A SIMPLE TECHNIQUE FOR SNR-SCALABLE VIDEO CODING USING MACROBLOCK SPIRAL RE-ORDERING

Łukasz Błaszak, Marek Domański

Poznań University of Technology, Institute of Electronics and Telecommunications, Poznań, Poland

ABSTRACT

Scalable representation of video mostly consists of some layers that may correspond to various spatial and temporal resolution. In order to match exactly the channel throughput, fine granularity scalability may be used. It is usually achieved by quality (SNR) scalability that scales the bitstream of an enhancement layer. In this paper, a very simple and straightforward technique of fine granularity scalability is considered. It consists in partitioning of a stream of encoded macroblocks. The new issue is re-ordering of macroblocks according to the spiral scan prior to compression. This re-ordering improves subjective video quality perceived by a viewer that watches scaled video. The advantages of such a technique are: simplicity of the scheme and preservation of the standard bitstream syntax.

1. INTRODUCTION

Scalability is a research and application issue in video coding since quite long time. Nevertheless, recent wide-spreading of wireless and heterogeneous networks yields scalability to become more and more challenging issue. Huge number of scientific papers and international standardization activities prove the importance of scalability. Currently, ISO and ITU are working on an international standard of scalable video coding [1,2] that will be an extension of the AVC/H.264 video compression standard [3]. The technology for scalable compression is described in Scalable Video Model (SVM) [2] that is likely to be the basis for the standard mentioned above. This technology comprises sophisticated tools for spatial, temporal and quality scalability together with motion-compensated temporal filtering.

Here, an alternative low-complexity tool for quality scalability is proposed. This tool, also proposed and considered for standardization [4,5], is based on partitioning of the stream of the macroblocks in a spatial and/or temporal enhancement layer. Such enhancement layer is constituted by data that are necessary to reconstruct video with higher resolution. Partitioning of this bitstream splits it into a quality enhancement layer base-quality layer (Fig. 1). The idea is to represent a part of macroblocks in the enhancement layer. As an arbitrary percentage of all macroblocks may be transmitted in the quality enhancement layer, it is in fact a very simple tool suitable to implement fine grain scalability.

The scalable video coder structure from Fig. 1 is quite general and fits to the current Scalable Video Model with motion-compensated temporal filtering (MCTF) [1,2,6] as well as to simpler proposals where temporal scalability is obtained by B-frame skipping [7-9].

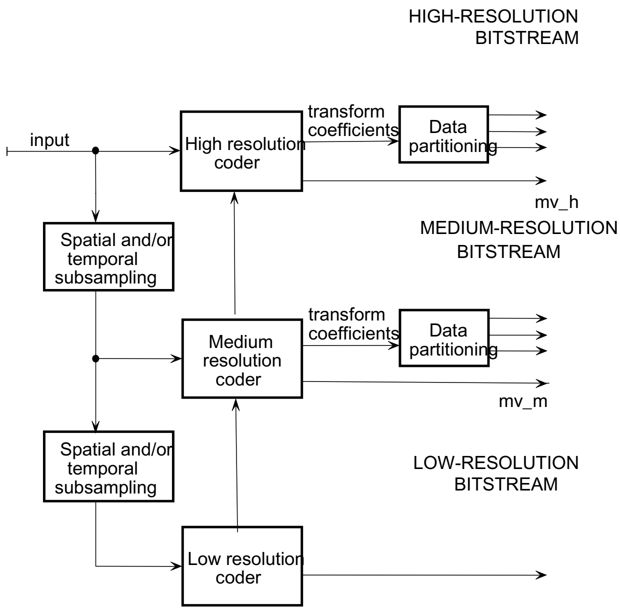


Fig. 1. General model of a scalable video encoder with mixed scalability.

The new idea is to re-order the stream of macroblocks according to the spiral scan (Fig. 2b) rather then to the traditional raster scan (Fig. 2a). This idea has been already mentioned by the authors in [4,10] but now it will be considered in details.

0-7803-9332-5/05/$20.00 ©2005 IEEE

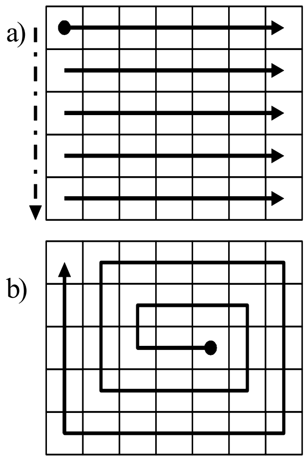


Fig.2. a) Raster scan of macroblocks,
b) spiral scan of macroblocks.

