

Multicore and Mesh Network-based Parallel Performance Evaluation using Intra Prediction Algorithms

Yungho Choi¹ and Neungsoo Park^{2*}

¹*Dept. of Electrical Engineering, Konkuk University, Seoul, Korea*

²*Dept. of Computer Science and Engineering, Konkuk University, Seoul, Korea*

¹*yunghoch@konkuk.ac.kr, ^{2*}neungsoo@konkuk.ac.kr*

Abstract

So far, many parallel algorithms have been developed under the assumption that a high performance multicore processor uses a bus for inter-core communications. However, this assumption begins to change as the number of processing cores is increased and thus, higher connectivity among cores is required. So, in this paper, three HEVC intra prediction algorithms are ported into a mesh network-based multicore system by using a wavefront-style parallelization. By analyzing parallel performance, this paper shows that UDIP best fits in the mesh network-based multicore system (almost 2 times faster than other algorithms).

Keywords: *Interconnects, Parallel execution, Intra prediction, Mesh*

1. Introduction

Video compression technology plays a crucial role in implementing multimedia services by enabling to store and transmit video data with less resource [1, 2, 3]. Such compression technologies include MPEG2, MPEG4, and H.264/AVC[4], which reduce data redundancy in subsequent video images. Since, these days, the need for higher image quality of multimedia services has been increased. To accommodate this need, JCTVC has been working on a new draft of a video compression technology, i.e., HEVC (High Efficiency Video Coding). This standard is designed for high resolution videos such as UDTV, which might not be efficiently compressed by previous compression technologies. To enhance compression efficiency, HEVC provides larger and more variable coding block sizes, adaptive loop filter, larger transform kernel size, etc. Despite its high compression efficiency, due to its high coding complexity and tools' dependencies, it might suffer from a long coding time.

When the suitability of algorithms for HEVC is determined, their performances are evaluated under an assumption that the single thread performance of processors will be continuously increased and thus, can support their complicated tools. This is not true because the single thread performance of commercial processors has been at a standstill for a while, migrating into multicore processors. For the feasibility of HEVC, it is important to efficiently implement HEVC in an environment of multicore systems to speed up its complicated coding processes.

This paper explores the performance of intra prediction algorithms proposed for HEVC in a mesh-based multicore system. To do so, this work parallelizes each intra prediction algorithm by employing a wavefront scheme [5] which breaks off dependency chains among coding blocks and enables to execute an intra prediction algorithm concurrently in many cores. Through evaluations, this work will provide an insight on how much parallelism each

* Neungsoo Park is the corresponding author. TEL: +82-2-450-4081

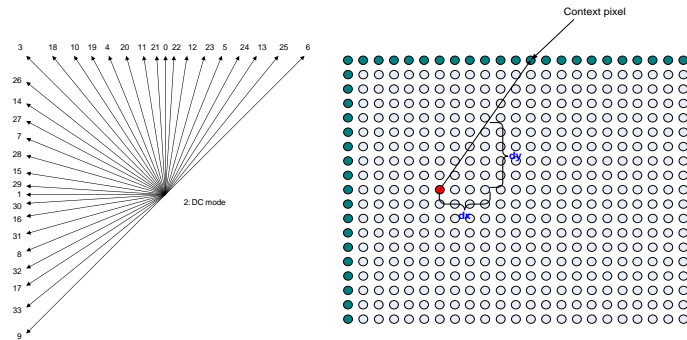
intra prediction algorithm has and how well each algorithm can exploit mesh-based multicore systems.

The remainder of this paper is organized as follows: Section 2 briefly summarizes intra prediction algorithms to be explored. Section 3 describes how to port intra prediction algorithms into a multicore system for their parallel processing performance. In Section 4, a performance evaluation on algorithms is given. In last, this work will be concluded in Section 5.

