

SCALABLE RATE-DISTORTION-COMPUTATION HARDWARE ACCELERATOR FOR MCTF AND ME

Yi-Hau Chen, Ching-Yeh Chen, Chih-Chi Cheng, Liang-Gee Chen

DSP/IC Design Lab.,
 Graduate Institute of Electronics Engineering,
 National Taiwan University, Taipei, Taiwan
 Email: {ttchen, cychen, ccc, lgchen}@video.ee.ntu.edu.tw

ABSTRACT

Motion-Compensated Temporal Filtering (MCTF) is an innovative prediction scheme for video coding and the core technology of scalable extension of H.264/AVC. The first MCTF and ME hardware work is in this paper. The proposed hardware not only can support the various coding schemes in JSVM and H.264 but also can adapt itself to provide rate-distortion-computation scalability. With the frame-level searching range data reuse and frame-interleaved MB pipelining scheme, external memory bandwidth are reduced 33%, and 10 Kbits buffer are saved. The proposed MCTF/ME hardware is of 350K gate counts and 30KB internal buffer, which can perform four-level MCTF or H.264 P-/B-frames at CIF Format.

1. INTRODUCTION

In recent years, the open-loop MCTF prediction scheme has been widely developed to achieve efficient scalable video coding. The main concept of MCTF is to perform discrete wavelet transformation in the temporal direction. For more details, please refer to [1]. MPEG has identified a set of applications that require scalable and reliable coding technologies. Currently, the scalable extension of H.264/AVC with MCTF is adopted as Joint Scalable Video Model (JSVM) 2.0 [2]. The lifting-based MCTF scheme is the core technology to provide scalable video coding. The MCTF can not only provide a variety of efficient scalabilities because its open-loop structure can prevent traditional drift problems but also improve the compression efficiency compared to H.264/AVC.

In a system-on-chip (SoC) design, system resources including system memory bandwidth and battery power are shared by integrated modules, like video coder, audio processor, and communication system. These integrated modules may be executed at the same time, and it makes the available system resources dynamic. If a rate-distortion-computation (R-D-C) video core which can provide the best rate-distortion performance under different available resources is integrated into SoC, the system can execute this video core regardless of the available system resource. For example, once the available system memory bandwidth or processing time is short, the R-D-C video core can adapt itself to meet the constraints of system resources. In this paper, we propose the first combined MCTF/ME core and utilize the relation between MCTF and ME to provide various R-D-C points.

In our previous work [3], the VLSI architecture of MCTF and the usage of system resources are analyzed. In this paper, we discuss the strategies for combining traditional closed-loop ME and open-loop MCTF and evaluate the impact on hardware cost. Since MCTF

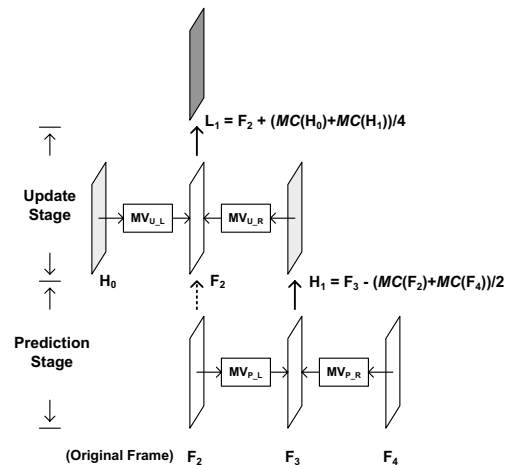


Fig. 1. The 5/3 MCTF scheme. $MV_{P,L}$ and $MV_{P,R}$ represent the motion vectors from the left and right neighbor frames for the prediction stage, respectively, and so present $MV_{U,L}$ and $MV_{U,R}$ for the update stage. The light gray frames (H) are the highpass frames, and the heavy gray frames (L) are the lowpass frames.

scheme induces high external memory bandwidth (EMB), the frame-level searching range (SR) data reuse scheme is adopted to reduce EMB. Besides, we propose a frame-interleaved macroblock(MB)-pipelining scheme to save hardware area, and the design strategies for update stage are also introduced. About 33% EMB and 10 Kbits buffer can be saved. The proposed scalable R-D-C MCTF/ME hardware is of 350K gate counts and 30KB internal buffer while supporting all JSVM and H.264/AVC coding schemes. Its processing ability is CIF Format with SR: $[-32, 32)$ for four level MCTF or H.264 P-/B-frames.

