P4-14

Hardware Oriented Algorithm Analysis and Modification for High Definition AVS Video Encoder VLSI Implementation Digest of Technical Papers

Hai bing Yin, Hong gang Qi, Don Xie, Wen Gao

Abstract--In AVS video coding standard, some algorithms consume huge computation with relatively little coding performance contribution, and some algorithms create data dependencies that are harmful for efficient hardware pipeline. This paper focuses on hardware oriented algorithm analysis and modification. Motion estimation and mode decision algorithms are reviewed and modified to a hardware friendly configuration for high definition (HD) AVS video encoder VLSI implementation. The resulting performance penalties are simulated and analyzed.

I. INTRODUCTION

AVS is the national audio and video coding standard of china. Dedicated AVS video encoder chip is highly desired for consumer applications such as DTV and PVR.

Parallel pipelining technology is widely used in VLSI implementation to improve the data processing throughput [1]. However, intrinsic data dependencies in video coding algorithms disturb the normal pipelining rhythm. Moreover, some algorithms ill-suited for hardware implementation contribute trivial coding performance in HD video cases. Thus, it is very crucial to tailor the algorithms in a hardware friendly configuration for VLSI implementation [2].

II. AVS VIDEO CODING ALGORITHM

AVS is also MPEG-like video coding standard with ME/MC/DPCM and VLC. In AVS, intra prediction with 5 luminance and 4 chrominance modes is done on 8x8 blocks, and only 16x16, 16x8, 8x16, and 8x8 partition modes are used in variable block size motion estimation (VBSME). The available temporal prediction modes include forward, backward, bidirectional (symmetric), and direct. As a result, there are totally fifty MB inter-prediction modes. All intra and inter candidate modes are selected by the mode decision (MD) when rate distortion optimization (RDO) is turned on.

Strong data dependencies exist in AVS coding algorithms. At the block level, one block intra prediction can't initiate until its left, up blocks have finished reconstruction; In RDO based MD algorithm, the tasks for rate and distortion estimation should be processed in turn. The block level spatial motion vector prediction is also context-dependent with the up, left, and up right blocks.

In HD AVS video coding, there are so many MBs to be processed in one second. More awfully, large search window (SW) is desired resulting in high off-chip memory access bandwidth burden and high parallelism of circuit structure.

This work was supported in part by NSFC 60802025 and 60833013.

Also, RDO based MD is challenged with high hardware parallelism due to abundant coding modes.

Our hardware oriented algorithm modifications will mainly focus on motion estimation and RDO based MD.

III. MOTION ESTIMATION ALGORITHM

A. Hierarchical IME

Three-level multiresolution ME (MMEA) is proposed and illustrated in Fig.1. Integer pixel SW [-32, 32]× [-32,32] is used as the illustration example. All pixels in the current MB and the reference pixels in the SW are decomposed into three levels. The middle level (L1) is 4:1 downsampled from the finest level (L0), and the coarsest level (L2) is 4:1 downsampled from level L1. Integer pixel ME (IME) is performed by three successive stages refinement. At stage 1, full search (FS) is performed to check all candidate motion vectors (MV) at level L2 shown using black grids to cover the whole SW. Four winner MVs are selected and used as the refinement centers of level L1 at stage 2, in which four-way parallel searches are performed centered about these four centers respectively. VBSME is not considered at the stage 1 and stage 2, in which FS is performed on the basis of macroblock. Only 16 pixels and 64 pixels attend in SAD calculation at these two stages due to downsampling.

There is strong correlation between MVs of different size blocks within a MB. Almost all MVs of the blocks with different size estimated by FS within the whole SW is located within a small local range (LSW). Thus, VBSME can be performed only within this well-selected LSW instead of the whole SW with negligible performance degradation. If the refinement winner obtained at the IME stage 2 and stage 1 is used as the LSW center, extensive simulations reveal that the LSW size larger than [-16, 16]×[-16,16] is enough with small performance degradation [4].

Then, VBSME is performed at IME stage 3 at level L0 within this LSW in Fig.1-(c) to accelerate the search speed.

B. Simplified Bidirectional Symmetric FME

In AVS, a novel "symmetric" temporal prediction is adopted to save the MV bits. In this mode, only forward MV (mvFw) is coded, and backward MV (mvBw) is estimated proportionally. mvBw and mvFw are all 1/4 pixel MVs. If this mode is adopted in IME and FME, the interpolation computation will be very high, and the normal FME pipeline rhythm is also disturbed.

