This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2617332, IEEE Transactions on Circuits and Systems for Video Technology

Fast Mode Decision Based on Grayscale Similarity and Inter-View Correlation for Depth Map Coding in 3D-HEVC

Jianjun Lei, *Member, IEEE*, Jinhui Duan, Feng Wu, *Fellow, IEEE*, Nam Ling, *Fellow, IEEE*, Chunping Hou

***Abstract*—The 3D extension of HEVC (3D-HEVC) significantly improves the coding efficiency of 3D video at the expense of computational complexity. This paper presents a novel fast mode decision algorithm for depth map coding based on the grayscale similarity and inter-view correlation. First, depth map grayscale similarity is adopted to judge whether the reference frame could assist the coding of current frame. When the difference of average grayscale between the co-located CU and the current CU is smaller than the similarity threshold, the depth level of current CU will be restricted by that of the coded reference CU. Second, the grayscale similarity and inter-view correlation are jointly used for dependent views to achieve early decision on the best prediction unit (PU) mode. The mode decision procedure will be early determined when the co-located CU, which has a grayscale similarity with the current CU, selects Merge or Inter 2N×2N as the best prediction mode. Moreover, when the corresponding CU in the independent view selects Merge or Inter 2N×2N as the best prediction mode, the current CU will skip other PU modes checking based on the strong inter-view correlation. Finally, different strategies are proposed for P-frames and B-frames of dependent views in view of the characteristics of different prediction structure. For B frames, the PU mode information of the coded independent view is utilized as reference to skip the unnecessary mode decision processes. For P frames, the spatial-temporal correlation is considered in the process of early mode decision to determine whether choosing the Merge mode or Inter 2N×2N as the best mode. Experimental results show that our proposed scheme achieves considerable time saving with negligible degradation of coding performance.**

***Index Terms*—Video coding, visual communication, depth map coding, 3D-HEVC, grayscale similarity, inter-view correlation, early termination**

Manuscript received February 9, 2016; revised August 2, 2016 and September 17, 2016; accepted October 10, 2016. Date of publication October, 2016. This research was supported in part by the Natural Science Foundation of China (No. 61271324, 61520106002, 61471262, 91320201).

J. Lei, J. Duan, and C. Hou are with the School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: jjlei@tju.edu.cn).

F. Wu is with the School of Information Science and Technology, University of Science and Technology of China, Hefei, 230026, China (e-mail: fengwu@ustc.edu.cn).

N. Ling is with the Department of Computer Engineering, Santa Clara University, Santa Clara, CA 95053 USA (e-mail: nling@scu.edu).

I. Introduction

THREE-DIMENSIONAL (3D) video systems are increasingly popular due to their real world visual experience with depth perception. Meanwhile, advances in 3D content acquisition and display technologies have pushed the development of 3D video [1], [2]. Multiview-video-plus-depth (MVD), which is formed by multiview texture video and its corresponding depth maps representing the geometric information of scene, has been a fundamental and efficient 3D data format [3], [4,] [5]. It allows for synthesis of much more virtual views by exploiting Depth-Image-Based-Rendering (DIBR) [6], [7]. However, MVD data is more complicated to process due to the introduction of the depth map. Therefore, high efficiency is important for the practical applications of MVD systems.

ITU-T and MPEG set up the Joint Collaborative Team on Video Coding (JCT-VC), and developed the new generation high efficiency video coding standard (HEVC) [8]. HEVC achieves significant compression efficiency and promotes the development of video coding technology [9]. For the purpose of developing an advanced 3D video coding standard, the Joint Collaborative Team on 3D Video Coding Extension Development Working Group (JCT-3V) was created in 2012 by ITU-T and MPEG [10]. After the standardization of 3D-AVC, JCT-3V focused on developing a 3D extension of the HEVC video coding standard and drafted the test model of 3D-HEVC [11], [12]. 3D-HEVC introduces new prediction techniques and coding tools to improve the efficiency of MVD data coding [13].