This paper is organized as follows. In Section 2, the MCTF schemes are introduced. Then, R-D-C properties of various coding schemes and its design challenges will be discussed in Section 3. In Section 4, the scalable R-D-C MCTF hardware architecture is proposed, and the experimental results are shown in Section 5. Section 6 will conclude this paper.

2. MOTION-COMPENSATED TEMPORAL FILTERING

MCTF is to perform wavelet transform in the temporal direction with motion compensation (MC). The coding performance and coding de-

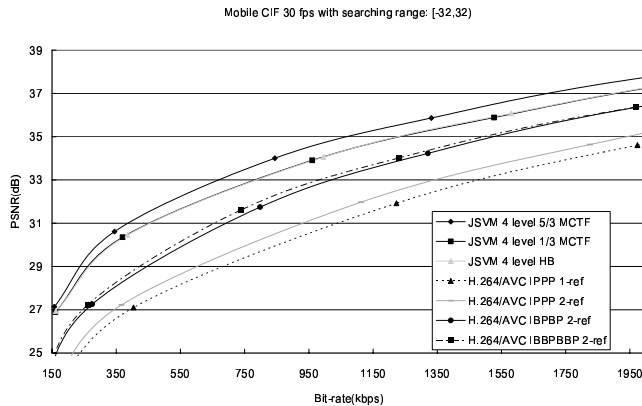


Fig. 2. PSNR comparison for JSVM 2.0 and H.264/AVC main profile for mobile sequence CIF format 30fps with searching range: [-32, 32). JSVM 2.0 represents the dedicated coding points.

Table 1. The Required Operations and Statistic of ME operation and external memory bandwidth (EMB) for JSVM and H.264/AVC.

Coding Scheme	Required Operations			ME	EMB
	ME	MC	Update	times/sec	MB/s
4 level 5/3 MCTF	Y	Y	Y	58.5	71.62
4 level 1/3 MCTF and HB	Y	Y	N	58.5	40.9
IPPP w 1 ref.	Y	Y	N	30	24.05
IPPP w 2 ref.	Y	Y	N	60	42.02
IBPBP w 2 ref.	Y	Y	N	60	42.02
IBBP w 2 ref.	Y	Y	N	60	42.02

lay depend on the adopted filter. MCTF is usually implemented by 5/3 or 1/3 filter with lifting scheme, because of its good performance and perfect reconstruction.

The 5/3 and 1/3 MCTF can be simply illustrated by Fig. 1, in which only two lifting stages are involved. The prediction stage is using even frames to predict odd frames by ME, and the residual frames are the highpass frames. The update stage is using the high-pass frames to update the even frames by MC, and the derived frames are the lowpass frames. The 1/3 MCTF can be performed by skipping the update stage of 5/3 MCTF and take even frames as lowpass frames directly. The multi-level MCTF scheme can be derived by recursively performing one-level MCTF on the L-frames in a bottom-up order. Furthermore, in JSVM, a coding scheme called Hierarchical B-frames (HB) is also introduced to provide H.264/AVC compatible scalable coding bitstreams. HB is to perform multi-level MCTF in a top-bottom way and its coding performance is quite similar to 1/3 MCTF.

3. SYSTEM ANALYSIS OF SCALABLE R-D-C MCTF/ME HARDWARE

3.1. Analysis of Rate-Distortion-Computation Properties

Figure 2 shows the coding performances of JSVM and H.264/AVC. In Fig. 2, four level 5/3 MCTF outperforms 1/3 MCTF and HB due to the update stage, and the performances of JSVM are better than those of H.264/AVC. These schemes have same basic operations,

ME and MC, and their computation complexities are quite similar as shown in Table 1 [3]. Their different coding performances result from different coding flows. Therefore, a core module of ME and MC can be configured to support either open-loop MCTF or closed-loop ME for different coding performance without degrading the hardware utilization.

To achieve R-D-C scalability, we analyze different coding schemes' R-D performance and EMB which is an important issue for system design. In general, they have different R-D performances and EMBs. From Fig. 2 and Table 1, four level 5/3 MCTF has the best coding performance with highest EMB while IPPP with one reference frame has the worst coding performance with the lowest EMB. It means that a combined MCTF/ME core can adapt current coding scheme to provide best video quality under available bit-rates and system resources. Thus, a scalable rate-distortion-computation MCTF/ME can always operate efficiently in an SoC.

