A CONSTANT RATE BANDWIDTH REDUCTION ARCHITECTURE WITH ADAPTIVE COMPRESSION MODE DECISION FOR VIDEO DECODING

Liu Song[1], Dajiang Zhou[1], Xin Jin[2] and Satoshi Goto[1]

[1]Graduate School of Information, Production and System, Waseda University
[2] Information Technology Research Organization, Waseda University
2-7 Hibikino, Kitakyushu, Fukuoka, 808-0135, Japan
phone: + (81) 80-3974-3207, email: liusong@toki.waseda.jp

ABSTRACT

The high definition video capture and playback is becoming increasingly popular in battery-operated handheld devices. Video coding standards like H.264/AVC achieves high coding efficiency but increases the system power consumption which is limited for battery-operated devices. This paper proposes a new adaptive external memory bandwidth reduction scheme to reduce power consumption. An efficient memory organization and a variable length coding scheme is embedded in this scheme. Moreover, an adaptive compression strategy is proposed to adaptively choose the best mode for reference frame compression. The adaptive mode decision unit optimizes the compression mode according to the bandwidth availabilities and image pattern complexity. Experimental results show that the proposed scheme achieves up to 70% bandwidth savings with 1~2dB image quality losses.

1.　　INTRODUCTION

Video codec systems, including encoders and decoders for MPEG-2, H.264/AVC and etc., usually rely on large external DRAMs for buffering mass data such as the reference frame. The estimated SDRAM memory access bandwidth for motion estimation and motion compensation in a 720p30 H.264 encoder is around 380 MB/s assuming number of reference frames = 2, motion vector search range of (64 Horizontal, 32 Vertical)[2]. In hardware implementation, the process causes significant power consumption. As more and more portable consumer electronics video devices, e.g. camera phones and personal media players, are becoming popular, memory bandwidth will become a critical factor of the whole system cost, especially for those battery-operated consumer electronic video devices with high-definition (HD) display capability.

In HD video decoders for the standards like H.264/AVC, the most bandwidth consuming part comes from frame access which is needed in the process of motion compensation. In previous works, many techniques have been developed to reduce the frame access bandwidth consumption. In [2-3], reference frames are compressed before they are stored into DRAM, and decompressed when they are fetched for reading. In [5-6], memory organizations for storing reference frames are optimized to increase the hit rate of the memory on chip.

However, in previous works, generally, only one bandwidth reduction ratio is provided which cannot adapt to different system states and applications. Moreover, a constant compression ratio cannot always achieve the best reference frame compression efficiency for different image contents and at the meantime preserving the decompressed image quality under a specific bandwidth constraint. On the other hand, variable compression ratios cause the problem that the bit rate of each frame is inconstant. As a result, the convenience of allocating constant density of the storage device or/and allocating constant bandwidth of accessing the storage device during a pre-scheduled time slot is lost.

In this paper, an adaptive bandwidth reduction scheme is proposed. The proposed scheme provides arbitrary bandwidth reduction ratios according to the user requirement and is applicable to different applications. Moreover, the proposed scheme achieves the target by applying different compression methods to different image pattern complexities while maintaining the constant frame bit rate.

An efficient memory organization and a variable length coding scheme are embedded in the proposed scheme. The memory mapping with a predefined data fetching scheme keep the row, bank and group address unchanged so that the random access to each partition on the group level can be ensured. There are no fluctuations on memory storage size as we save the DRAM traffic while keeping the original storage mapping. The variable length coding scheme achieves good coding efficiency by processing each unit differently.

Moreover, several compression modes are provided to adapt to different bandwidth availabilities and image complexity patterns. To optimize the mode decision, an adaptive mode decision scheme with real-time bandwidth consumption feedback is introduced. The adaptive mode decision controller is to decide, on a processing unit, whether to choose the lossless compression to the unit, the lossy compression to the unit or very lossy compression to the unit with more scarification of image quality.

