CACHE-BASED INTEGER MOTION/DISPARITY ESTIMATION FOR QUAD-HD H.264/AVC AND HD MULTIVIEW VIDEO CODING

Pei-Kuei Tsung, Wei-Yin Chen, Li-Fu Ding, Shao-Yi Chien, and Liang-Gee Chen

DSP/IC Design Lab, Graduated Institute of Electronics Engineering
National Taiwan University, Taipei, Taiwan

ABSTRACT

To provide more vivid perception, more and more advanced features, like the 4k×2k resolution and the multiview functionality, are emerging for TV. For a multiview video coding (MVC) encoder, motion and disparity estimation (ME/DE) take at least half the hardware requirement. To solve these challenges, a cache-based integer ME/DE algorithm is proposed. With a cache memory as the search window buffer, a predictor-centered ME/DE algorithm is presented. The search range can be reduced to ±16 pixels with less than 0.1dB quality drop compared with full search algorithm. Based on this algorithm, an integer ME/DE chip design is realized. It can reduce 82% on-chip SRAM and 39% system bandwidth. Moreover, the search candidate requirement is also reduced by 79%. As the result, an ME/DE chip design for 4k×2k quad-HD H.264 and HDTV MVC is implemented.

Index Terms—Multiview video coding, H.264, motion estimation, quad high definition, cache memory

1. INTRODUCTION

For advanced TV applications, the vivid perception quality is required. Therefore, a large video resolution, like quad-HD, is recommended. In addition, multiview video can bring the viewers 3D and real perceptual experience by projecting multiple views from different viewing angles to users simultaneously. As the display technology evolves, lots of related applications, like 3D-TV [1] and free-viewpoint TV (FTV) [2] are emerging. However, the computation requirement for MVC sequences is huge especially in the high definition (HD) video specifications. For example, 82.4TOPS computing power and 54.6TB/s memory access are required to encode a 3-view 1080p video with full search algorithm motion estimation (ME). In order to make the real-time multiview applications practical, a hardware accelerator and an efficient multiview video coding (MVC) scheme are needed. MPEG 3D Audio/Video (3DAV) Group is working toward the standardization of MVC. The joint multiview video model (JMVM) is released by the MPEG 3DAV Group as the reference software and the research platform [3]. In the JMVM, H.264/AVC is adopted as the base layer. In addition, the disparity estimation (DE), the most significant feature in JMVM, can effectively exploit the inter-view redundancy and saves 20% to 30% of bit rates. In HD multiview applications, extra on-chip memory and system bandwidth are required to store and load the reference frame data. It makes the hardware implementation for real-time applications more difficult. In this paper, a cache-based ME/DE algorithm for H.264 and MVC is proposed. By adopting the cache memory as reference buffer and the ME/DE algorithm, the SRAM is reduced by 82%, the bandwidth is reduced by 39% and the PE requirement is reduced by 79 % compared with the previous H.264 design [8].

The remaining of this paper is organized as follows: the problem statement and the proposed cache-based ME/DE algorithm is described in Sec. 2. Then, Sec. 3 shows the experimental results and Sec. 4 introduces the implementation result and the comparison. Finally, Sec. 5 summarizes this paper.

2. DESINGE CHANLLENGES AND PROPOSED CACHE-BASED ME/DE ALGORITHM

In an MVC system, all the viewing channels can be divided into two types: the primary views and the secondary views. When processing primary views, the conventional H.264 coding scheme is adapted. On the other hand, disparity-compensated prediction is adopted as the coding tool for the secondary views. In an MVC system, DE is processed like multiple reference frame ME in H.264 to simplify the design.

Our main design challenge is to reduce the on-chip memory requirement. In current H.264 hardware designs, the ME engine processes the current block with an on-chip memory buffering the searching area centered by the zero motion vector (MV). Based on this searching pattern, lots of schemes reusing the overlapped searching area between adjacent block, like level-C and level-C+ [9], can be implemented. These data reusing schemes can save the off-chip memory bandwidth with the price of on-chip memory size. However, in HD video applications, the searching area in each reference frame grows up with the resolution. Take the

1080p stereo view video for example. If the ME search range is [±96, ±64] and the DE search range is half the ME search range, the total memory buffer is 134 Kbyte if the level-C data reusing scheme is adopted. Its area is equal to the area of 2M logic gates while using SRAMs in TSMC 90nm memory compiler. In order to reduce this huge memory requirement, a predictor-centered search with the corresponding cache-based memory management is proposed in this section.