3.2. System Design Challenges

Although a scalable R-D-C MCTF/ME hardware can be configured to meet available EMB, each scheme's EMB should be reduced as much as possible. In our previous works [3] [4], frame-level data reuse schemes are proposed to reduce EMB. To further reduce EMB, open-loop properties of MCTF should be considered. On the other hand, frame-level data reuse schemes will induce buffer overhead by conventional H.264/AVC design [5], such area overhead should also be saved as possible.

As mentioned in Section 2, three main operations, ME, MC, and update stage, are necessary for building a scalable R-D-C MCTF/ME hardware accelerator. Table 1 shows the required operations for various coding schemes. Since the update stage is only performed in 5/3 MCTF scheme, it is inefficient to build an independent module which supports all operations of update stage. To improve the utilization of whole system, the cost of update stage should be minimized.

4. PROPOSED SCALABLE R-D-C MCTF/ME HARDWARE ARCHITECTURE

In Section 3.2, the design challenges of scalable R-D-C MCTF/ME hardware are discussed. In this section, we apply frame-level searching range (SR) data reuse methods, which are extended from [3] and [4], to further reduce the EMB of MCTF. Then, we propose a frame-interleaved MB-pipelining scheme to reduce the hardware cost from frame-level SR data reuse. Finally, the design strategies of update stage are explored.

4.1. Frame-level Searching Range Data Reuse

In our proposed architecture, for processing P-/B-frames, two SR buffers are required to store two reference frames simultaneously. To efficiently utilize these SR buffers, Double Current Frames (DCF) scheme [4] is adopted in frame-level SR data reuse. DCF can efficiently reduce the EMB by sharing SR data belonged to two frames' current blocks. It makes the EMB of loading searching range data to be about half compared to conventional scheme. Moreover, modified DCF (m-DCF) [3] makes DCF to be suitable for fractional ME/MC as shown in Fig. 3(a).

For open-loop MCTF schemes, since ME takes inputted current frames instead of reconstructed frames as reference frames of closed-loop coding, we can cascade several m-DCF sets together and

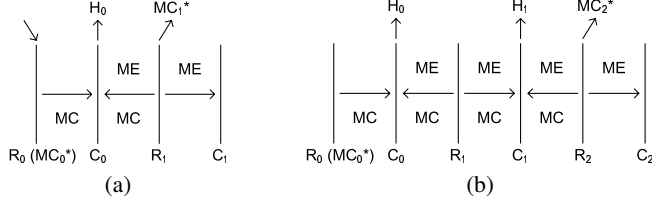


Fig. 3. The frame-level searching range data reuse scheme (C: Current frame; R: Reference frame): (a) modified double current frame; (b) Extended modified double current frame

process these current frames simultaneously due to no data dependency issues. In order to fully utilize the two SR buffers in our proposed architecture, we cascade two m-DCF sets as extended m-DCF as shown in Fig. 3(b). Then, the EMB of storing and loading the left intermediate MC blocks (MC_1^*), derived from R_1 to decide H_1 , can be further reduced. Besides, the extended m-DCF can be also applied on B-frames in H.264/AVC and reduce the EMB of closed-loop coding as well.

4.2. Proposed Frame-interleaved MB Pipelining Scheme

Due to the variable block size ME (VBSME) and Lagrangian mode decision in H.264/AVC, the operations of ME are divided into two pipeline stages, integer ME (IME) and fractional ME (FME) [5]. Since the prediction core of MCTF is quite similar to ME in H.264/AVC, the design of processing element in our previous work [6] is adopted with minor modifications for proposed MCTF/ME hardware. In Fig. 4, the notations of processed frames are the same as those of Fig. 3(b). Figure 4(a) shows the ME schedule of extended m-DCF scheme by directly applying MB pipelining scheme. For all current frames, the FME will start to process $n - th$ MBs only after all $n - th$ MBs finish IME processes. Therefore, the ME schedule in Fig. 4(a) needs to buffer six current MBs and five motion vector (MV) sets which induce a large internal buffer size as shown in Fig. 5.

To reduce the hardware cost caused from data dependency of MB pipelining scheme, we propose a frame-interleaved MB pipelining scheme to shorten the data life time of current blocks and MV data. The schedule of the proposed frame-interleaved MB pipelining scheme is shown in Fig. 4(b). Once the computation of one current block is finished, the IME and FME will enter next pipeline stage and the related results will be propagated to next processing module. Therefore, only two current MB buffers and two MV buffers are needed for IME and FME, respectively, and 10Kbits buffer, surrounded by dotted region in Fig. 5, can be saved.