The adaptive mode decision controller serves for the following purposes. Firstly, the variable length causes the on-chip/off-chip data transfer bit rate inconstant. The controller adapts mode decision to maintain the constant transmission bit rate. Secondly, to optimize the mode decision on the bandwidth availability and image complexity pattern, parameters such as different predetermined thresholds and ini-

tial mode are adjusted. Thirdly, the strategy to improve subjective image quality is also integrated in the controller. Fourthly, the adaption between the frames is also utilized. The mode used in compressing the current reference frame is predicted from the modes utilized in the previous frame. Experimental results reveal that the proposed scheme is efficient for different bandwidth compression ratios and can be operated across a broad range of systems. The proposed scheme achieves up to 70% bandwidth savings with 1~2dB image quality loss.

The rest of the paper is organized as follows. The proposed scheme is explained in detail in Section 2. The experimental results and comparison with other works are given in Section 3. The conclusions are described in Section 4.

2.   MEMORY ORGANIZATION

Before compressing, a reference frame is first divided into an integer number of partitions with each partition of 16x2 pixels. One partition refers to a group of reference pixels, which is processed as a basic unit for compression and decompression. In order to apply the variable length compression to each partition, an appropriate memory mapping is required with which random access can be ensured through low complexity control scheme. A memory organization suitable for variable length compression has been proposed in [4]. To apply the optimum partition size to the memory organization, the proposed modified memory organization is shown in Fig.1.

As shown in the figure, in a reference frame, 128×64 pixels are stored in the same row address of DRAM. Moreover, one bank is divided into 8 groups, each of which contains 8 vertically adjacent partitions as shown in Fig.1. By means of this layout, a neighbouring bank can be pre-activated for a row switching so that switching penalties can be reduced. As the partition length of each partition after compression is not constant, fetching the partition length information is necessary. The extra bandwidth consumption spent on fetching partition length information is negligible and the partition length fetching scheme is simple [4]. Since the purpose of this work is to save the DRAM traffic instead of storage size, it is possible to keep the compressed mapping of row, bank and group addresses unchanged from the original one, so as to randomly locate each compressed partition to the group level. Then inside one group, partitions are compressed and organized compactly as depicted in Fig.2. The compressed partitions occupy less space than original ones, so that there is some unused space in most groups. Though the unused space does not add to DRAM data transfer, it decreases the opportunity for consecutive vertical access and consequently leads to extra scheduling overhead of the DRAM controller. To overcome this problem, the compressed partition in the upper half (group 0~3) of a bank are organized to stick to the lower boundary of their groups, while the lower half (group 4~7) stick to the upper boundary, so that a vertical burst access can be sequential even if it crosses the group boundary. (Fig.2)
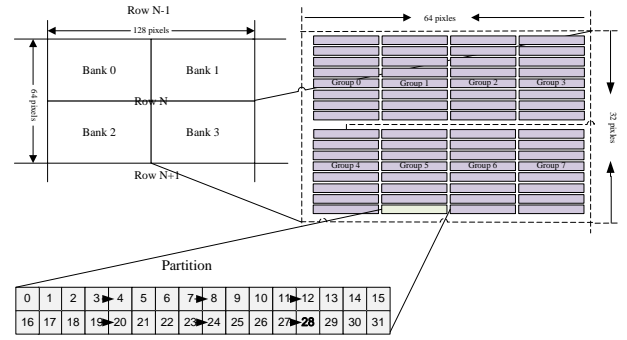

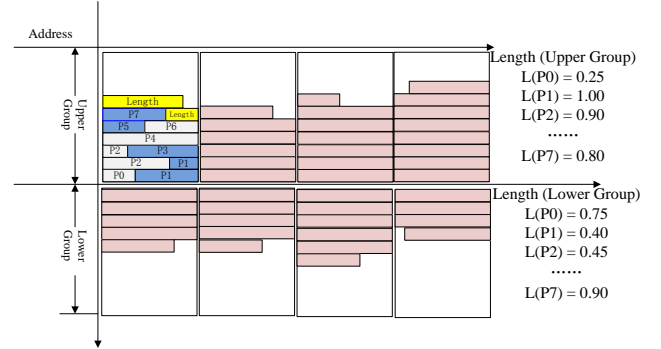Fig.1.   Memory organizations for reference frames