2.1. Overview of the Proposed Cache-Based Algorithm

Figure 1 (a) shows the memory access patterns in primary views of the proposed algorithm. MVs from the neighboring MBs are set as the initial hints first where the hint means the start candidate for ME. The MVs of the left, top-left, top, top-right MBs, zero MV, and the MV predictor defined in H.264 for the 16x16 mode as the initial hints for primary views. The rate-distortion costs of these hints are evaluated, and the hint with the lowest cost is chosen as the best hint. The best hint is then set as the center of the refining range.

Fig. 1 (b) shows the access patterns in secondary views. Besides using hints from the neighboring MBs, MVs from inter-views are also taken as the initial hints. Figure 2 shows the main concept of the inter-view hint generation. After gathering the MVs from the block matching result in the DE stage, inter-view MV hints can be extracted.

The SRAM organization is another important design issue. Assuming that the smallest addressable unit in the cache word is a 4×4-pixel block. When a non-word-aligned 16×16 MB is needed, at most a word-aligned 20×20 pixel block is loaded. Thus, each hint contains 4×4 candidates instead of only one candidate in the cache-based flow since a 4×4 search can be performed for each hint. After the best hint is known, a word-aligned range centered by the best hint word is loaded as the refinement block. According to [4], ±2 pixel refinement is sufficient for inter-view predictors. In the proposed algorithm, the refinement range in the secondary view is modified to 4×4 pixels to fit the cache word. That is, the size of refinement range in secondary views is just the same as the size of the initial hint. Thus, no additional refinement is needed in secondary views.

2.2. Data Prefetch Scheme

To achieve parallel data processing, all variable block sizes share the same hints and the refinement area in the proposed algorithm. The result of the best hint in 16×16 mode is used to all the variable block sizes. Because all the hints in primary views are MV results in neighboring MBs, the data can be known and prefetched to the cache in advance by the MB-level pipeline scheduling shown in Fig. 3. However, the refinement area cannot be prefetched because

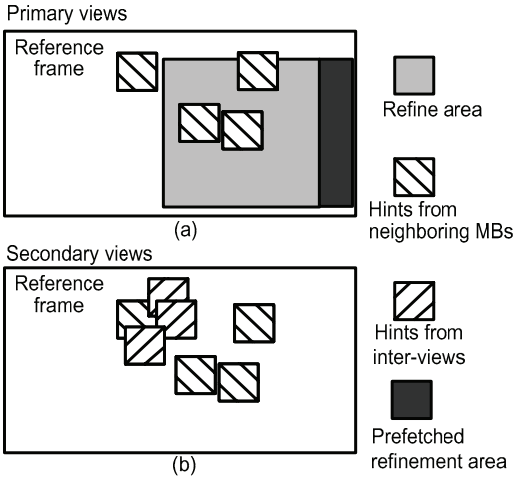


Fig. 1. Memory access patterns in cache-friendly ME.

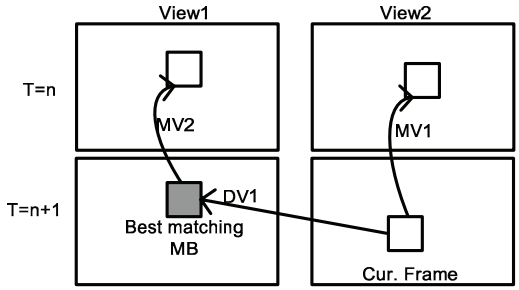


Fig. 2. Concept of the inter-view hint generation. Use MV_2 as the hint for MV_1.

