Diamond scanning order of image blocks for massively parallel HEVC compression

Adam Łuczak, Damian Karwowski, Sławomir Maćkowiak, and Tomasz Grajek

Chair of Multimedia Telecommunication and Microelectronics,
Poznan University of Technology, Polanka St. 3, 61-131 Poznan, Poland,
(aluczak,dkarwow,smack,tgrajek)@multimedia.edu.pl,

Abstract. In all hybrid video coders the raster scanning order of macroblock has been used. Such order scheme make parallel processing of macroblock very difficult. or even impossible The paper presents alternative method of macroblock ordering called "diamond scan". This scheme of macroblock ordering allows for strong parallelization of macroblock processing and to have benefit from multicore processors (GP and GPU). Applying presented scheme leads to create a more efficient software and hardware implementations, in terms of energy consumption and efficient use of available equipment.

1 Introduction

Dynamic development of multimedia services led to elaborating numerous video compression techniques [1]. Greatest popularity was gained by technology of hybrid compression of a video that found practical applications in many international and commercial video compression standards [1,2].

The main idea of hybrid video coding is dividing images into non-overlapping rectangular blocks, and coding data of individual blocks by the use of intra prediction, motion-compensated prediction, transform coding and entropy coding of residual data. This approach was extensively improving in video encoders of successive generations. Compression efficiency of encoders was gradually increasing by the application of more and more sophisticated context-based coding mechanisms, in which information from neighboring blocks is used when encoding data of the current block.

The context-based coding paradigm improves compression performance of encoder, but also increases its complexity significantly. Additionally, it makes performing computations in parallel in a video codec extremely difficult. From that reason, the commonly used solution in contemporary video encoders is processing image blocks sequentially one by one in the so-called raster scanning order, i. e. from left to right and from top to bottom of an image. This imposes severe limitations to structure of video encoder that can not be efficiently realized on available multi-processor and hardware-based platforms (ASIC, FPGA) with parallel processing of data.

The above mentioned features of context-based coding together with fact of continuous increasing the spatial and temporal resolution of a video material

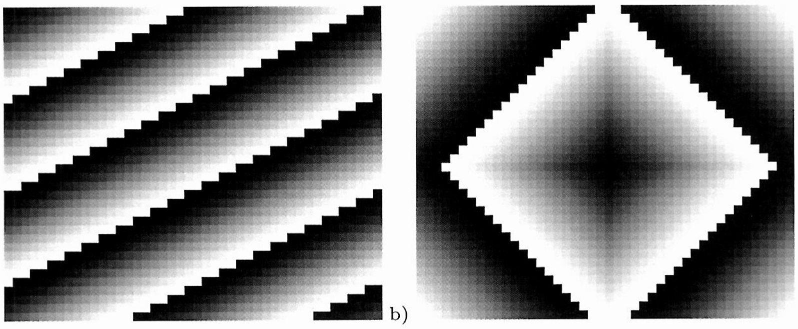

Fig. 1. CU's scanning order for a) WaveFront and b) Diamond.

make real-time video coding extremely difficult. From that reason, alternative order of coding image blocks is additionally investigated in new forthcoming technology of High Efficiency Video Coding (HEVC) (known as wavefront processing) with possibility of parallel processing of image data [2,4,6]. Nevertheless, abilities of performing calculations in parallel are still limited for wavefront scheme within HEVC.

The question arises, whether it is possible to modify existing technologies of hybrid video compression to perform computations in a massively parallel way. The authors propose new mechanism of image blocks ordering in encoder that increases degree of computations parallelism with still high compression performance of a video encoder. The proposed mechanism was explored in context of coding tools of forthcoming HEVC video compression technology. Description of the proposed method, detailed methodology of experiments and results were presented in the paper.