Fig.2.   Compressed mapping of partitions

3.   VARIABLE LENGTH CODING

Each partition is processed as a basic unit for compression and decompression. DPCM (Differential Pulse Code Modulation) is first performed to code value difference between adjacent pixels. Since DPCM within a partition have a strong locality, the conventional codes, especially the universal codes, are not so efficient for them. Therefore, we apply an efficient variable length coding scheme [1]. The maximum absolute DPCM value (MADV) of the sub-partition (Here we refer to 2x2 pixels) is utilized to decide the coding patterns A, B, C, D, E and F. The same DPCM value is coded differently according to the determined pattern. By applying variable bit rates for different image contents, image quality is ensured. Moreover, to increase the compression efficiency for different application cases, lossless and lossy codes are provided in the variable length coding table as shown in Table I. As listed in the Table I, S indicates the sign bit, lossy coding is highlighted in yellow and denoted by LY and the lossless coding is denoted by LS.

4.   ADAPTIVE MODE DECISION CONTROLLER WITH MODE OPTIMIZATION

Variable length coding achieves good coding efficiency by utilizing the data characteristic of different image complexity patterns. Therefore, the frames with fewer details can be compressed using fewer bits, while, others with higher complexity are compressed using more bits. As a result, an adap-

tive mode decision controller is proposed to control the compression ratio of each reference frame to a constant value according to the provided bandwidth limit.

The adaptive mode decision controller is to determine whether and how to compress each partition. Specifically, several compression modes are provided which are dedicated for different compression ratios. The adaptive mode decision controller determines the optimum mode for each partition according to the bandwidth availability. The controller adjusts the compression mode according to the accumulated bits of compressed partitions and the image complexity pattern. The control flow of adaptive mode decision controller in video decoding is shown in Fig.3.

Table I   Variable length coding table

MADV		0	1	2	3-4	5-8	9-20
Pattern		A 00	B 01	C 10	D 110	E 110	F 11110
0	LS		1	01	001	0001	00001
	LY		NO	1	01	001	0001
±1	LS		0S	1S	01S	001S	0001S
	LY		NO	1	01	001	0001
±2	LS			00S	10S	010S	0010S
	LY			0S	1S	01S	001S
±3	LS				11S	011S	0011S
	LY				1S	01S	001S
±4	LS				000S	100S	0100S
	LY				00S	10S	010S
±5	LS					101S	0101S
	LY					10S	010S
±6	LS					110S	0110S
	LY					11S	011S
±7	LS					111S	0111S
	LY					11S	011S
±8	LS					0000S	1000S
	LY					000S	100S
±9	LS						1001S
	LY						100S
±10	LS						1010S
	LY						101S
±11	LS						1011S
	LY						101S
±12	LS						11000S
	LY						110S
±13	LS						11001S
	LY						110S
±14	LS						11010S
	LY						110S
±15	LS						11011S
	LY						111S
±16	LS						11100S
	LY						111S
±17	LS						11101S
	LY						111S
±18	LS						11110S
	LY						0000S
±19	LS						11111S
	LY						0000S
±20	LS						00000S
	LY						0000S

4.1   Adaptive Compression of Reference Frame

There are eight compression modes designed to achieve different compression ratios shown in Table II. We call mode 7 is more aggressive than mode 0 as it reduces more bandwidth though with larger image quality loss. The mode check starts from the initial mode. And the mode goes from the less aggressive modes to the more aggressive modes.

To show the compression efficiency of different compression modes, we implement the algorithm in JM 15.1 reference software and test different sequences using each mode individually. Using the encoding parameters listed in Table III, six sequences are coded and the averaged bandwidth compression efficiency (consumed in data reading and writing, respectively), which is defined as the ratio of the compressed bandwidth to the original uncompressed bandwidth. Compression mode efficiency over 4 QP values is shown in the Table IV. Table IV shows the proposed compression scheme and mode can reduce the bandwidth to an extent from 20% to 50%.