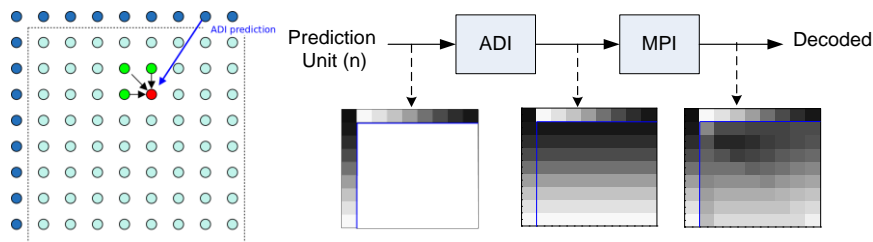
2. HEVC Intra Prediction Algorithms

This work considers three intra prediction algorithms, i.e., Unified Directional Intra Prediction (UDIP) [6, 7], Combined Multi-parameter Intra Prediction (CIP) [2], Planar Intra Prediction (Planar) [7], shown in Fig. 1. UDIP is an intra prediction algorithm which provides more various pixel prediction directions due to larger coding block sizes. This enables to accommodate more various directional image patterns, resulting in a higher prediction accuracy. CIP is similar to UDIP except for providing an additional low pass filtering-like process using multi parameters, thus better predicting natural images. Planar utilizes the most right-bottom pixel of the coding block to predict pixel values impacted by pixels of right and/or bottom sides, which suits for planar images.

Since these algorithms behave differently in terms of data accessing pattern, communication among processing nodes, coding complexity, etc. in a mesh network-based multicore system, it is important to analyze their parallel processing characteristics. For this analysis, the next section describes how to port intra prediction algorithms into a mesh network-based multicore system by parallelization.



(a) Unified directional intra prediction



(b) Combined Multi-parameter intra prediction

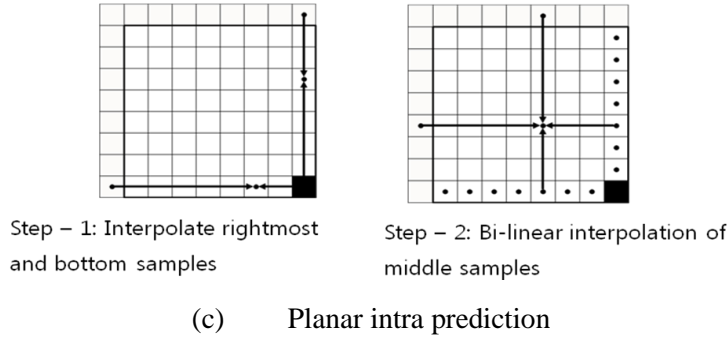


Figure 1. Intra Prediction Algorithms for HEVC

3. Porting of Intra Prediction Algorithms into a Multicore System

To evaluate the parallel performance of each intra prediction algorithm, this work ports intra prediction algorithms into a multicore system consisting of 64 processing cores. These processing cores are connected using a mesh network and on-chip caches. Regarding cache coherence, only local caching is allowed and thus, remote data accesses are expensive. This might decrease the parallel processing performance of an intra prediction algorithm requiring to frequently access remote data. Each processing core in the multicore system has 3 way VLIW pipeline for instruction level parallelism. Additionally, the multicore system provides 4 DDR controllers which mitigate the bottleneck problem of off-chip memory accesses by increasing off-chip memory bandwidth.