2 Proposed 'diamond' mechanism of image blocks ordering

Currently available context-based solutions for hybrid video compression technology put the requirement to process image blocks sequentially (for raster scanning order) [1, 7] or enable relatively moderate level of computations parallelism in encoder (in the case of wavefront processing) [2, 3]. Authors proposed new mechanism of image blocks ordering – called the diamond scanning scheme, in order to enable massive parallelism of computations in a video codec. What is very important, the idea of context-based coding that is essential from the compression performance viewpoint is still in use for the proposed image block scanning scheme.

The main idea of the diamond scanning scheme is to group image blocks into separate coding zones (marked with successive numbers), with the first coding zone located in the center of an image as presented in figure 2.

The size of an image block can be arbitrary determine depending on the compression technology used and desired level of computations parallelism. In

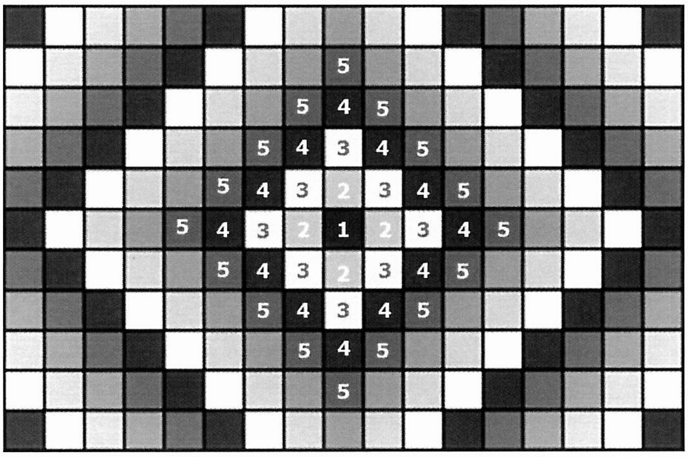

Fig. 2. Diamond scanning order.

this proposal, the image block is a coding unit (CU) of HEVC video compression technology [6], with fixed size of 16x16 luma samples. The coding zone is of diamond shape and collects image blocks that can be processed simultaneously on multi-processors. In turn, in order to take advantage of context-based coding, individual coding zones are processed sequentially, one by one, beginning from the center of an image and going towards the image margins. In this way, when processing a coding zone, data of previously processed coding zones can be used as a context information for more efficient video compression.

In general, the level of computations parallelism increases for successive coding zones for the proposed mechanism. Number of blocks that can be processed simultaneously is not fixed, but is higher than in the case of wavefront processing. Detailed numbers for individual zones were presented in figure 3. What is extremely important, significant percentage of blocks have access to 3 context blocks, which ensures high compression performance of a video encoder. This and other aspects of the proposed mechanism will be investigated in depth in later sections of the paper.

3 Diamond scanning - experimental results

In order to explore the impact of the diamond blocks ordering on various aspects of video compression, the own software of video encoder was elaborated. The encoder was implemented from scratch in C programming language with the use of coding tools taken from HEVC video compression technology (intra-frame prediction, transform and quantization) together with the proposed mechanism of diamond blocks ordering. The proposed method of image blocks ordering was investigated in terms of three aspects that are extremely important in video compression. First of all, what is the number of neighboring blocks that is available for context-based coding mechanism? Secondly, what level of computations parallelism is possible for the proposed method? And finally, to what extent the proposed method affects compression performance of a video encoder?

Methodology The level of computations parallelism and the number of neighboring blocks, which are available for context-based coding can be easily calculated for a given method of blocks scanning. Nevertheless, The impact of image blocks ordering on compression performance of encoder can be tested only experimentally. In order to conduct such experiments, the mentioned software of video encoder was additionally equipped with two alternative methods of image blocks ordering, i. e. raster scanning scheme and wavefront scanning scheme (besides the default diamond scanning scheme).

The preliminary analysis revealed that in the case of inter-coded frames the influece of the method of image blocks ordering on compression performance is negligible. Motion estimation and compensation is realized in time and not in space, thus the location of a block within an image is not important in the process of motion vector calculation.