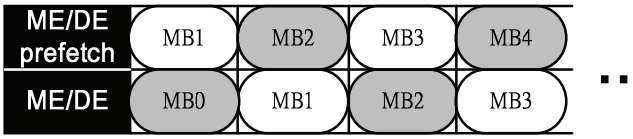


Fig. 3. MB-level pipeline scheduling

it needs the best hint result from the current IME stage. In the proposed algorithm, the refinement area of the current MB is assumed to be next to the refinement area of the left MB. Therefore, the refinement area can be prefetched by a level-C manner, as the dark region in Fig. 1 (a), in the previous pipeline. In the current pipeline stage, only the data that not prefetched needs to be reloaded. In the proposed flow, the on-chip cache size mainly depends on the size of refinement area and cache hit-rate. According to the analysis in our previous work [5], an 8 Kbyte memory buffer can achieve 99.13% hit-rate if the required refinement range is set as ±16 pixels.

Being different from primary views, there are two kinds of the MV hints in secondary views. Hints from neighboring MBs can be prefetched just as hints in primary views. However, inter-view hints from the neighboring views cannot be prefetched in the same way because the value of these hints depends on the DE result. However, the bandwidth over-

head resulted from these hints is only several cache words for reloading a miss-prefetched hint.

2.3. Refinement Area Reduction

Since the required cache size is dominated by the size of the refinement area, an intuitive way to save the memory size and bandwidth is to reduce the refinement area. However, the refinement size affects the R-D performance considerably. Large refinement area can compensate the weak of initial refining center and maintain the R-D performance. In other words, an accurate initial refinement center can reduce the refinement area. Therefore, the motion information preserving scheme is proposed to get more accurate initial hints and refining centers. In the proposed scheme, motion information is saved and reused in the intra-coded MBs. The MV predictor defined in standard H.264 is derived from the MV field. As a result, when an MB is intra-coded, its motion information is not encoded, and no MV is available. However, if the MV pointing to the best matched block is stored even if the intra mode wins the inter/intra mode decision, the MV can still be used as a predictor for neighbor MBs. Therefore, motion information is reused instead of being discarded even if the block is intra-coded.

3. EXPERIMENTAL RESULT

The proposed cache based ME/DE algorithm is implemented by modifying the JMVM 4.4. First, the requirement of refinement range in primary views is analyzed in both the proposed cache based algorithm and the typical zero-centered algorithm described in Sec. 2. Figure 4 shows the experimental result. The horizontal axis in Fig. 4 represents the size of search range, and the vertical axis represents the quality drop compared with the quality under the largest refinement. The quality drop saturates very fast in the proposed algorithm. There is only less than 0.1 dB quality drop even if the search range is ±16. It is because in the most of cases, MVs have small variance with neighboring MBs. Even in the high motion cases, MVs are still similar if the MBs are in the same object. However, the zero-centered algorithm cannot reflect this similarity. According to Fig. 4, the zero-centered search needs a ±128 pixels searching range to compensate the quality drop. However, the proposed one only needs a ±16 pixels search range. Figure 5 express the effect from the proposed refining center decision. The content in the testing sequence is a tractor with high speed. The performance of the algorithm without motion information preserving is much worse than the proposed one. In Fig. 5, intra mode dominates the mode decision in most of the frame. It is because the cost in ME is trapped in the local minimum due to the bad initial hints. On the other hand, the proposed algorithm can still provide the accurate and robust initial hints. According to this advantage, the requirement of the refinement range can remain

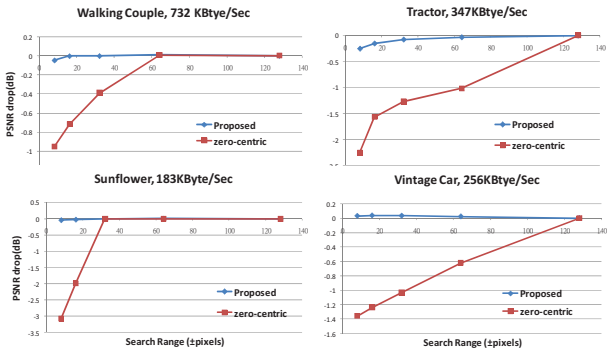


Fig. 4. Comparison of required search area between zero-centered full search and the proposed algorithm.

