Signal Processing: Image Communication 47 (2016) 271–279

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

A fast inter coding algorithm for HEVC based on texture and motion quad-tree models

Fen Chen, Peng Li, Zongju Peng*, Gangyi Jiang, Mei Yu, Feng Shao

Faculty of Information Science and Engineering, Ningbo University, Ningbo, Zhejiang, China

ARTICLE INFO

Article history:
Received 30 January 2016
Received in revised form
25 June 2016
Accepted 6 July 2016
Available online 7 July 2016

Keywords:
High efficiency video coding (HEVC)
Fast inter coding
Motion property
Quad-tree models

ABSTRACT

In order to reduce the encoding complexity of the emerging high efficiency video coding (HEVC) standard, a fast inter coding algorithm for HEVC based on texture and motion quad-tree models is proposed. First, the coding units are classified into motion or static blocks according to a motion/static detection algorithm. Then, different encoding schemes are adopted for each type of block. In particular, we can make full use of statistical properties and temporal correlations to determine the depth range and prediction mode of static largest coding tree units (LCTUs). Moreover, fast coding unit size and mode decision are made by means of quad-tree models and spatiotemporal correlations for the motion LCTUs. Finally, we merge the above schemes together to gain better performance. Experimental results show that the proposed overall algorithm, compared with the original HEVC encoding scheme, reduces encoding time by 47.5% on average with a Bjonteggard delta bit rate increase of only 1.6% for various test sequences under random access condition. Compared with a state-of-the-art algorithm, the proposed method can save more encoding time while maintaining comparative rate distortion performance.

1. Introduction

The demand for high definition (HD) and ultra-HD (UHD) multimedia services in TV broadcasting, internet video, or even mobile applications is already imminent. However, the problems of high-bandwidth video transmission and large-capacity storage required by these HD and UHD videos have remained major research subjects in recent years. To address these problems, the Joint Collaborative Team on Video Coding (JCT-VC) developed the high efficiency video coding (HEVC) standard [1] in January 2013, which is a successor to H.264/AVC (Advanced Video Coding) [2]. The goal of HEVC is to achieve about a 50% bit-rate reduction over high profile H.264/AVC given the same objective video qualities [3].

The hierarchical quad-tree structure based on the coding tree unit (CTU) has been adopted in HEVC. The CTU is one of the most powerful tools available for improving HEVC coding efficiency. It includes the coding unit (CU), prediction unit (PU), and transform unit (TU) to describe the overall encoding process. HEVC intra prediction adopts a quad-tree structure based coding technique and multiple-angle intra prediction to improve coding efficiency, while the high-precision motion compensation based on variable PU size, adaptive motion vector prediction, and merging techniques are employed in the inter coding. These new techniques significantly improve the compression efficiency of HEVC [4]. However, they lead to intensive computational complexity, which hinders real-time applications of HD and UHD videos [5]. Therefore, it is important to reduce the complexity of the HEVC encoders as much as possible without video quality degradation.

HEVC inter coding makes full use of temporal correlation to reduce coding redundancy, further improving compression efficiency on the basis of intra coding. Because of the introduction of new techniques in HEVC, many fast algorithms designed for H.264/AVC cannot be applied to HEVC in a straightforward manner. Recently, there have been quite a few related methods aiming to reduce the complexity of the HEVC encoders. They can be roughly divided into three categories: fast CU size decision, fast mode decision, and combination of both. In this paper, we focus on combining fast decision for both CU size and prediction mode. Gweon et al. [6] proposed a fast coded block flag (CBF) algorithm that utilized the zero CBF of the current PU to early terminate the mode decision of the next PU. Cho et al. [7] proposed an early termination method based on the property of the Skip mode. Yang et al. [8] proposed an early Skip determination method based on motion vector difference and CBF. The above algorithms [6–8] take advantage of special encoding parameters to terminate the CU or PU mode decisions early, reducing the encoding complexity to some extent. However, the number of CUs satisfying the conditions of these special encoding parameters is small, the encoding time saving is quite limited.

* Corresponding author.
E-mail address: pengzongju@126.com (Z. Peng).

http://dx.doi.org/10.1016/j.image.2016.07.002
0923-5965/

There have been many investigations about fast mode decisions for inter prediction in HEVC. Shen et al. [9] presented an adaptive inter-mode decision algorithm for HEVC that jointly utilizes inter-level and spatiotemporal correlations. Based on statistical analyses, three methods were proposed: early Skip mode decision, prediction size correlation-based mode decision, and rate-distortion (RD) cost correlation-based mode decision. Lee et al. [10] put forward an early Skip mode decision method by utilizing its distortion characteristics after calculating the RD cost of a $2N \times 2N$ Merge mode. Ahn et al. [11] proposed a fast encoding scheme for HEVC inter coding which fully utilizes spatiotemporal encoding parameters to early determine CU size and Skip mode. This scheme uses sample-adaptive-offset parameters as the spatial encoding parameter to estimate the texture complexity and uses motion vectors, TU size, and CBFs as the temporal encoding parameters to measure motion complexity.

