demand can be ensured due to application of scalable video coding systems.

Unfortunately, the scalable coding schemes provided by MPEG-2 and MPEG-4 are not satisfactory in some aspects, like coding efficiency and bandwidth adaptation flexibility. Although MPEG-4 [3] has adopted Fine-Granularity-Scalability (FGS) as a tool for precise tuning a bitstream to channel payload, its coding efficiency is not satisfactory because of lack of temporal prediction in the enhancement layer.

There were many attempts to improve spatially scalable coding of video. Great expectations are related to the inherently scalable wavelet-based techniques [18, 19], which have been successfully exploited for flexibly scalable still image compression in the new international standard JPEG 2000 [20]. Unfortunately, in video coding, motion-compensated wavelet-based schemes are not as successful as in still image compression, because they are usually still not able to outperform coding efficiency of classic block-based motion-compensated hybrid techniques at reasonable processing power and memory requirements.

Another approach has been proposed by the authors who introduced a concept of spatio-temporal scalability being a mixture of spatial and temporal scalability [14,15,16]. This approach was quit successful but mixing this technique with

FGS provides even more flexible structure of the encoder.

The FGS technique enables the receiver to decode properly a video stream even it is received with limited bitrate, e.g. because of temporal bandwidth fluctuations.

The encoder proposed in this paper has extended ability of adaptation to network throughput variations.

The encoder is able to produce three layers. Every layer is dedicated to different bandwidth channel and thanks to FGS has also some ability of bitrate adaptation below the current layer bitrate.

The goal of the work is to achieve total bitrate related to both layers of a scalable bitstream possibly close to the bitrate of single-layer coding. The assumption is that high level of compatibility with the MPEG-2/H.263 video coding standards would be ensured. In the paper, the MPEG-2 video coding standard is used as reference but the results are also applicable to the MPEG-4 and H.26L systems with minor modifications. In particular, it is assumed that the low-resolution base layer bitstream is fully compatible with the MPEG-2 standard or H.263 recommendation.

There are three type of scalability described in references: temporal scalability, spatial scalability and SNR scalability [12,13]. The proposed encoder is hybrid motion-compensated encoder with spatial and temporal scalability.

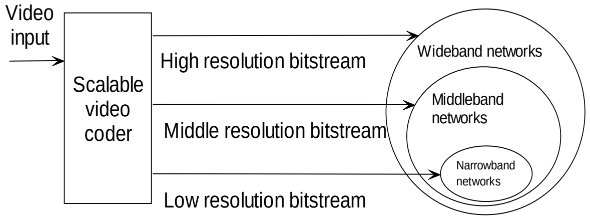

Fig.1. Multi-layer scalable video coding system in heterogeneous network that consists of sub-networks characterized by different throughputs.

2. SPATIO-TEMPORAL SCALABILITY

The proposed scalable coder consists of two or three motion-compensated hybrid coders (Fig. 2) that encode a video sequence and produce two or three bitstreams corresponding to two or three different levels of spatial and temporal resolution. In this structure, it is also possible to encode a video sequence without temporal decomposition (as denoted in Figs. 2 and 3 using dotted lines).

In two-layer version, the lower layer has reduced spatial and temporal resolutions. In the H.263-based coder the temporal resolution reduction is achieved by partitioning of the second P-frames (B-frames for the MPEG-2 based coder): each second frame is not included into the low resolution layer. The full resolution coder produces a bitstream which can be partitioned appropriate to the network throughput.