4.2   Mode Decision Optimization between the Frames

The mode decision optimization is first done on the scale of frames. The initial mode is set according to the required bandwidth ratio. A look up table, as listed in Table V is set to provide the initial mode decision. The table is based on the statistical compression ratio of different sequences shown in Table IV. As listed in Table IV, the predetermined bandwidth limit assumption is set on the consideration of Write (As shown in Table IV, Read compression ratio is usually 5% more than Write compression ratio and bandwidth limit can be set accordingly.)

Table II Compression modes

MODE	Compression Method
mode 0	Only Lossless coding is used
mode 1	ABCD use Lossless coding, EF use Lossy coding
mode 2	ABC use Lossless coding, DEF use Lossy coding
mode 3	AB use Lossless, CDEF use Lossy coding
mode 4	Quantized by 2 , then ABCDEF use Lossy coding
mode 5	Quantized by 4 , then ABCDEF use Lossy coding
mode 6	Quantized by 8 , then ABCDEF use Lossy coding
mode 7	Quantized by 16 , then ABCDEF use Lossy coding

Table III Test sequences and encoding parameters for mode compression experiments

Test Sequences	720p (1280×720)	Cyclists
		ShuttleStart
		Mobcal_ter
	1080p (1920×1080)	PedestrianArea
		RushHour
		BlueSky
Profiles	High Profile	
Coding Structure	IPBBP	
QP (I/P/B)	22/23/24,27/28/29,32/33/34,37/38/39	
Number of Reference Frames	4	
FrameRate	30	
NumberBFrames	2	
FramesToBeEncoded	30/720p, 30/1080p	
Search Range	32	

Table IV Compression efficiency for different mode

| Average CR	4 QP values (I)	Mode 0 (%)		Mode 1 (%)		Mode 2 (%)		Mode 3 (%)		Mode 4 (%)		Mode 5 (%)		Mode 6 (%)		Mode 7 (%)	
		Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read\	Write	Read	Write	Read
Cyclists	I:22/27/32/37	38.46	44.19	36.85	41.81	36.29	41.13	34.81	39.68	31.22	36.84	26.71	30.30	23.24	25.34	17.44	19.11
ShutleStart	I:22/27/32/37	28.47	33.73	27.90	32.60	27.64	32.15	26.96	31.19	25.36	29.42	23.44	25.67	22.45	23.56	16.14	17.00
Mobcal_ter	I:22/27/32/37	60.57	64.99	55.96	60.05	55.28	59.39	54.44	58.58	53.16	57.46	43.96	47.57	34.06	36.72	24.53	26.47
PedestrainArea	I:22/27/32/37	38.13	42.69	36.08	39.99	35.52	39.33	34.49	38.35	32.20	36.44	26.24	29.07	20.73	22.60	15.79	16.83
RushHours	I:22/27/32/37	36.17	41.41	34.42	38.89	33.90	38.21	32.95	37.22	30.77	35.35	25.72	28.67	20.54	22.48	16.12	17.20
BlueSky	I:22/27/32/37	36.34	46.98	34.29	43.39	33.98	42.85	33.36	42.01	32.21	40.61	28.19	33.77	24.53	27.40	19.94	21.25
Average		39.69	45.66	37.58	42.79	37.10	42.18	36.17	41.17	34.15	39.35	29.04	32.51	24.26	26.35	18.33	19.64
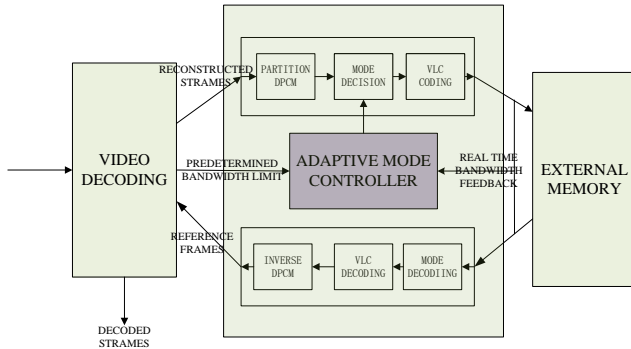
Fig.3. Control flow of adaptive mode decision controller in video decoding

Table V Look up table for initial mode decision

Predetermined bandwidth limit for Write (%)	(40, 100]	[35, 40)	[25, 35)	[0, 25)
Initial Mode	Mode 0	Mode 1	Mode 3	Mode 6

The mode decision controller starts to the compress reference frame using the initial mode. However, it is not always the mode which can achieve the desired performance. To optimize the mode between the frames, we have the engine to record the accumulated bandwidth consumption in previous frame. The mode decision unit compares the previous accumulated bandwidth consumption with the predetermined bandwidth limit and does the optimized mode decision at the beginning of current frame. Specifically, if accumulated bandwidth is more/less than the predetermined bandwidth limit, mode up/down decision for current frame is made on the basis of initial mode.