4.3. Design Strategies for Update Stage

In JSVM, the update stage is separated into two operations, deriving update motion vectors and update MC, and the operating unit is 4×4 block. Based on the discussion in Section 3.2, the hardware cost of update stage should be minimized as possible. The EMB of update stage should be considered as well since update MC makes high EMB under direct implementation as shown in Table 2 due to fractional MC (FMC).

After examining the hardware resources in Fig. 5, we can separate the working periods of prediction stage and update stage so that the two SR SRAM and the FME module will be free during update stage. To minimize the hardware cost, we reuse the FME module's

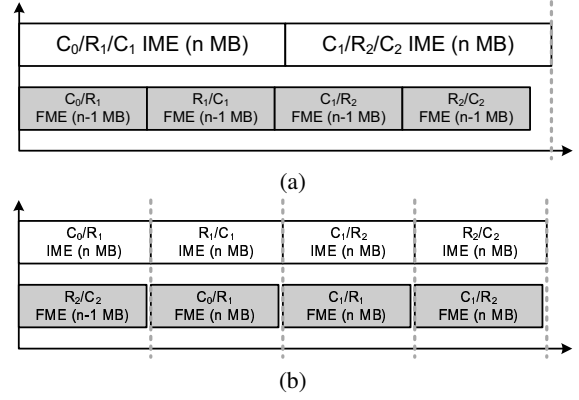


Fig. 4. The ME schedule compatible to extended m-DCF (C: Current frame; R: Reference frame): (a) ME schedule of MB pipelining scheme; (b) ME schedule of proposed frame-interleaved MB pipelining scheme

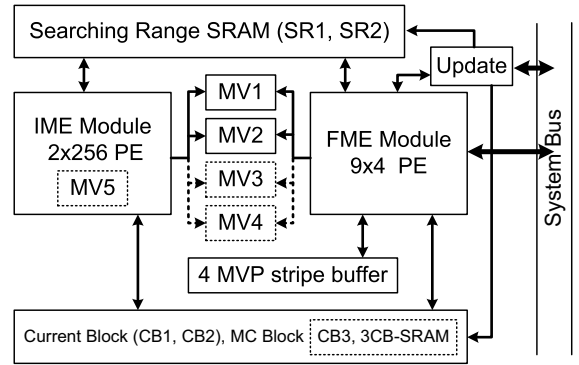


Fig. 5. The block diagram of scalable R-D-C MCTF/ME hardware based on proposed frame-interleaved MB pipelining scheme, the region surrounded by dotted lines are the extra buffer while applying MB pipelining scheme (SR : Searching Range; CB: Current Block; MC: Motion Compensated; MV: Motion Vector; MVP: Motion Vector Predictor)

FMC engine and MV buffer. Only weighting mechanism is added to original FME module for adaptive weighting update MC. Besides, the update module in Fig. 5 is designed to statistic the prediction MVs for deriving update MVs.

To reduce EMB, the concept of MB-level SR data reuse schemes are applied. For deriving update MVs, the intermediate results of update MVs are stored in one SR SRAM by level-D data reuse scheme [7] due to the large storage size of SR SRAM. In update MC, the ME-based MC with level C+ data reuse scheme [4] is proposed to treat highpass frames as reference data in ME for MB-level data reuse. Moreover, DCF is also applied to further reduce EMB of H-frame. A side effect of ME-based MC is regular external memory access, so the efficiency of external memory can be improved compared to irregular access.

5. EXPERIMENTAL RESULTS

The specification of the proposed scalable R-D-C MCTF/ME hardware accelerator is CIF format 30fps with SR: [-32, 32]. Table

Table 2. External Memory Bandwidth and Required Working Frequency Comparison of Proposed scalable R-D-C MCTF/ME hardware for MCTF and H.264/AVC for CIF 30fps sequence with searching range: [-32,32)

Coding Scheme		5/3 MCTF				IPPP		IBPBP	IBBP
		1 level	2 level	3 level	4 level	w 1 ref.	w 2 ref.	w 2 ref.	w 2 ref.
Prediction	Original	33.04	37.53	39.78	40.90	24.05	42.02	42.02	42.02
EMB(MB/s)	Proposed	33.04	35.32	34.21	32.54	24.05	42.02	24.05	30.04
Update	Original	10.74	19.96	26.49	30.72	-	-	-	-
EMB(MB/s)	Proposed	6.04	10.58	13.61	15.50	-	-	-	-
Total	Original	43.78	57.49	66.27	71.62	24.05	42.02	42.02	42.02
EMB(MB/s)	Proposed	39.08	45.90	47.82	48.04	24.05	42.02	24.05	30.04
	Reduction Ratio(%)	10.74	20.17	27.84	32.93	0	0	42.76	28.51
Required Working	Prediction	41.82	49.70	52.94	54.21	29.09	52.73	52.80	52.78
Frequency(MHz)	Update	2.23	3.82	4.85	5.49	-	-	-	-
	Total	44.05	53.52	57.79	59.70	29.09	52.73	52.80	52.78

Table 3. Summary of Implementation Results of Proposed scalable R-D-C MCTF/ME Hardware.