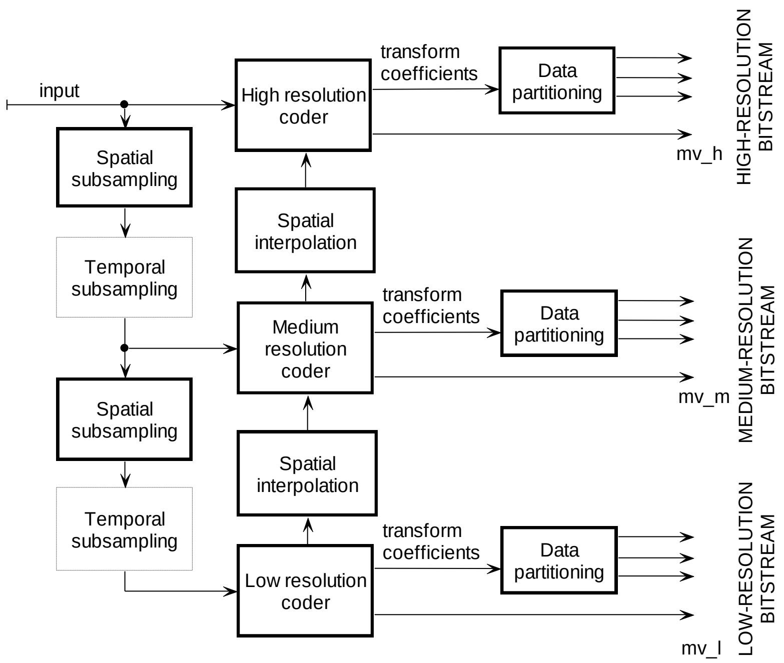

Fig. 2. A generic block diagram of the scalable coder as proposed in this paper.

.

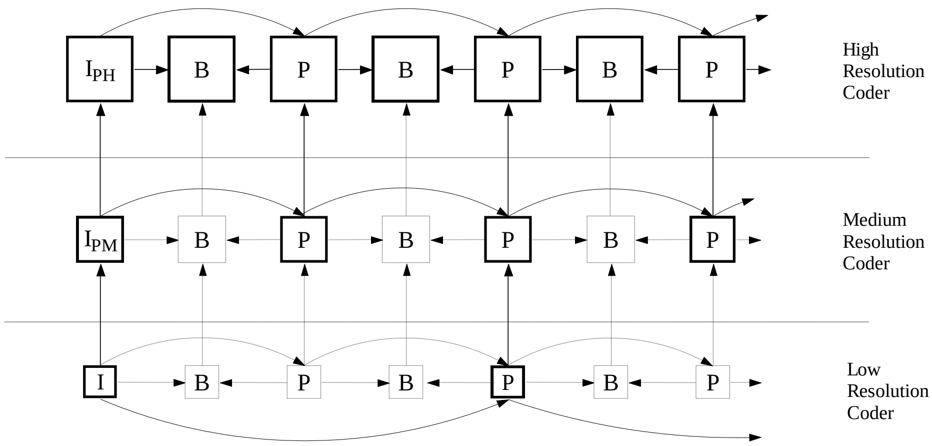

Fig.3. Exemplary structure of the video sequence.

3. SCALABLE CODER

The coder produces a scalable bitstream, which consists of two or three layers: the low resolution layer with reduced both temporal and spatial resolution and the high resolution layer corresponding to full resolution in time and space. In the three layered system, bitstream produced in the low resolution coder is described by fully compliant MPEG-2 or H.263 standard syntax. The proposed structure covers different spatial resolutions: QCIF in the low resolution layer, CIF in the middle resolution layer and 4CIF in the high resolution layer.

The motion vectors *mv_l* for the low-resolution images are estimated independently from those vectors estimated for the higher resolution layers. Motion vectors *mv_l* are transmitted for the low resolution layer. The other parts of the coder produce bitstreams related to other layers. In particular, motion is estimated for full-resolution images and full-frame motion compensation is performed. Therefore the number of motion vectors in the upper layer is four times that of the actual layer.

For each layer, in order to adapt to fluctuations of the channel bandwidth, the FGS technique can be applied [22].

4. EXPERIMENTAL RESULTS

In order to test the proposed structure a verification model has been prepared as software written in C++ language. The most important feature is its flexibility allowing tests of different variants of the coding algorithm. The software is currently available for progressive sequences. The performance of the two loop structure has been tested for various bitrates.