Different situation takes place for intra-coded frames. In this case, an image block is coded with reference to data of neighboring image blocks that were already coded. It means that the way of blocks ordering can play significant role in encoder. This aspect was explored by testing the efficiency of HEVC intra prediction mechanism (plane predictor, DC predictor and set of 17 angular predictors) for those three methods of image blocks ordering: raster, wavefront and diamond. For every methods, the 'quality' of prediction was tested by calculating the known in compression PSNR measure between the original and predicted (or reconstructed) frames. Experiments were done for two scenarios: 1) without quantization and without residual signal and 2) with quantization and with adding the residual data. In the first scenario, the prediction is performed on the basis of original image samples, and PSNR is calculated between original and predicted frames. The second scenario refers to standard encoding procedure, i. e. the PSNR is calculated between the original and reconstructed frames.

Due to no signal quantization in the first scenario of experiments (refer to table 3), it allows to determine the boundary (the best) case, because it answers the question what is the best prediction of an image that can be achieved with a given prediction tools and a method of blocks ordering. The second scenario shows the performance of a video encoder working in a lossy compression mode.

In general, in the case of raster and wavefront scanning schemes, the context data (neighboring CUs) is located to the left and above the currently coded CU. For the diamond ordering method, the location of context data is not fixed. Depending on the location of the CU within an image the location of context data changes. From that reason, the prediction tools had to be modified in order to be able to perform prediction from arbitrary direction.

Analysis of computations parallelism: The known solutions described in a literature are oriented for the entropy coding parallelization. This is achieved by splitting an image into entropy slices. Each entropy slice is composed from one

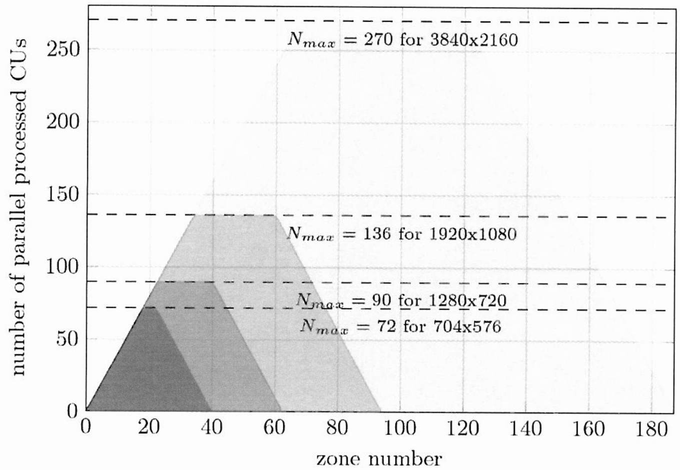

Fig. 3. Number of parallelly processed CU's for given zone.

to several CU rows. Mostly, it is possible to process only a few units of CU at a time.

A better solution seems to be the wavefront [5], which introduces the new way of the CU scanning order (figure 1). This scanning mode provides the same number of contexts, as traditional raster scanning.

Unfortunately, the disadvantage of this scanning method is slow increase of the number of CU units processed in parallel. For example, in the theoretical case of an image with the 2:1 aspect ratio, the maximum number of CU (N_{max}) which can be processed simultanously is equal to the vertical size (V_{CU}) of an image counted in CUs. During encoding, the number of parallelly processed CU is increasing from 1 to N_{max} and decreasing to 1. In the real case when the image has an aspect ratio 16:9, the N_{max} is never reached.

Authors' proposal was designed to achieve a massive parallelization of CU encoding. In addition, high level of parallelism is being quickly achieved during image encoding process. These capabilities are particularly important in the phase of encoding. In contrast to the decoder, the encoder must review all combinations of coding methods in order to find the best of them.

Proposed method of scanning order has two main features: 1) the starting point is placed in the center of an image, 2) the CUs have maximum of 3 possible contexts. In result, the number of CU units processed in parallel grows rapidly. Its number is increasing more then 4 times faster than for wavefront method, quickly achieving high value (see figure 3).