Cell Library		TSMC Artisan 0.18um				
Working Frequency		60 MHz				
Module	IME	FME	Update	Control	Total	
Gate Count	146429	122770	18509	64697	352405	
Buffer Type		Buffer Size(bits)				
SR Buffer	16 Dual-port SRAM 240×32					
MVP Buffer	4 Single-port SRAM 88×16					
CB Buffer	2×2048					
MC Buffer	2048					

2 shows the comparison of required EMB and working frequency for various coding schemes of proposed scalable R-D-C MCTF/ME hardware under above specification. In this table, the EMB and required frequency of 1/3 MCTF and HB are the same with those of prediction stage of 5/3 MCTF, and the original implementation is to apply DRF directly.

By frame-level SR data reuse scheme, the prediction EMB can be reduced, and the reduction ratio increases as the level of MCTF increases. Moreover, the EMBs of H.264/AVC schemes with B-frame are also saved. By ME-based MC scheme with level C+ data reuse, the update EMB can be reduced about 50% compared to that of conventional FMC in worst case. For 5/3 MCTF schemes, 10% to 30% EMB is saved. It is worth to note that the reduction ratio of EMB can be larger for a specification with higher SR.

For the R-D-C properties, the proposed hardware accelerator can adaptively choose the most efficient coding schemes under different available working frequencies and EMBs from Table 2. For example, when the working frequency and EMB are both critically limited, the proposed accelerator can select IPPP coding scheme. However, when only the EMB is limited under 36 MBytes/s, the coding scheme with the best R-D performance among all H.264/AVC and 1/3 MCTF schemes should be applied.

The proposed design is synthesized by SYNOPSIS Design Vision with TSMC Artisan 0.18μm process and the working frequency is 60MHz. Table 3 shows the synthesized results. The total gate

count is about 350K, and the internal buffer is about 30KB. The full search ME, VBSME, Lagrangian mode decision and motion vector predictor in JSVM are supported.

6. CONCLUSION

In this paper, the properties of R-D-C for MCTF and ME are discussed, and system design issues are explored. The first MCTF/ME hardware architecture is proposed to achieve various R-D-C coding points and coding schemes. With the frame-level SR data reuse scheme, proposed frame-interleaved MB pipelining scheme and update stage's design strategies, the total EMB is largely reduced and hardware cost is saved. The implementation results also demonstrates the R-D-C scalability of proposed hardware. In the future, we will extend this work to multi-scale pyramid MCTF scheme in JSVM.

7. REFERENCES

- [1] D. Taubman, "Successive refinement of video: fundamental issues, past efforts and new directions," in *International Symposium on VCIP*, 2003, pp. 791–805.
- [2] *Joint Scalable Video Model (JSVM) 2.0 Reference Encoding Algorithm Description*, ISO/IEC JTC 1/SC 29/WG 11 N7084, Apr. 2005.
- [3] C.-Y. Chen, C.-T. Huang, Y.-H. Chen, C.-J. Lian, and L.-G. Chen, "System analysis of VLSI architecture for motion-compensated temporal filtering," in *Proc. of ICIP*, 2005, pp. 992–995.
- [4] C.-T. Huang, C.-Y. Chen, Y.-H. Chen, and L.-G. Chen, "Memory analysis of VLSI architecture for 5/3 and 1/3 motion-compensated temporal filtering," in *Proc. of ICASSP*, 2005.
- [5] T.-C. Chen, Y.-W. Huang, and L.-G. Chen, "Analysis and design of macroblock pipelining for H.264/AVC VLSI architecture," in *Proc. of ISCAS*, May 2004.
- [6] Y.-W. Huang and etc., "A 1.3TOPS H.264/AVC single-chip encoder for HDTV applications," in *Proc. of ISSCC*, Feb. 2005.
- [7] J.-C. Tuan, T.-S. Chang, and C.-W. Jen, "On the data reuse and memory bandwidth analysis for full-search block-matching VLSI architecture," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 1, pp. 61–72, Jan. 2002.