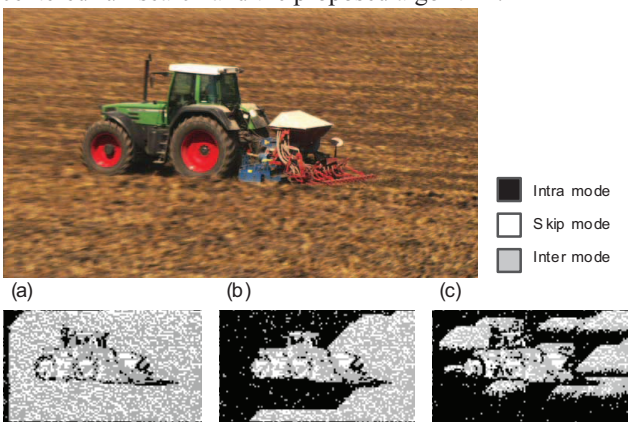


Fig. 5. Block type distribution at Tractor frame #502: (a) search range 16; with proposed refining center decision (b) search range 16; without proposed algorithm (c) search range 4; with proposed algorithm.

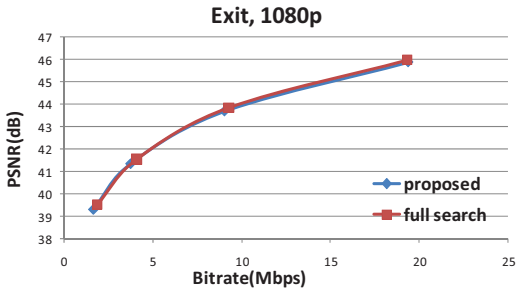


Fig. 6. The Rate-Distortion comparison in secondary views.

small, just like that of slow motion cases. Figure 6 shows the rate-distortion performance in the secondary views. As shown in Fig. 6, although no refinement is performed in the secondary views, there is only less than 0.1dB drop in these views due to the robust inter-view hints.

4. IMPLEMENTATION AND COMPARISON

Figure 7 shows the block diagram of the architecture. The location of hints is generated by snake-scan address generator. After computing the cost for each hint by SAD trees, the

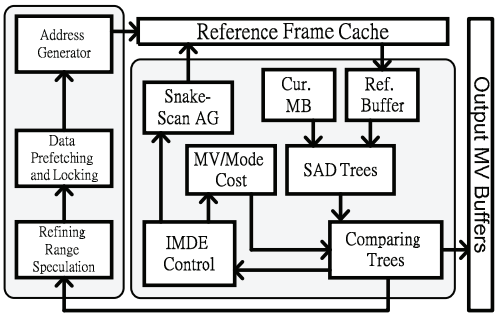
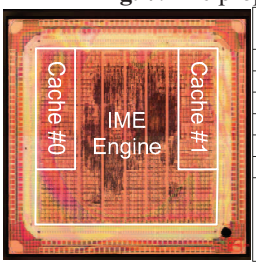


Fig. 7. The proposed ME/DE architecture.



Technology	TSMC 90 nm CMOS LOGIC Low Power LowK Cu 1P9M 1.2&2.5V
Package	COB256 S1
Core Size	2.1 mm x 2.1 mm
Logic Gate Count	230K (2-input NAND gate)
SRAM	64000bits
Max Frequency	300MHz
Power	265mW@300MHz, 1.2V
Processing Capability	4096x2160, single view Super-HDTV H.264;1080p, stereo views HDTV MVC; 720p, four views HDTV MVC

Fig. 8. The chip layout and the specification of the prediction core design.

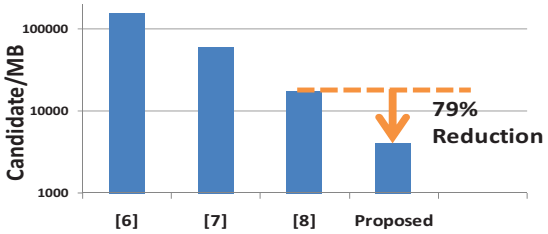


Fig. 9. The candidate requirement in 4k×2k resolution.

