J. Vis. Commun. Image R. 45 (2017) 170–180

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Fast depth map mode decision based on depth–texture correlation and edge classification for 3D-HEVC ☆

Qiuwen Zhang [a,*], Na Zhang [a], Tao Wei [b,c], Kunqiang Huang [a], Xiaoliang Qian [d], Yong Gan [a]

[a] College of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China
[b] China National Digital Switching System Engineering & Technological R&D Center, Zhengzhou 450002, China
[c] Computer College, Henan Institute of Engineering, Zhengzhou 451191, China
[d] College of Electric and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China

ARTICLE INFO

Article history:
Received 24 December 2016
Revised 17 February 2017
Accepted 3 March 2017
Available online 6 March 2017

Keywords:
3D-HEVC
Depth map
Mode decision
Edge classification

ABSTRACT

The 3D extension of High Efficiency Video Coding (3D-HEVC) has been adopted as the emerging 3D video coding standard to support the multi-view video plus depth map (MVD) compression. In the joint model of 3D-HEVC design, the exhaustive mode decision is required to be checked all the possible prediction modes and coding levels to find the one with least rate distortion cost in depth map coding. Furthermore, new coding tools (such as depth-modeling mode (DMM) and segment-wise depth coding (SDC)) are exploited for the characteristics of depth map to improve the coding efficiency. These achieve the highest possible coding efficiency to code depth map, but also bring a significant computational complexity which limits 3D-HEVC from real-time applications. In this paper, we propose a fast depth map mode decision algorithm for 3D-HEVC by jointly using the correlation of depth map-texture video and the edge information of depth map. Since the depth map and texture video represent the same scene at the same time instant (they have the same motion characteristics), it is not efficient to use all the prediction modes and coding levels in depth map coding. Therefore, we can skip some specific prediction modes and depth coding levels rarely used in corresponding texture video. Meanwhile, the depth map is mainly characterized by sharp object edges and large areas of nearly constant regions. By fully exploiting these characteristics, we can skip some prediction modes which are rarely used in homogeneity regions based on the edge classification. Experimental results show that the proposed algorithm achieves considerable encoding time saving while maintaining almost the same rate-distortion (RD) performance as the original 3D-HEVC encoder.

1. Introduction

Three-dimensional (3D) video has received more and more attention by the development of 3D content acquisition and display technologies in recent years. Multi-view video plus depth (MVD) is one of the most promising 3D video representations to support depth perception of 3D scenes [1,2]. In the MVD systems, a small number of captured texture video and its corresponding depth map are coded and the resulting bitstream packets are multiplexed into a 3D video bitstream [3,4]. After decoding texture video and depth map, much more intermediate virtual views suitable for displaying the 3D content on a free-viewpoint display can be synthesized using a depth-image-based-rendering (DIBR) technique [5]. Especially, the additional intermediate virtual view quality (rendered by the DIBR) highly depends on the coding result of depth map. Thus, high efficient depth map coding is most crucial to realize the 3D video practical applications.

As a result, efficient depth map coding has been recently investigated by Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) established by international organization for standardization (ISO) MPEG and the international telecommunication union (ITU) VCEG, and 3D extension of high efficiency video coding (3D-HEVC) is developed for the effective compression of depth map data [6]. Different from the conventional texture video compression [7], it is key point to preserve the depth sharp object edges rather than the depth map visual quality. Based on this characteristic, 3D-HEVC introduces several new prediction modes and coding tools to preserve the sharp object edges in depth map coding, such as the Depth Modeling Mode (DMM) [8],

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author at: College of Computer and Communication Engineering, Zhengzhou University of Light Industry, No. 5 Dongfeng Road, Zhengzhou 450002, China.

E-mail address: zhangqwen@126.com (Q. Zhang).

http://dx.doi.org/10.1016/j.jvcir.2017.03.004
1047-3203/

segment-wise depth coding (SDC) [9], and motion parameter inheritance (MPI) [10]. Meanwhile, a computationally expensive exhaustive quadtree coding structure of HEVC is also used in 3D-HEVC depth map coding. Hence, the 3D-HEVC mode decision is required to be checked all combinations of conventional HEVC prediction modes [11–13] with the additional new tools to find the one with least rate distortion cost in depth map coding. These techniques achieve the highest possible coding efficiency but require a very high computational complexity, which limit 3D-HEVC from the practical applications. Therefore, fast algorithms, which can reduce the complexity of depth map coding without sacrificing rate distortion (RD) performance, are extremely necessary for 3D-HEVC real-time applications.