The idea of spiral scan is very similar to that of water-ring scan [11,12] that has been also proposed for fine grain scalability. Water ring scan has been proposed as a modification of the FGS bitplane coding [11,12] while, here, the spiral scan is considered together with modified context-based predictions in AVC/H.264.

The goal of this paper is to show how the spiral scan of macroblocks may be used in order to obtain quality (SNR) scalability using minimum complexity effort and with no modification of the hitherto used bitstreams. Full compatibility with AVC/H.264 [9] is guaranteed in the base layer which is also important standardization requirement [1].

2. SPIRAL SCAN

The idea of the spiral scan is related to the properties of the human visual system such as focusing on a part of the image instead of the whole image. Nothing to say, that the human observer concentrates himself or herself on the most interested region of the image. In the most of the cases, a cameraman shots video in such a way that most interesting region is centered in a picture. Therefore, the outer parts of images may be represented with lower bitrate lower quality if the overall bitrate must be reduced. Often, the reduced quality of the outer parts of images is even not perceived by a viewer. This effect may be exploited in scalable video coding, where low-bitrate bitstreams are embedded in a high-bitrate bitstream. In such a case, the high quality image portion should be located in the area of interest rather than on the top of a picture (Fig. 3). For this purpose, the spiral scan of macroblocks may be used and the respective bitstream may be cut after arbitrary number of macroblocks, thus giving good quality in the area of interest in images (Fig. 3b). In the decoder, the other macroblocks are reconstructed from a low-quality base layer. The standard macroblock order would yield an impractical location of the high quality area on the top of images (Fig. 3a).

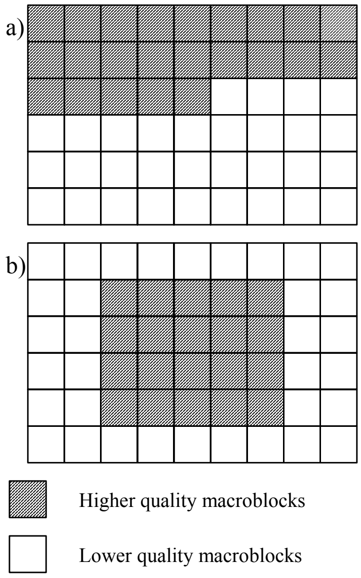


Fig.3. A decoded image using
a) raster scan, b) spiral scan.

In the efforts toward an efficient scalable video coding standard, the AVC/H.264 serves as reference for compatibility and coding efficiency comparisons. In AVC codec, coding of macroblocks depends on various contexts that are used to predict numerous types of information. A context consists of neighboring macroblocks or blocks (Fig.4).

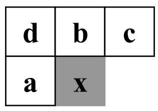


Fig. 4. The standard context.

As compared to the raster scan, the contexts must be modified according to the current processing direction related to the spiral scan.

For an AVC/H.264-based codec with spiral scan, the predictions of all data elements have to be modified. The following prediction tools need adaptation to spiral scan:

- prediction of (4×4)-pixel, (8×8)-pixel and (16×16)-pixel luma blocks for intra frame coding,
- prediction of chroma blocks for intra frame coding,
- motion vectors prediction for all block sizes,
- prediction of macroblock encoding parameters,
- context prediction for CABAC and CAVLC coding:
    - block-based prediction,
    - bit-based prediction.

A typical example of such context adaptation is shown in Fig. 5.

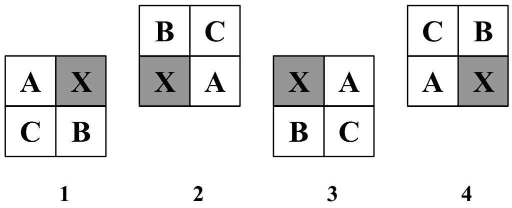
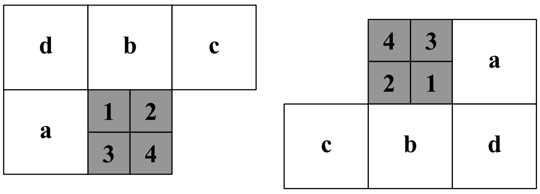


Fig. 5. Possible neighbourhood for intraframe (16×16) prediction and CABAC coding.

Modified neighborhood implies also modification of coding order of blocks inside a macroblock. An example of possible order is shown in Fig. 6.



Standard order of block coding for 8x8 block size.

Example of modified order of block coding for 8x8 block size.

Fig. 6. Block order inside microblocks.