3D-HEVC inherits the key coding technologies of HEVC, especially for independent view which entirely adopts the HEVC compliant codec. In addition to motion-compensated prediction (MCP), 3D-HEVC also introduces disparity-compensated prediction (DCP) for dependent views referring to previously decoded frames of the basic view [14], [15]. For depth data, 3D-HEVC incorporates inter-component prediction technology and additional coding tools. A depth map is a grayscale Luma-only image in which each sample indicates the relative distance from the object to the camera plane in the 3D space. The geometry information given by the depth data is exploited in the rendering process using DIBR [16]. Depth maps have quite different characteristics than the texture videos, which consist of large quantity of smooth regions separated by

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2617332, IEEE Transactions on Circuits and Systems for Video Technology

sharp edges. Consequently, the efficiency of depth coding can be improved by regarding its properties. 3D-HEVC has included several additional coding tools for depth map coding, such as the Depth Modeling Modes (DMMs), motion parameter inheritance (MPI), and View Synthesis Optimization (VSO). Conventional intra prediction and transform coding in HEVC produce significant artefacts at sharp edges in depth maps and affect the quality of synthesized intermediate views. DMM partitions a depth block into two non-rectangular regions using two different partition types, namely Wedgelets and Contours [17]. Each partitioned region is represented by a constant partition value (CPV). DMM achieves a better representation of edges and more accurate predictions, which is integrated as an alternative to the intra modes. Furthermore, MPI as an inter-component tool extracts the motion information from the coded texture video for depth map coding [18]. 3D-HEVC has achieved great coding efficiency for dependent views and depth data by introducing inter-view and inter-component coding tools. Meanwhile, the computational complexity of 3D-HEVC has increased exponentially due to the introduction of the wedgelet pattern decision in DMM, complex Rate Distortion Optimization (RDO) process, and VSO process for depth video coding [19], [20]. It is obviously necessary to research fast coding algorithms to reduce computational complexity of 3D-HEVC and simultaneously ensure the coding performance.

Currently, many fast coding algorithms have been proposed for 3D-HEVC. Chi *et al*. presented a fast CU size decision algorithm which jointly takes advantages of the inter-view correlation and quad-tree structure constraints for 3D-HEVC [21]. In [22], Mora *et al*. proposed the inter-component coding tools, the texture quad-tree initialization (QTI), and the depth quad-tree limitation (QTL) to simultaneously achieve time saving and coding gains. Tohidypour *et al*. proposed a fast method that adaptively adjusted the motion search range and introduced early termination for the inter/intra prediction mode search [23]. This method mainly exploited the inter-view correlation of motion homogeneity, prediction modes, RD cost and the disparity between different views. In [24], [25], Tohidypour *et al*. reduced the computational complexity of 3D-HEVC by utilizing inter-view information of encoded neighboring blocks to predict the mode of blocks in dependent views. Shen *et al*. proposed a fast and efficient mode decision algorithm, which exploits not only the correlation among spatial and temporal neighboring CUs but also the relevance of inter-views and inter-levels [26]. In [27], Zhang *et al*. proposed a fast decision algorithm for dependent texture views to reduce the encoding time of 3D-HEVC, which utilized the inter-view correlation to early decide Merge mode and terminate CU splitting. In [28], [29], [30], Zhang *et al*. proposed several fast algorithms to reduce the computational complexity of the procedures of motion estimation and disparity estimation.

Some fast algorithms have also been presented for depth coding in 3D-HEVC. Zhang *et al*. utilized a classification method to partition the depth block to replace the DMM mode when coding depth map [31]. Sanchez *et al*. proposed to reduce the complexity of depth intra prediction of the 3D-HEVC by exploiting a Simplified Edge Detector (SED) algorithm to exclude the unnecessary DMM evaluations [32]. Gu *et al*. selectively skipped intra Segment-wise DC Coding (SDC) mode in full RD cost calculation to accelerate the depth intra coding [33]. Merkle *et al*. proposed a simplification of the Wedgelet search for DMM modes 1 and 3 to reduce the encoder/decoder complexity [34]. A fast DMM mode selection algorithm for depth intra coding was proposed by Gu *et al*. in [35], which utilized the selection results of MPMs to early terminate the DMM full-RD cost calculation to achieve time saving. In [36], Gu *et al*. proposed a fast selection algorithm to skip unnecessary Bi-Partition modes by using the RD cost value calculated in HEVC intra Rough Mode Decision (RMD) as the mode selection threshold. Zhang *et al*. presented a low-complexity depth map compression method for HEVC-based 3D video coding, which includes early mode decision termination, adaptive search range motion estimation, and fast disparity estimation [37]. In [38], Li *et al*. presented a fast mode decision algorithm for 3D-HEVC based on the depth information to limit the required coding level to accelerate the encoding process.