Analysis of context availability Changing the way of CU scanning order cause the change of availability and location of the contexts. While in the practice, the position of the context is only a technical problem, whereas a change in the number of available contexts may affect the efficiency of compression.

Histogram of contexts number for all three scanning modes (raster, wavefront, diamond) has been shown in table 1. As has been said, in the wavefront scanning mode availability of contexts is the same as for raster scanning mode. While in the proposed method of CU ordering, there are only 1 or 3 available contexts, which suggests that the quality of intra prediction can be poorer. However, this is the cost of a strong parallelization of calculations. The question is, if such cost is acceptable ?

Table 1. Histogram of available contexts for various scanning schemes.

Sequence resolution	Scanning mode	Number of CU's with N contexts					Total number of CU's
		N=0	N=1	N=2	N=3	N=4	
704x576	Raster	1	43	35	35	1470	1584
	WaveFront	1	43	35	35	1470	1584
	Diamond	1	78	0	1505	0	1584
1280x720	Raster	1	79	44	44	3432	3600
	WaveFront	1	79	44	44	3432	3600
	Diamond	1	123	0	3476	0	3600
1920x1080	Raster	1	119	67	67	7906	8160
	WaveFront	1	119	67	67	7906	8160
	Diamond	1	186	0	7973	0	8160

Analysis of coding efficiency In order to estimate the impact of the proposed block scanning order on the coding efficiency, the large set of test sequences has been encoded using various values of quantization parameter (Q) and for three scanning orders. The results of average PSNR (for 250 frames per test) are shown in tables 3 and 3.

How one can see the differences between PSNR for raster and diamond scanning order are very small (see table 3). What is noteworthy, although proposed scheme reduces the maximum number of available contexts from 4 to 3, often gives a better average PSNR than raster scanning mode.

This phenomenon has been investigated by the authors. It was shown that the starting point in the center of the image and growth of the coded area in the direction of image sides cause improvement the efficiency of intra-frame prediction. This gain compensates the loss associated with reduced amounts of CU contexts. The matrix sequence is a good example of this phenomenon (see table 3), which contains a scene with tunnel. The scene is characterized by the prospect of converging in the center of the picture, which is also the starting point of encoding process. In this case, as it was turned out, intra-frame predictions worked the best. The obtained PSNR gain was about 0.8 dB.

The results obtained for the set of test sequences show that the proposed method of scanning does not affect both the quality of the intra-frame prediction

Table 2. Average PSNR difference between sequences processed with different kind of scanning methods.

Test sequence	Sequence resolution	$\Delta PSNR(PSNR_{DIAMON} - PSNR_{RASTER})[dB]$		
		No quant	Q=28	Q=48
Blue Sky	1080p	−0.031	+0.010	0.000
Pedestrian Area	1080p	−0.115	+0.021	+0.002
Sun Flower	1080p	+0.227	+0.223	+0.185
Spin calendar	720p	+0.002	−0.021	−0.025
Crew	720p	+0.003	+0.520	+0.039
Raven	720p	−0.049	+0.044	+0.024
Harbour	720p	+0.009	−0.007	+0.007
Matrix	576p	+0.831	+0.769	+0.724

Table 3. Average PSNR for *Crew* sequence processed with different kind of scanning order.

	Results for Crew.yuv 1280x720				
	$Q = 8$	$Q = 16$	$Q = 24$	$Q = 32$	$Q = 40$
$PSNR_{RASTER}$	51.220	42.861	36.854	32.515	30.440
$PSNR_{WAVEFRONT}$	51.220	42.861	36.854	32.515	30.440
$PSNR_{DIAMOND}$	51.215	42.841	36.857	32.569	30.338

and the compression efficiency. In some cases it may even increase prediction quality and give a substantial profit in PSNR.