A number of fast algorithms [14–19] have been proposed to reduce the depth map computational complexity for previous video coding standards (such as H.264/AVC and its extension of multi-view video coding (MVC)), achieving significant time savings with acceptable video quality degradation. However, these fast algorithms are not directly applicable to the new standard 3D-HEVC, which high computational complexity is intrinsically related to the use of new prediction modes and coding tools in 3D-HEVC encoder. Recently, several studies on the reduction of depth map coding complexity have been reported for 3D-HEVC encoders in the literature. A fast depth map wedgelet partitioning scheme is presented in [20] based on adaptively utilizing the mode with minimal cost in rough mode decision of HEVC intra prediction. Fast depth mode decision algorithms are proposed in [21,22] to early terminate the unnecessary prediction modes with full RD cost calculation in 3D-HEVC. A fast mode decision algorithm based on simplified edge detector is proposed in [23] to reduce the complexity of the 3D-HEVC depth intra prediction. Fast depth mode decision algorithms are employed in [24,25] to selectively omit unnecessary DMM based on the pre-calculated RD costs of the HEVC intra modes and the edge classification. A fast depth intra mode decision is introduced in [26] to reduce the depth intra complexity in a smooth region. A fast algorithm is designed in [27] based on the early SKIP mode detection and the prediction size correlation based-mode decision to reduce 3D-HEVC encoding time for real-time applications. A flexible block partitioning is employed in [28] to efficiently represent the depth map smooth areas delimited by sharp edges. A fast depth map coding algorithm is proposed in [29] to reduce the computation complexity of the 3D-HEVC encoder by utilizing early Skip and early DIS scheme. A fast depth map quadtree structure determination scheme is designed in [30] to terminate the quadtree-based partition process of coding tree unit(CTU) as early as possible. A fast mode decision algorithm based on the grayscale similarity and inter-view correlation is proposed in [31] to reduce the complexity of depth map coding by skip unnecessary mode checking within the mode decision process. Two fast algorithms including the squared euclidean distance of variances (SEDV) and probability-based early depth intra mode decision (PBED) are presented in [32] to speeding up the most time-consuming intra mode processes in 3D-HEVC depth map coding. Effective early termination and intra mode decision algorithms are also developed in our previous work [33,34] to reduce the depth map coding complexity of 3D-HEVC encoders. The aforementioned algorithms are well developed for depth map coding achieving significant time savings in 3D-HEVC. However, most of these fast algorithms have not adequately exploited the correlation of depth map and texture video. Additionally, the characteristics of the depth map are not fully studied. This situation results in a limited time saving. Furthermore, most of the previous fast depth map algorithms are not designed for the recent 3D-HEVC test model HTM-16.0 [4]. There is still some room for further reduce mode decision complexity of the 3D-HEVC depth map coding.

To further relieve the computation complexity of depth map mode decision, this paper proposes a fast depth map mode decision algorithm for 3D-HEVC encoders based on the depth–texture correlation and edge classification. The main idea of the proposed algorithm is that the correlation of depth map–texture video and the edge information of depth map are used to analyze the current depth map coding unit (CU) prediction mode and early skip unnecessary variable-size mode decision. It consists of four fast mode decision strategies: adaptive depth map coding levels determination, early depth map SKIP/Merge mode detection, fast depth map inter mode size decision and Selective depth map intra prediction. Extensive experimental results demonstrate that the proposed fast mode decision algorithm can significantly reduce the depth map encoding time of 3D-HEVC while maintaining almost the same RD performance as the original encoder.

The rest of this paper is organized as follows. The proposed fast depth map mode decision algorithm is detail in Section 2. Simulation results and conclusions are given in Sections 3 and 4, respectively.

2. Proposed fast depth map mode decision algorithm

2.1. Adaptive depth map coding levels determination based on the depth–texture correlation

3D-HEVC inherits an advanced quadtree-based coding approach from HEVC, wherein a picture is divided into coding tree unit (CTU) [35]. The CTU can then be split into four CUs, and the CU is the basic unit of region splitting used for inter/intra prediction, which allows recursive subdividing into four equally sized blocks. A specified maximum coding level is set to limit the CU split recursion. In the joint model of 3D-HEVC, a complex RD optimization process is performed all the possible coding level to find one with the minimum RD cost and determine the best coding mode for a CU at that level. This technique achieves the highest possible depth map coding efficiency, but requires a very high encoding time. Since the depth map and texture video represent the same scene at the same time instant (they have the same motion characteristics), there is a high correlation among the coding levels from depth map and texture video. Based on this concept, if we can exploit depth map and texture video correlations to determine depth map CU coding level, the time-consuming process of computing RD costs on unnecessary coding level can be skipped in 3D-HEVC depth map coding.

Depth map is used to represent the same scene at the view point and same time instant captured by the corresponding texture video cameras. Therefore, CU coding level of depth map and texture video are closely linked. The optimal depth map coding level of a certain CU is the similar or very close to the coding level of its corresponding texture video CUs due to the two CUs have the same motion characteristics. The coding levels of the corresponding texture video CU affects the coding level determination process of the current depth map CU. Thus we can make use of the texture video coding information to analyze current depth map CU properties and early skip unnecessary depth map coding level.

On the basis of these observations, we propose to analyze the depth map CU coding level using the coding information from the co-located texture video CU. The co-located texture video CU are described as in Fig. 1. D_c denotes the current depth map CU, C_c denotes the co-located CU in the texture video and C_l, C_u, C_{ul} and C_{ur} its left, up, upleft and upright CU in corresponding texture video as in Fig. 1.

According to the coding information correlation with the mode maps of encoded texture video frames, we define a set of mode predictors (Ω) for D_c as follows,

$$\Omega = \{C_c, C_l, C_{ul}, C_u, C_{ur}\} \tag{1}$$