4.3   Mode Decision Optimization inside the Frames

The mode decision optimization between the frames does the mode check on the scale of frame. However, it cannot change the mode precisely and quickly in some cases. Therefore, the mode check inside the frame is also implemented. The mode check inside the frame is done on a smaller scale of unit, specifically, several blocks of partitions. According to current experimental results, 3~4 groups with a mode check show the optimum result. The accumulated bit rates of 3~4 groups are compared with the prede-

termined bandwidth. Specifically, mode up/down decision for following partitions is made.

Moreover, to further improve the image quality subjectively, we also apply the region of interest method. As we know, the image part with fewer details is not as important as the image part with more details. Therefore, the image part with many details should be ensured with high image quality consequently, we select higher compression modes (mode up decision) for the image part with fewer details, while use lower modes (mode down decision) for the areas with more details. The whole mode decision optimization flow is shown in Fig.4.

5.   EXPERIMENTAL RESULTS

To evaluate the performance of the proposed scheme, the proposed scheme is implemented in H.264 reference software JM15.1. The test sequences together with the encoding parameters are the same with that listed in Table III except that the bitstream is generated using only QP (I/P/B) equal to 27/28/29. The detailed encoding parameters are shown in Table VI. The previous works based on a constant bit rates scheme for each partition (PRE [1]), a min-max-scalar-quantization based approach proposed in [2] (MM) and an embedded reconstruction patterns based method proposed in [3] (ERP) are also implemented and compared.

Table VI Test sequences and encoding parameters for proposed architecture

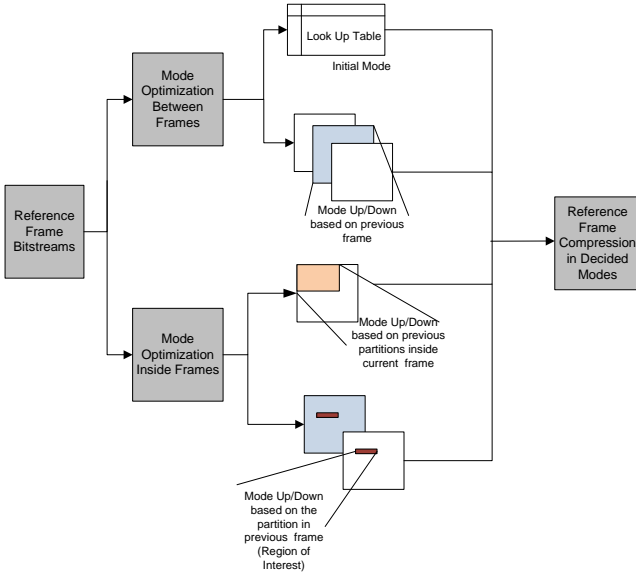| Test Sequences	720p (1280×720)	Cyclists
		ShuttleStart
	1080p (1920×1080)	PedestrianArea
		RushHour
Profiles	High Profile	
Coding Structure	IPBBP	
QP (I/P/B)	27/28/29	
Number of Reference Frames	4	
FrameRate	30	
NumberBFrames	2	
FramesToBeEncoded	100/720p, 60/1080p	
Search Range	32	

Fig.4.   Mode decision optimization flow

Table VIII shows algorithm efficiency of the proposed work. As listed in the Table VIII, the proposed work achieved up to 70% bandwidth reduction with little image quality loss. Moreover, the proposed scheme achieved desired bandwidth reduction ratios (predetermined bandwidth limit) accordingly so that the scheme can be applied in different system conditions.