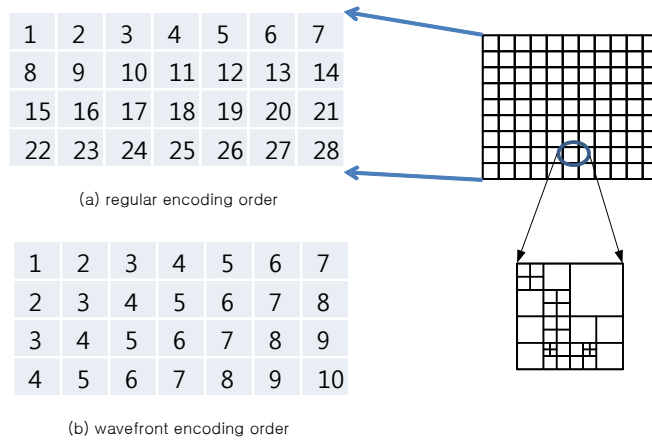


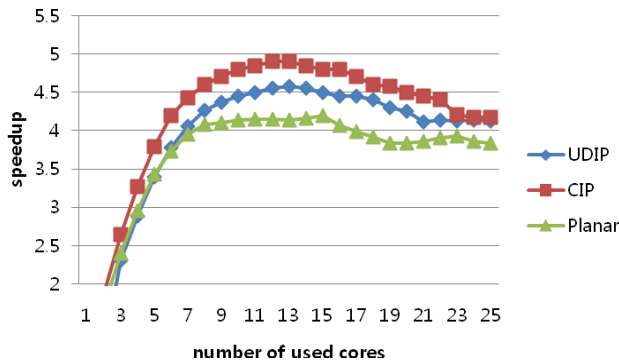
Figure 2. Wavefront-style Parallelization of an Intra Prediction Algorithm

To parallelize intra prediction algorithms, a wavefront-style coding block encoding order [5] is employed. In an intra prediction, spatially neighboring pixels encoded previously are needed to predict macroblock pixel values. However, if coding blocks are encoded in a regular encoding order shown in Figure 2(a), coding block level parallelism cannot be achieved. For example, in Figure 2(a), neighboring coding block 10 and 11 cannot be intra-predicted concurrently because, to predict coding block 11, coding block 10's reconstructed picture image is required, which is available only after coding block 10 is encoded. a wavefront-style coding block encoding order [5] can resolve this problem by changing encoding order as

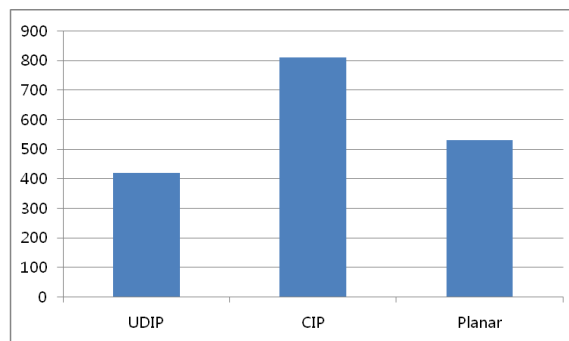
shown in Figure 2(b) and thus, enabling a coding block level parallelism. For example, in Figure 2(b), when coding block 4s are encoded, the reconstructed pictures of coding block-3s are available because they are previously encoded. In the next section, using a multicore system and a wavefront-style coding block encoding order, an evaluation on a multithreaded performance of intra prediction algorithms is given to identify the algorithm providing the highest parallel processing performance in a multicore system.

4. Mesh Network-based Multicore system-based Parallel Performance Evaluation of HEVC Intra Prediction Algorithms

This section evaluates the parallel processing performance of three intra prediction algorithms by parallelizing the algorithms implemented in TMuC1.0 [8] using a pthread API. The parallel algorithm employed for these intra prediction schemes is the wavefront algorithm [5] that enables a parallel processing at the level of coding blocks, shown in Figure 3. For a parallel processing performance evaluation, each algorithm is executed in a multicore system under the following conditions: PeopleOnStreet(2560x1600) sequence, 64x64 maximum coding unit size, Intra-only configuration, 100 frames, QP=32. The execution time and the speedup of each intra prediction algorithm are given in Figure 3. Speedup is calculated by dividing single-thread execution time of an algorithm with its parallel processing time.