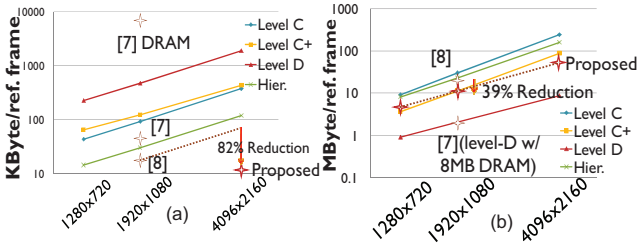


Fig. 10. Comparison between the proposed algorithm and the prior-art: (a) On-chip SRAM (b) Bandwidth

refining center can be speculated. Then, the address generator detects the missing part of the refinement area and reloads it from the cache memory. At the same time, the data prefetching block predicts and prefetchs the refinement area for next MB. Based on the proposed algorithm and architecture, a chip design is implemented. The detail of the specifications and the die photo are shown in Fig.8. The processing capability supports the integer ME for 4096× 2160 pixels quad-HD single view H.264 coding, HDTV 1080p stereo-view MVC, and HDTV 720p four-view MVC. The comparisons with the previous H.264 design [6][7][8] are shown in Fig. 9 and Fig. 10. Compared with [8], the on-chip SRAM area is reduced by 82% and bandwidth is reduced by 39% under the same video resolution. Moreover, the search candidates requirement is about 79% reduced. Thus, the proposed chip can support a larger processing capability to quad-HD specification with similar or even smaller hardware resource.

5. CONCLUSION

A cache-based integer ME/DE algorithm and chip design is proposed in this paper. By use of the cache memory as the reference frame buffer, a cache-based and predictor-centered ME/DE algorithm is realized. By applying the proposed refining center decision algorithm, the required searching range is reduced to ±16 with less than 0.1 dB quality loss. Then, with the inter-view hint generation, the refining range in secondary views can be reduced to 4×4 pixels. The memory requirement, including on-chip SRAM size and external memory bandwidth, is reduced because of the refinement area reduction. Compared with the latest H.264 encoder chip [8], the on-chip SRAM size is reduced by 82% and the system bandwidth is reduced by 39%. Furthermore, the search candidate requirement is also 79% reduced under the same video resolution. Thus, the larger processing capability, 4k×2k quad-HD specification, can be supported with similar or even smaller hardware resources than the prior-art. As a result, an IME accelerator for HD MVC and quad-HD H.264 video coding is proposed.

6. REFERENCES

[1] A. Smolic and P. Kauff, "Interactive 3-D video representation and coding technologies," *Proceedings of the IEEE*, vol. 93, no. 1, January 2005

[2] M. Tanimoto, "Free viewpoint television - FTV," *Proceedings of 2004 Picture Coding Symposium*

[3] MPEG-4 Video Group, "Joint Multiview Video Model (JMVM) 1.0," Number ISO/IEC JTC1/SC29/WG11 N8244, July, 2006, Klagenfurt, Austria.

[4] L.-F. Ding et al, "Fast Motion Estimation With Inter-view Motion Vector Prediction for Stereo and Multiview Video Coding," *Proceedings of ICASSP 2008,* pp. 1373-1376

[5] W.-Y. Chen et al, "Algorithm and Architecture Design of Cache System for Motion Estimation in High Definition H.264/AVC," *Proceedings of ICASSP 2008,* pp. 2193-2196

[6] Y.-W. Huang et al, "A 1.3TOPS H.264/AVC single-chip encoder for HDTV applications," *in Proceeding of IEEE ISSCC, 2005*, pp. 128-129.

[7] Z. -Liu et al, "A real-time 1.41w H.264/AVC encoder SOC for HDTV 1080p," *in IEEE International Symposium on VLSI Circuit Digest of Technical Papers, 2007*, pp. 12-13

[8] Y.-K. Lin et al, "A 242mw 10mm^2 1080p H.264/AVC high-profile encoder chip," *in Proceeding of IEEE ISSCC, 2008*, pp. 314-315

[9] C.-Y. Chen, C.-T. Huang, Y.-H. Chen, L.-G. Chen, "Level C+ Data Reuse Scheme for Motion Estimation with Corresponding Coding Order," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 4, pp553-558, April 2006.