The proposed work is also compared with previous relative works. The results of previous works are shown in Table VII. Both the proposed work and previous works can be implemented on encoder side and decoder side. The results are compared with algorithm implemented only on the decoder side. There are two reasons for this. First, the implemented system is syntax compatible with H.264 standard and can decode any H.264 compressed bit stream. Secondly, as the previous works [1], [2], [3] and the proposed work apply the lossy way to do reference frame compression, the error propagation happens when there is a mismatch between encoder and decoder. In previous works [2] and [3], the algorithm is implemented in the core video coding loop (that is, the algorithm is implemented both on the encoder side and decoder side). As a result, the error robustness cannot be seen in previous results. As listed in Table VII and Table VIII, even if the algorithm is not implemented the core coding loop and there is error propagation on the decoder side, the proposed work shows better compression efficiency and less image quality loss than previous works.

6.   CONCLUSIONS

In this paper, the constant bit rate bandwidth reduction architecture with adaptive compression mode decision is proposed. The proposed scheme can provide arbitrary bandwidth reduction ratios according to the user requirement and is applicable in different applications. The proposed scheme applies a variable compression way to different image pattern com-

plexity at the same time maintaining the constant frame bit rate for each frame. The experiment results show the proposed scheme achieves up to 70% bandwidth savings with little image quality loss and is error robust in the occasion that the algorithm is only implemented on the decoder side.

Table VII   PSNR Comparison among the previous works
(Implemented only on decoder side with mismatch)

(dB)	PRE[1] -25%	PRE[1] -50%	MM[2] -25%	MM[2] -37.5%	MM[2] -50%	ERP [3] -50%
Cyclists	-0.56	-8.04	-6.18	-10.27	-17.65	-11.47
ShutleStart	-0.27	-4.71	-5.33	-8.36	-15.55	-10.01
PedestrainArea	-0.29	-8.15	-3.08	-4.59	-9.71	-7.30
RushHours	-0.49	-9.76	-3.72	-5.78	-12.04	-7.81

Table VIII   Performance of proposed work
(Implemented only on decoder side with mismatch)

	Desired BR	50% (-50%)	45% (-55%)	40% (-60%)	35% (-65%)	30% (-70%)
Cyclists 720p	Real BR	50.00	45.00	40.00	36.40	30.19
	PSNR	0dB	0dB	0dB	-0.5dB	-1.54dB
ShuttleStart 720p	Real BR	50.00	45.00	40.00	35.00	30.04
	PSNR	0dB	0dB	0dB	0dB	0dB
PedestrianArea 1080p	Real BR	50.00	45.00	40.00	35.91	31.95
	PSNR	0dB	0dB	0dB	-0.37dB	-1.03dB
RushHour 1080p	Real BR	50.00	45.00	40.00	35.35	30.71
	PSNR	0dB	0dB	0dB	-0.15dB	-0.84dB

7.   ACKNOWLEDGEMENT

REFERENCES

[1] Liu Song, Dajiang Zhou, Xin Jin, Peilin Liu, Satoshi Goto, "An adaptive bandwidth reduction scheme for video coding", in Proc. IEEE International Symposium on Circuits and Systems, 2010

[2] M.Budagavi and M.Zhou, "Video coding using compressed reference frames," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 2008, pp.1165-1168.

[3] Y.V.Ivanov and D.Moloney, "Reference frame compression using embedded reconstruction patterns for H.264/AVC decoder," in Proc. 3rd Int Conf. Digital Telecommunications, 2008, pp.168-173.

[4] Xuena Bao, Dajiang Zhou, Satoshi Goto, " A lossless frame recompression scheme for reducing DRAM power in video encoding," IEEE International Symposium on Circuits and Systems, 2010

[5] T. Song and T. Shimamoto, "Reference frame data compression method for H.264/AVC", in IEICE Electronic Express ,vol.4, No. 3, 2007, pp. 121-126, 2007..

[6] P.Zhang, W.Gao, D. Wu, D. Xie, "An Efficient Reference Frame Storage Scheme for H.264 HDTV Decoder," in Proc.ICME,pp.361-364,July,2006.