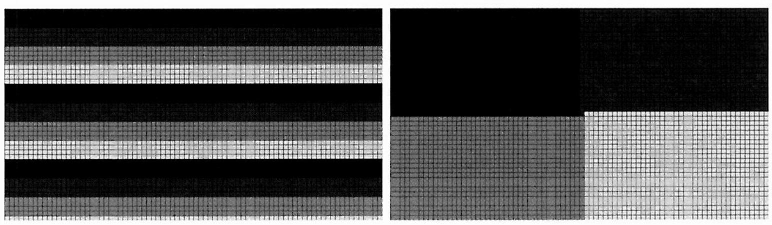

Fig. 4. Exemplary arrangements of entropy slices.

4 Scanning method and the entropy coding

It should be noted that the proposed method of parallelism introduction refers only to the coding/processing units but does not include entropy coding. This topic is out of the scope of this paper, but authors have also a concept of parallel entropy coding and i few words will be drafted. In general entropy encoding requires a different approach. In the case of a raster scan mode of CU, we have only one bitstream, the wavefront scanning mode requires as many bitstreams as there are CU rows in a video image. This two solutions have some drawbacks. For raster scanning the entropy coding is a bottleneck of encoder and decoder but for wavefront with a lot bitstreams (e.g. 17 for FullHD) the problem of inter-slice synchronization arises. For diamond scanning mode authors proposes a star scanning mode with 4 entropy slices, one for each quarter of an image, which is a compromise between one and large number of bitstreams.

5 Conclusions

The new method of image blocks ordering has been proposed. The method introduces strong parallelism of computations of encoder and decoder. The analysis prooved that application of the new scanning order does not impair the efficiency of video encoding. The proposed metod was explored in the context of HEVC video compression technology. Nevertheless, the method can also be used for any hybrid video coding technique (i.e. AVC, VC-1, AVS)[2,8,9].

Thanks this, in real-time software and hardware implementations, more complex and accurate compression algorithms can be used. As result it is possible to achieve better coding efficiency. In the future proposed scheme of CU scanning should let to compose a natively multi-threaded coding algorithms or even standards.

References

1. J. W. Woods, Multidimensional Signal, Image, and Video Processing and Coding, Academic Press, 2012
2. ISO/IEC 14496-10:2010, Coding of Audio-Visual Objects – Part 10: Advanced Video Coding, December 2010.
3. ITU-T Q6/16 and ISO/IEC JTC1/SC29/WG11: Joint call for proposals on video compression technology. MPEG Document N11113, Kyoto, Japan, January 2010.
4. Mesa, M.A.; Ramirez, A.; Azevedo, A.; Meenderinck, C.; Juurlink, B.; Valero, M.; , "Scalability of Macroblock-level Parallelism for H.264 Decoding," Parallel and Distributed Systems (ICPADS), 2009 15th International Conference on , vol., no., pp.236-243, 8-11 Dec. 2009
5. G. Clare, F. Henry, S. Pateux,"Wavefront Parallel Processing for HEVC Encoding and Decoding", JCTVC-F274 / m20694, 6th Meeting: Torino, IT, 14-22 July, 2011.
6. Blaszak, L.; Domanski, M.; , "Spiral coding order of macroblocks with applications to SNR-scalable video compression," Image Processing, 2005. ICIP 2005. IEEE International Conference on , vol.3, no., pp. III- 688-91, 11-14 Sept. 2005
7. K. McCann, B. Bross, I. K. Kim, S. Sekiguchi, W. J. Han: HM5: High Efficiency Video Coding (HEVC) Test Model 5 Encoder Description. JCTVC- G1102, JCT-VC Meeting, Geneva, November 2011.
8. Society of Motion Picture and Television Engineers: VC-1 Compressed Video Bitstream Format and Decoding Process. SMPTE 421M-2006, 2006.
9. Audio Video Coding Standard Workgroup of China (AVS): The Standards of People's Republic of China GB/T 20090.2-2006, Information Technology – Advanced Coding of Audio and Video – Part 2:Video, 2006.