In the IME stage, although mvFw is integer pixel accuracy,

978-1-4244-4316-1/10/\$25.00 ©2010 IEEE

its corresponding mvBw is 1/4 pixel accuracy. Some cycles are desired to finish the 1/4 pixel interpolation, so this extra cycle consumption challenge the throughput of one candidate MV each cycle at level L0 in IME, which is highly desired to cover the large SW. Thus, symmetric mode is only used in FME in our work, and the forward IME result is used as the symmetric mode FME refinement center. There are eight 1/2 pixel and eight 1/4 pixel candidate MVs to be refined in FME. This extra interpolation computation is acceptable and has no conflict with FME pipeline rhythm.



Fig.1 Illustration of the proposed three-level MMEA.

IV. RDO BASED MODE DECISION

RDO based MD contributes to AVS performance considerably, but its computation complexity is very high.

The reconstructed I frames are used as the reference anchor for the whole GOP, so RDO based MD algorithm is employed for I frame intra prediction. Simple SAD based MD algorithm is employed for intra mode selection in P and B frames. Reconstructed pixels are used as reference pixels for intra prediction in both I, and P/B frames.

Different modes have different selection probabilities. Fig. 2 shows the probability statistics of ten typical 720P format test sequences. The skip/direct mode occurs with the highest probability in general. Also, some modes occur with small probabilities, although the modes vary in different test sequences. So, we can preselect three candidate modes with the largest probabilities based on the SAD criterion. Then, the pre-selected candidate modes, the selected intra mode, and the skip/direct mode are checked using RDO based MD.



V. SIMULATION RESULTS AND CONCLUSIONS

The modified algorithms are tested using four sequences "city", "Sailormen", "Night", and "Spincalendar" of 720P format. IPBBPBB format with GOP length 15 and search range 256×192 are used for 1/4 pixel VBSME with

bidirectional search supported. All inter and intra coding modes and RDO based MD are supported.

We select the algorithm with full search ME and RDO based MD without any simplification as the anchor. Several simplification measures are taken in the proposed algorithm as follows. The modified ME algorithm is marked with A1. The simplified MV prediction in [2] is adopted in our work to break the data dependency in spatial MV prediction. This simplified MV prediction and the proposed symmetric FME algorithms are marked with A2. The simplified SAD based intra mode MD algorithm is marked with A3 in P and B frames. The modified inter mode MD algorithm is marked with A4. The proposed whole algorithm is equivalent to the case that switches A1, A2, A3 and A4 are both turned on.

The PSNR degradation [5] results of the proposed algorithm compared with the anchor are given in Fig.3. According to the results, the PSNR degradation is generally no larger than 0.2dB in most test sequence with low/medium motion and texture such as "city" and "Sailormen". The high motion or texture sequences such as "Night" and "Spincalendar" have relatively high PSNR degradation, and the largest degradation occurs in the case of "Spincalendar", and this maximal degradation is approximately 0.4dB. Because the anchor is the optimal algorithm with full search and RDO based MD without any simplification, so the PSNR degradation of the proposed algorithm is acceptable with hardware friendly compatibility.

The proposed algorithm modifications are well suited for hardware implementation for both Jizhun profile AVS and main profile H.264 video encoders.



Fig.3 PSNR comparison between the proposed and the anchor algorithms

Reference

- Zhenyu Liu, Yang Song, Satoshi Goto etc. HDTV 1080P H.264/AVC Encoder Chip Design and Performance Analysis, IEEE Journal of Solidstate Circuits, Vol.44, no.2, Feb,2009.
- [2] Tu-Chih Wang, Yu-Wen Huang, Hung-Chi Fang and Liang-Gee Chen, "Performance analysis of hardware oriented algorithm modifications in H.264," in IEEE ICASS, Hong Kong, China, April, 2003.
- [3] B. C. Song et al, "Multi-resolution block matching algorithm and its VLSI architecture for fast motion estimation in a MPEG-2 video encoder," IEEE Trans. CSVT vol. 14, no. 9, pp.1119-1137, 2004.
- [4] H. B. Yin et al, VLSI Friendly ME Search Window Buffer Structure Optimization and Algorithm Verification for High Definition H.264/AVS Video Encoder, IEEE Conf. Multimedia and Expo, June 28-July 3, 2009..
- [5] Calculation of Average PSNR Differences between RD-Curves ITU-T VCEG, 2001, Proposal VCEG-M33.