Three basic series of experiments have been performed for constant quality coding, corresponding to approximately 500kbps for non-scalable H.263 and 4Mbps for non–scalable MPEG–2 coding of SDTV signals:

a) H.263-based experiments have made with CIF sequences with and without GOPs in the enhancement layer. The coder used was built on the H.263 baseline coder.

b) MPEG-2-based experiments used 4CIF sequences with the GOP length of 12 for both layers.

The overall coding performance is summarized in Table 1.

The FGS technique can be used to bitrate reduction but it also produces drift [21]. Nevertheless drift is not a significant problem in many applications [15]. In particular, the MPEG-2-related coders use mostly relatively short independently coded groups of pictures (GOPs), thus preventing drift from significant accumulation.

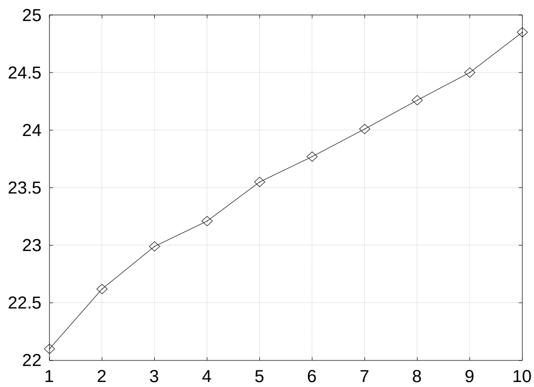

Fig. 4. Average values of PSNR [dB] plotted versus the number of nonzero DCT coefficients allocated. Results for the *Cheer* test sequence, the high-resolution bitstream (H.263 GOP=6, 630Kbps).

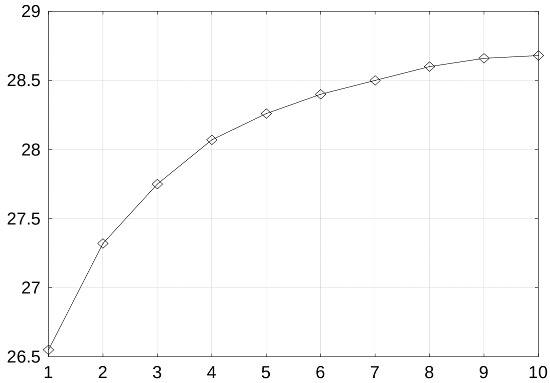

Fig.5. Average values of PSNR [dB] plotted versus the number of nonzero DCT coefficients allocated. Results for the *Cheer* test sequence, the high-resolution bitstream (MPEG-2 GOP=12, 2.35Mbps).

5. CONCLUSION

In this paper, three modifications of the two-layer video codec have been tested. All three variants provide spatio-temporal scalability mixed with data partitioning. The experimental data prove that the coder is able to produce bitstreams tuned to the bitrates of the network. The experimental results prove that acceptable bitrate overheads can be achieved by flexible combinations of spatial/temporal/SNR scalability with fine granularity. Such combination of scalability modes is very advantageous for practical applications of the wireless transmission.

Table I. Experimental results

		H.263 without GOP in high resolution layer		**H.263 (GOP=6)**		**MPEG-2 (GOP=12)**	
		Cheer	*Bus*	*Cheer*	*Bus*	*Cheer*	*Bus*
Single-layer coder	Bitstream [Kbps]	402.33	414.61	651,63	482,64	3910	3930
	Average luminance PSNR [dB]	25.62	27.57	26.87	27.77	30.66	33.54
Proposed scalable coder	Low resolution layer bitstream [Kbps]	98.28	99.70	97.18	99.61	1260	1270
	Low resolution layer average PSNR [dB] for luminance	23.19	26.06	22.92	25.61	30.43	34.42
	High resolution layer bitstream [Kbps]	315.41	302.61	629,25	421.08	2410	3280
	Average PSNR [dB] for luminance recovered from both layers	25.58	27.59	26.87	27.77	30.66	33.59

ACKNOWLEDGEMENT

The work has been partially supported by the IST Programme of the EU under contract number IST-2000-26467 (MASCOT).