The above mentioned changes do not influence standard bitstream syntax. Of course, both the encoder and the decoder must use the same scan. If both scans would be allowed, an additional flag should be transmitted in a bitstream, e.g. in picture headers.

There is no problem with establishing slices within an image with the spiral scan (Fig. 7).

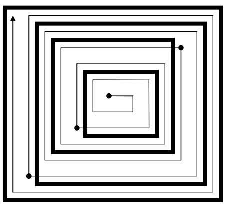


Fig. 7. Typical structure of slices in an image with the spiral scan.

3. SNR SCALABILITY USING THE SPIRAL SCAN

The spiral scan may be used in a very straightforward way in order to obtain scalability of the bitstream. One may simply cut the enhancement stream of macroblocks that represent an image. In the decoder, the missing macroblocks are reconstructed from the low-resolution base layer. Thus, reduced is quality of the outer portions of an image (Fig. 3b). It is worthy to stress that such an implementation of scalability is extremely simple, does not influence coding efficiency, and does not increase the computational effort by encoding and decoding.

4. EXPERIMENTAL RESULTS

Some experiments have proved that application of the spiral scan does not influence compression efficiency of single-layer (nonscalable) AVC/H.264 nonscalable video codecs [13]. Here, we report the experiments on Scalable Video Model (SVM) [2] augmented with spiral scan tools.

The coding scenario was the following:

1. Base layer: QCIF, 15 Hz, 64 kbps, fully AVC-compilant (raster scan);
2. 1st enhancement layer: QCIF, 15 Hz, 128 kbps, spiral scan;
3. 2nd enhancement layer: CIF, 30Hz, spiral scan;
4. 3rd enhancement layer: 4CIF, 30Hz, spiral scan.

Four-layer coding was applied for 4CIF test sequences, while the last layer was absent in the experiments with CIF sequences. For all layer, the results for the spiral scan have been compared to those with the raster scan (Tables 1, 2).

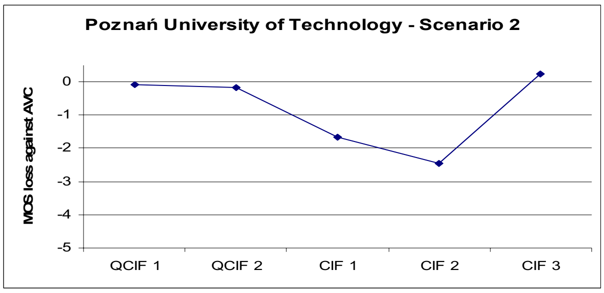
The results prove that the objective quality measured by PSNR values is the same for both codecs, while often the subjective quality was nicer for the raster scan. For the spiral scan, good subjective quality was verified by the independent tests in standard testing points used in MPEG [14]. Fig. 8 shows loss of subjective quality in a multi-layer scalable codec (simplified SVM) with spiral scan against the reference nonscalable AVC/H.264 codec.

Table 1. Experiments with 4CIF video sequences.

Sequence	Raster scan codec		Spiral scan codec	
	Bitrate [kbit/sec]	Luma PSNR [dB]	Bitrate [kbit/sec]	Luma PSNR [dB]
City 4 Layers	63.53	33.537		
	126.97	36.329	128.32	36.330
	511.37	35.254	511.60	35.218
	2045.05	33.882	2029.97	33.843
Crew 4 Layers	95.20	33.160		
	192.09	36.091	193.41	36.079
	763.08	36.553	766.22	36.542
	3059.08	36.966	3065.78	36.964
Harbour 4 Layers	96.67	30.466		
	192.97	33.202	193.22	33.162
	763.96	31.754	762.22	31.725
	3082.53	33.332	3083.54	33.325
Soccer 4 Layers	95.53	33.887		
	192.22	37.151	192.34	37.119
	762.28	36.189	762.12	36.158
	3065.39	36.115	3065.78	36.097

Table 2. Experiments with CIF video test sequences.