(a) Speedup of intra prediction algorithms



(b) Execution time of prediction algorithms (ms/frame)

Figure 3. Parallel Performance of Intra Prediction Algorithms

As shown in Figure 3(a), CIP shows the highest speedup, which means the highest parallelism and the best exploitation of the given mesh network-based multicore system (up to 25% higher speedup than other algorithms). As the number of processing cores increases, the speed-up of algorithms is saturated around 8 cores. The one of the reasons is the local cache policy of the given multicore system which does not allow to cache remote data homed in another cores. Therefore, as the number of cores increases, the data traffic between cores rapidly aggravates and makes the underlying mesh network saturated. This problem decreases the speed-up for each algorithm, significantly when more than 13 cores concurrently process coding block intra predictions. This indicates that, to implement high performance HEVC, a careful optimization of communications among processing cores is required.

Figure 3(b) shows parallel execution times of algorithms. As shown, UDIP can be executed up to two times faster than other algorithms. This means that UDIP has the lowest coding complexity and does not suffer from mesh network congestion significantly, while, although CIP shows the highest speed-up, its coding complexity is much larger than other algorithms, ending up with the longest execution time.

5. Conclusion

This work parallelizes three intra prediction algorithms for a mesh network-based multicore system, and, using these implementations, evaluates the parallel performance of intra prediction algorithms. Through this work, it is shown that, when intra prediction algorithms are parallelized and executed concurrently in multiprocessor systems, the network connecting processing cores might be a performance limitation by being saturated.

Overall, UDIP can be best executed in the given multicore system (almost 2 times faster than other algorithms). Additionally, in spite of the highest speedup, the high coding complexity of CIP increases its execution time. In this work, drawn is a conclusion that, in selecting an algorithm for HEVC, taking into accounts both the coding complexity and the parallelizability of algorithms is necessary.

References

- [1] M. Kim and I. Jee, "H.264 Encoding Technique of Multi-view Image expressed by Layered Depth Image", the Journal of IWIT, vol. 10, no. 1, (2010), pp. 81-90.
- [2] B. Bae and Y. Kim, "A Study on Controlling of Target Transfer Rate in Terrestrial DMB Systems", the Journal of IWIT, vol. 10, no. 3, (2010), pp. 19-26.
- [3] Y. Im and E. Kang, "MPEG-2 Video Watermarking in Quantized DCT Domain", the Journal of IWIT, vol. 11, no. 1, (2011), pp. 81-86.
- [4] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC video coding standard", IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 7, (2003), pp. 560-576.
- [5] Z. Zhao and P. Liang, "Data partition for wavefront parallelization of H.264 video encoder", IEEE International Symposium on Circuits and Systems, (2006), pp. 4-2672.
- [6] K. McCann, W. Han, I. Kim, J. Min, E. Alshina, T. Lee, J. Chen, V. Seregin, S. Lee, Y. Hong, M. Cheon and N. Shlyakhov, "Video Coding Technology proposal by Samsung", JCTVC-A124, Dresden, German, (2010) April.
- [7] K. Ugur, R. Andersson and A. Fuldseth, "Video Coding Technology Proposal by Tandberg, Nokia and Ericsson", In: JCTVC-A119, Dresden, German, (2010) April.
- [8] HEVC Reference Software TMuC1.0, https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/HM-1.0.

Authors



Yungho Choi

He received his BS and PhD degrees from Yonsei University and University of Southern California, in 1991 and 2001, respectively. He was a senior Engineer in Intel, Shrewsbury, USA. Currently, he is an associate professor in the Department of Electrical Engineering, Konkuk University, Seoul, Korea. His research interests include multiprocessor systems, high performance codec architectures, interconnection networks as well as image recognition.



Neungsoo Park

He received his Ph.D. in Electrical Engineering from the University of Southern California in 2002. He also received his BS and MS in Electrical Engineering from Yonsei University, Seoul, Korea, in 1991 and 1993 respectively. Currently, he is an associate professor in the Department of Computer Science and Engineering, Konkuk University, Seoul, Korea. Prior to joining Konkuk University, he was a senior engineer in Samsung Electronics Co. Ltd., Korea. His research interests include parallel computing, computer architecture, embedded system, high performance computing for signal processing, VLSI computations, and multimedia systems, etc.