6. REFERENCES

[1] ITU-T, *Video coding for narrow telecommunication channels at < 64kbit/*s, Recommendation H.263, 1996.

[2] ISO/IEC International Standard 13818, *Information Technology - Generic Coding of Moving Pictures and Associated Audio Information.*

[3] International Organization For Standardization, “MPEG-4 Video Verification Model v. 18.0”, ISO/IEC JTC1/SC29/WG11 N3908, Pisa, January 2001.

[4] D. Wu, Y. Hou, Y. Zhang, “Scalable video coding and transport over broad-band wireless networks,” *Proc. of the IEEE*, vol. 89, pp. 6-20, January 2001.

[5] L. Hanzo, P. Cherriman, J. Streit, *Wireless Video Communications*, IEEE Press, New York, 2001.

[6] T. Tsunashima, J. Stampleman, V. Bove, ”A scalable motion -compensated subband image coder,” *IEEE Trans. on Communication,* vol. 42, pp. 1894-1901, 1994.

[7] F. Bosveld, ”Hierarchical video compression using SBC,” Ph.D. dissertation, Delft University of Technology, Delft 1996.

[8] H. Gharavi, W.Y. Ng, ”H.263 Compatible Video Coding and Transmission,” in *Proc. First International Workshop on Wireless Image/Video Communication*, pp. 115-120, Loughborough 1996.

[9] Senbel, H. Abdel-Wahab, ”Scalable and robust image compression using quadtrees,” *Signal Processing: Image Communication*, vol. 14, pp. 425-441, 1999.
[10] Shen, E. Delp, ”Wavelet based rate scalable video compression,” *IEEE Trans. Circuits Syst. Video Technology*, vol. 9, pp. 109-122, February 1999.
[11] U. Benzler, ”Scalable multi-resolution video coding using a combined subband-DCT approach”, *Picture Coding Symposium 1999*, pp. 17-20.
[12] M. Domański, A. Łuczak, R. Świerczyński, Skalowalne kodowanie sygnałów wizyjnych, Poznańskie Warsztaty Telekomunikacyjne, Poznań 1997, 1997.
[13] International Telecomunication Union, Radiocommunication Study Groups, Document 11-3/103-E, MPEG-2 spatial scalable video coding.
[14] M. Domański, A. Łuczak, S. Maćkowiak, „*Spatio-Temporal Scalability for MPEG*” IEEE Transactions on Circuits and Systems for Video Technology Vol. 10 No. 7, pp. 1088-1093, October 2000.
[15] M. Domański, A. Łuczak, S. Maćkowiak, "*On improving MPEG Spatial Scalability*", Proceedings of 2000 International Conference of Image Processing ICIP Vancouver, Kanada, pp. II-848 - II-851.
[16] M. Domański, A. Łuczak, S. Maćkowiak, "*Spatio-Temporal Scalability using modified MPEG-2 predictive video coding*", Proceedings of European Signal Processing X, Tampere, Finlandia 2000, pp. 961-964.
[17] R. Aravind, M. Reha Civanlar, A. Reibman, “Packet loss resilience of MPEG-2 scalable video coding algorithms,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 426-435, October 1996.
[18] M. Ghanbari, *Video coding*, London, IEE 1999.
[19] P. Topiwala (ed.), *Wavelet image and video compression*, Boston, Kluwer 1998.
[20] ISO/IEC IS 15444-1 / ITU-T Rec. T.800, “JPEG 2000 image coding system.”
[21] M. Domański, S. Maćkowiak, „Modified MPEG-2 video coders with efficient multi-layer scalability,” *Int. Conf. Image Processing ICIP 2001*, IEEE, vol. II, pp. 1033-1036, Thessaloniki, October 2001.
[22] J.-R.Ohm, M. van der Schaar, *Scalable Video Coding*, Tutorial material, *Int. Conf. Image Processing ICIP 2001*, IEEE, Thessaloniki, October 2001.