Algorithms and hardware design for key technologies in video coding

FU, Chen

(開始ページ / Start Page)
1
(終了ページ / End Page)
128
(発行年 / Year)
2024-03-24
(学位授与番号 / Degree Number)
32675甲第604号
(学位授与年月日 / Date of Granted)
2024-03-24
(学位名 / Degree Name)
博士(工学)
(学位授与機関 / Degree Grantor)
法政大学 (Hosei University)
(URL)
https://doi.org/10.15002/00030499

博士学位論文
論文内容の要旨および審査結果の要旨

論文題目	Algorithms and hardware design for key technologies in video coding
氏名	FU Chen
学位の種類	博士（工学）
学位番号	第 604 号
学位授与年月日	2024 年 3 月 24 日
学位授与の要件	法政大学学位規則第 5 条第 1 項第 1 号該当者（甲）
論文審査委員	主　査　周　金佳　　准教授
	副　査　和田　幸一　　教授
	副　査　尾川　浩一　　教授

１．論文内容の要旨

Video encoding plays a crucial role in compressing extensive video data, facilitating efficient storage and transmission. Context-Adaptive Binary Arithmetic Coding (CABAC) poses a bottleneck for the throughput in video coding due to its strong inter-symbol dependency, resulting in higher computational complexity and slower processing speed. Moreover, accessing off-chip DRAM for reference frame reading and writing dominate the power consumption of the entire video coding system. Finally, new techniques are desired to enhance video coding efficiency by achieving higher compression ratios while keeping video quality.

To achieve higher throughput, lower power consumption, and improved compression ratio, this dissertation develops three key technologies: 1) By employing strategic approaches, this research successfully overcomes throughput bottlenecks, achieves a more efficient CABAC component, and realizes a throughput of 2191Mbin/s, enabling real-time decoding for 8K UHDTV. 2) The proposed reference frame recompression scheme significantly reduces the traffic for reading/writing reference frames from DRAM, improving throughput/gate efficiency, achieving real-time processing at 8K@120fps, and greatly reducing DRAM power consumption. 3) By integrating deep learning-based dynamic talking-head generation technology, this research greatly enhances the realism and emotional expressiveness of online video conferencing. This approach also provides a novel solution for improving video compression ratio. This thesis includes the following chapters.

Chapter 1 [Introduction] gives a concise overview of the research background in video encoding, highlighting the significance of this study. It briefly outlines the methods proposed, summarizes the results, and discusses the contributions of these methods.

Chapter 2 [Algorithms and Hardware Design of Entropy Coding] focuses on the principles of entropy coding. Context-Adaptive Binary Arithmetic Coding (CABAC) as an entropy coding algorithm, is the throughput bottleneck of the whole video encoder due to its strong data dependencies. The chapter discusses the implementation and optimization of the entropy coding syntax element module at a hardware level and introduces the highly pipelined and parallel VLSI architecture of the CABAC encoder tailored for UHDTV applications. Firstly, this research proposes a prediction-based context model prefetching strategy, effectively eliminating the clock consumption of the contextual model for accessing data in memory. Moreover, a multi-result context model update (MCMU) is applied to reduce the critical path delay of context model updates in multiple bin/clock architecture. Furthermore, a pre-range update and pre-renormalize techniques are applied to reduce the multiplex binary arithmetic encoding's route delay due to the incomplete reliance on the encoding process. Moreover, to further speed up the processing, this research proposes to process 4 regular and several bypass bins in parallel with a variable bypass bin incorporation (VBBI) technique. Finally, a quad-loop cache is developed to improve the compatibility of data interactions between the entropy encoder and other video encoder modules. This chapter concludes with an evaluation of the proposed methods and a summary of the significant findings in entropy coding.

Chapter 3 [Algorithms and Hardware Design of Reference Frame Recompression] introduces the architecture of lossless reference frame re-compression and addresses current challenges. Accessing off-chip DRAM for reference frame reading and writing dominate the power consumption of the entire video coding system. The proposal of reference frame recompression can reduce the DRAM bandwidth, and then reducing the power of reference frame reading and writing. This chapter presents an adaptive detachable partition based RFRC scheme which can compress a variable-size to a fixed bit of access unit. Comparing with the conventional schemes which compress fixed-size partition to variable bits, this work can increase the compression ratio by greatly reducing the redundancy of the partitions' boundaries. Moreover, a high-speed coding algorithm is applied to reduce the complexity and increase the parallelism of compressing and decompressing. Finally, a new memory mapping strategy is utilized to support random access. This chapter presents the algorithm design for lossless FRC

and the tailored VLSI architecture design, demonstrating the effectiveness of the proposed hardware architecture in reducing data volume.

Chapter 4 [Algorithms of Generation based Video Coding] explores learning-based speech emotion recognition and audio-driven single-image talking face generation. It details a dual-level neural network architecture for speech emotion recognition and introduces a novel talking-head generation scheme that uses speech emotion and intensity recognition to enhance realism. By adopting this strategy, a whole video can be generated by one image and one audio. The chapter evaluates the proposed methods and discusses the main results, ablation experiments, and their implications for talking-head generation.

Chapter 5 [Conclusion] concludes the dissertation by summarizing the key contributions, discussing the implications of the findings, and suggesting future research directions in video encoding, reference frame re-compression, and dynamic talking-head generation.

As mentioned above, this research proposes algorithms and hardware design solutions for the key technologies in video coding. The implementation achieves higher throughput, lower power consumption, and an improved compression ratio for the video coding process.

２．審査結果の要旨

To enhance the throughput of video compression, reduce power consumption, and improve compression ratios, this dissertation puts forward algorithms and hardware designs for key techniques in video coding.

1. To improve the throughput, CABAC, which serves as the bottleneck for video coding throughput, is optimized in hardware by three proposals. The first proposal is a novel prediction-based context model prefetching. It can shorten the critical path, thereby increasing the working frequency. The second method utilizes multi-result context model updates (MCMU) and pre-renormalize techniques. It can further reduce the critical path and support parallel processing. The third strategy incorporates variable bypass bin incorporation (VBBI), further increasing the parallelism. These strategies can increase the working frequency and enhance the parallelism. As a result, the throughput is greatly improved to 2191Mbin/s, meeting real-time 8K UHDTV video encoding needs.

2. To reduce the power consumption of the video codec system, this dissertation optimizes the component responsible for accessing off-chip DRAM for reference frame reading and writing, which dominates the overall power consumption. An adaptive detachable partition-based reference frame recompression scheme is proposed. It can provide a high compression ratio by utilizing the detachable compression unit. The

compressor and decompressor respectively achieve maximum efficiency improvements of 1.6 times and 3.4 times in terms of samples/cycle/gate, surpassing the state-of-the-art results. The power of the reference frame reading and writing is reduced by more than 50%. The proposed architecture is also verified in a whole decoder. It can support real-time processing of 8K 120 frames/s for the low-delay-P configuration of H.265/HEVC with 300MHz.

3. To increase the compression ratio of conference videos, learning-based video generation techniques are proposed. A dual-level model for speech emotion recognition combined with dynamic talking-head generation is developed. This model offers a more nuanced and accurate emotion recognition, with an industry-leading accuracy rate of 73.5%. It marks a new direction in synthesizing emotionally expressive talking heads, enhancing the realism and expressiveness of online video conferencing.

Based on all of these, this examination committee is unanimous that the submitted doctoral thesis is fully qualified as a Doctor of Philosophy (Engineering).