In fact, similarly to H.264/AVC, the most time-consuming computation in HEVC is the inter prediction process, which includes motion estimation (ME). Therefore, many fast ME algorithms have also been proposed to reduce inter-slice coding complexity. Jou et al. [12] proposed a fast ME algorithm aimed at real-time video encoding, which adopts a predictive integer ME algorithm that selects the most probable search direction and employs a PU size-dependent fractional ME to reduce interpolation filtering. Lee et al. [13] proposed a fast ME algorithm using a priority-based inter mode decision method. This algorithm decided whether to perform ME by calculating the priority of each mode, reducing the encoding complexity of ME process by up to 55.51%. In addition, there are many algorithms utilizing parallel computing to reduce the computational complexity of encoders. Yan et al. [14] proposed a parallel computing framework for HEVC CU partitioning tree decision on many-core processors. The framework achieves 11 and 16 times speedup for 1920×1080 and 2560×1600 video sequences respectively. Yan et al. [15] also proposed a parallel framework to decouple ME for different partitions on many-core processors, and 30 and 40 times speedup for 1920×1080 and 2560×1600 video sequences have been achieved. The above algorithms [6–15] mainly focus on texture characteristics, motion properties, spatiotemporal correlations, or statistical information about the CU size and prediction modes to speed up HEVC encoders. However, they do not make full use of these information, and most of them cannot consider the motion and edge information of different CU depth levels. Hence, there is still considerable space for improvement in HEVC encoders.

To overcome the limitations of the above algorithms, this paper proposes a fast inter coding algorithm based on texture and motion quad-tree models, considering the motion and edge information of each CU depth level. First, the proposed algorithm classifies the current Largest Coding Tree Unit (LCTU) into motion or static blocks according to a motion/static detection algorithm. Then, different fast encoding schemes are adopted for each type of block. A fast encoding scheme for static LCTUs is performed to determine their depth range and candidate mode set using statistical properties and temporal correlations of CU size and prediction modes. For the motion LCTUs, a fast CU splitting scheme is proposed that utilizes a texture quad-tree model (TQM) and temporal correlations of CU size, while the proposed fast mode decision scheme uses a motion quad-tree model (MQM) and spatial correlations of the best prediction mode. This paper has two main contributions: 1) We propose two quad-tree models (MQM and TQM) by fully utilizing the motion characteristics as well as the texture edge information of each CU depth level. Further, we propose an algorithm to early determine the probable depth range and mode set based on these two models so that we can speed up the encoding process with negligible loss in RD performance. 2) We develop different strategies for motion and static blocks based on the statistical experiments and analysis.

The rest of this paper is organized as follows. In Section 2, the HEVC inter coding complexity and statistics of static and motion LCTUs are analyzed in detail. In Section 3, we first briefly introduce the overall framework of the proposed algorithm, and then present two encoding schemes targeting static LCTUs and motion LCTUs in detail, respectively. The experimental results and analyses are given in Section 4. Finally, conclusions are made in Section 5 along with a summary of the proposed algorithm.

2. Inter coding complexity and statistical analyses in HEVC

In the HEVC standard, intra or inter prediction coding is performed on a quad-tree based block structure. The recursive block in HEVC is called the CTU, which can be 64×64, 32×32 or 16×16 in size. Of course, CTU can further split into four equally sized CUs, which is the leaf node of a quad-tree partitioning of the CTU. In the encoding process, the best combinations of CUs, PUs, and TUs are determined in the sense of RD optimization (RDO). In HEVC intra coding, each 8×8 PU can be divided into four partitions of size 4×4. Therefore, the original HEVC encoding scheme needs to calculate the RD-cost value 341 ($4^0+4^1+4^2+4^3+4^4=341$) times in order to determine the best combination of CUs and PUs. As shown in Fig. 1, the LCTU recursively splits into sub-CUs from depth 0 to depth 3. Each CU can be further divided into PUs. The HEVC encoder has 12 different PU modes, including Skip mode, Merge mode, Inter_$2N \times 2N$, Inter_$2N \times N$, Inter_$N \times 2N$, Inter_$N \times N$, asymmetric modes (Inter_$N \times nU$, Inter_$2N \times nD$, Inter_$nL \times 2N$, and Inter_$nR \times 2N$), and Intra modes (Intra_$2N \times 2N$ and Intra_$N \times N$). Every PU except for the Intra PUs needs to perform ME to determine the best PU mode. The computational complexity of the ME process is quite high. In conclusion, CU size decision, inter mode decision, and the ME process lead to huge computational complexity of the Inter encoding process, limiting its use in real-time applications.

In order to design a fast HEVC encoding algorithm, we perform explorative experiments on the latest HEVC reference software (HM-16.9) and statistically analyze the experimental results. The Random Access-Main (RA-Main) profile was used and four QPs {22, 27, 32, and 37} are tested. RDO is enabled. The test sequences [16] provided by JCT-VC, including Traffic (2560×1600), Kimono (1920×1080), PartyScene (832×480), BlowingBubbles (416×240) and Johnny (1280×720), are used in the statistical experiment. We divide the test sequences set in two parts, and use one for the statistics and the other for the encoding. We find that the following phenomena and rules exist in HEVC.

(1) CU splitting results are different in various frames. Under the RA-Main test condition, a hierarchical B structure is used for encoding. The frame types include I frame, Generalized P and B (GPB) frames, reference B frames and non-reference B frames. Fig. 2 shows the optimal CU splitting of different frame types in the “BlowingBubbles” sequence. The CUs in the I and GPB frames have the smallest size, while those in the reference B slice have a relatively larger size, and those in the non-reference B slice have the largest size. The underlying reason of the phenomena is that the CU splitting is relative to the importance of these frames. I and GPB frames have the highest importance, followed by the reference B frames. Non-reference B frames have the least importance.

(2) Optimal depth distribution of static LCTUs in reference B frames is different from that in non-reference under different QPs. The extraction method of static LCTUs will be introduced in the succeeding section. The statistical results is shown in Fig. 3. The optimal depth range of most static LCTUs in both