Sequence	Raster scan codec		Spiral scan codec	
	Bitrate [kbit/sec]	Luma PSNR [dB]	Bitrate [kbit/sec]	Luma PSNR [dB]
Bus 3 Layers	95.18	29.242		
	190.48	32.259	191.04	32.246
	509.67	29.108	510.76	29.074
Football 3 Layers	192.27	32.157		
	383.52	36.144	383.61	36.099
	1024.43	33.463	1023.72	33.444
Foreman 3 Layers	48.02	31.760		
	95.87	34.776	95.43	34.633
	255.18	32.973	255.57	32.911
Mobile 3 Layers	64.50	25.204		
	129.05	28.161	129.05	28.066
	387.06	26.201	384.48	26.111



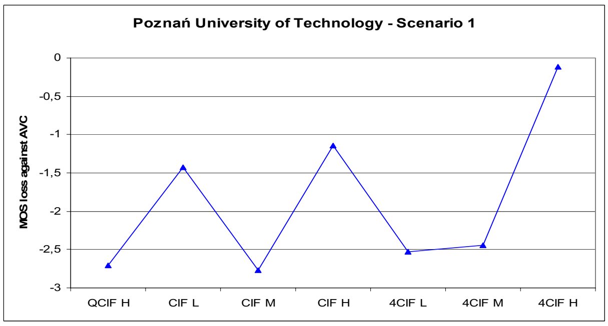


Fig. 8. Subjective quality loss for the sequences decoded from various layers of spatial and temporal resolution with the spiral scan with respect to that of the standard AVC/H.264 codec – results reported in [14] for the codecs and test material delivered by the authors.

5. CONCLUSIONS

Using the spiral scan, it is possible to obtain SNR scalability at no complexity increase and no coding efficiency loss against the respective nonscalable codec. The implementation of the scalable codec is extremely simple but in many sequences a viewer does not perceive any loss of quality due to extraction of up to 40% of bits, in some cases. The experiments have been done for the very demanding new codec that corresponds to Scalable Video Model built on top of the new AVC/H.264 codec. These experiments prove that application of the spiral scan does not decrease objective quality for spatial and temporal scalability.

REFERENCES

[1] "Working Draft 1.0 of 14496-10:200x/AMD1 Scalable Video Coding," ISO/IEC JTC1/SC29/WG11 Doc. N6901, Hong Kong, Jan. 2005.

[2] "Scalable video model version 3.0," ISO/IEC JTC1/SC29/WG11 Doc. N6717, Palma de Mallorca, Oct. 2004.

[3] "ISO/IEC 14496-10 Advanced Video Coding, 3rd Edition," ISO/IEC JTC1/SC29/WG11 Doc. N6540, Redmond, July 2004.

[4] Ł. Błaszak, M. Domański, R. Lange, A. Łuczak, "Response to SVC CE2 tasks: testing of SNR scalability technology," ISO/IEC JTC1/SC29/WG11 Doc. M11101, Redmond, July 2004.

[5] "Description of Core Experiments for Scalable Video Coding (SVC)," ISO/IEC JTC1/SC29/WG11 Doc. N6898, Hong Kong, Jan. 2005.

[6] H. Schwarz, D. Marpe, Th. Wiegand, "SNR-scalable extension of H.264/AVC", Proc. Int. Conf. Image Proc., Singapore, pp. 3113-3116, Oct. 2004.

[7] M. Domański, A. Łuczak, S. Maćkowiak, "On improving MPEG spatial scalability," IEEE International Conference on Image Processing, Vancouver BC, vol. II, s. 848-851, 2000.

[8] M. Domański, Ł. Błaszak, S. Maćkowiak, "AVC video coders with spatial and temporal scalability", Picture Coding Symp., Saint Malo, pp. 41-46, 2003.

[9] Y. Bao, X. Wang, M. Karczewicz, J. Ridge, "A low-complexity AVC-based scalable video codec," ISO/IEC J TC1/SC29/WG11 Doc. M11751, Hong Kong, Jan. 2005.

[10] Ł. Błaszak, M. Domański, R. Lange, "Modified AVC codec with SNR scalability based on macroblock hierarchy," International Workshop on Systems, Signals and Image Processing, IWSSIP'04, Poznań, pp. 15-18, Sept. 2004.

[11] G.H. Park, Y.J. Lee, W.-S. Cheong, K. Kim, J. Kim, Y. K. Lim, "Water ring scan method for H.26L based FGS," Tech. Rep. B094, ITU-T VCEG | ISO/IEC MPEG (JVT), Feb. 2002.

[12] G.H. Park, K. Kim, "Water ring scan method for FGS video coding schemes," *IEICE Trans. on Communications*, vol. E88-B, pp. 835-840, Feb. 2005.

[13] Ł. Błaszak, M. Domański, "Spiral scan in video compression," European Conference on Signal Processing EUSIPCO 2005, Antalya, Sept. 2005, to be published.

[14] V. Baronchini, T. Oelbaum, "Subjective test results for the CfP on scalable video coding technology," ISO/IEC JTC1/SC29/WG11 Doc. M10737, Munich, March 2004.