However, the existing fast algorithms for depth map coding have not adequately exploited the spatial, temporal, and inter-view correlation. The characteristics of the depth map are rarely mined and used as well. In this paper, to reduce the complexity of the 3D-HEVC, we propose a novel fast mode decision method based on grayscale similarity and inter-view correlation for depth map coding, which sufficiently utilizes the depth map's grayscale similarity among the inter-frames and the correlation between independent view and dependent views. The main contributions of this paper can be summarized as follows: 1) we take into account depth characteristics different from texture and mine the grayscale similarity of the depth map between the temporal reference frame and current frame, which is as a criterion to judge whether or not to early terminate current CU splitting; 2) for dependent views, the grayscale similarity and inter-view correlation are combined to early determine the best PU mode; 3) based on the distinct coding characteristics of the P-frames and B-frames, different mode decision methods are adopted to achieve respectable time saving and simultaneously assure the coding performance. Experimental results demonstrate the effectiveness of our proposed approaches.

The rest of the paper is organized as follows. Section II introduces the coding structure and mode decision process for 3D-HEVC. Section III analyzes the temporal correlation in the PU mode and CU splitting depth level based on the grayscale similarity and the inter-view correlation for dependent views in the depth map coding. Based on the analysis in Section III, a fast mode decision method is proposed for depth map coding in Section IV. Experimental results and conclusions are given in Sections V and VI, respectively.

II. Background

A. MVD Prediction structure

The prediction structure of three view texture videos and corresponding depth maps in 3D-HEVC are depicted in Fig. 1.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2617332, IEEE Transactions on Circuits and Systems for Video Technology

As a base view, namely independent view, View 0 is coded first and uses the hierarchical B picture (HBP). The other dependent views, View 1 and View 2, are subsequently coded. All intra-coded key pictures (I frames) are replaced by inter-coded pictures (P frames) using the inter-view prediction in dependent views. The remaining pictures with hierarchical B pictures use inter-view prediction and temporal prediction [39]. For the random access configuration, key frames will be inserted at each intra period, which is set as 24 in Fig. 1. T_i represents the texture video, D_i is the depth map coded after T_i, and i denotes the ith view.

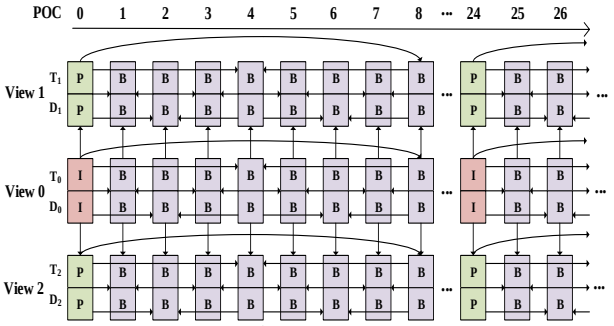


Fig. 1. The prediction structure of three view texture videos and depth maps in 3D-HEVC.

B. CU splitting and Mode decision

The same quad-tree coding structure is inherited from HEVC for both texture videos and depth maps in 3D-HEVC. Each frame is divided into slices, and then a slice is partitioned into a series of coding tree units (CTUs). As the basic coding unit, CTU can be split into four CUs. In addition, a CU can also be recursively split into four sub-CUs coming to the next depth level. The prediction information for the CU is carried by the prediction unit (PU), whose root node is the CU [40]. There are two PU types for intra prediction and eight PU types for inter prediction in the PU mode decision process. The encoders perform eleven PU modes for inter-frames, consisting of Merge mode (including Skip mode, which is a special Merge mode without residuals), eight Inter modes (2N×2N, N×N, N×2N, 2N×N, 2N×nU, 2N×nD, nL×2N, and nR×2N), and two Intra modes (2N×2N, N×N), as shown in Fig. 2.

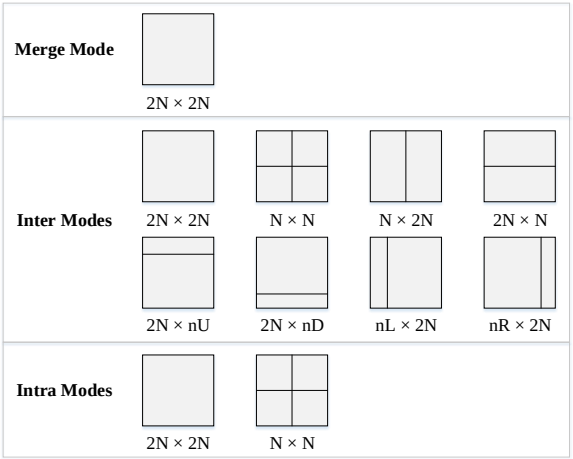


Fig. 2. PU modes in 3D-HEVC.

Considering the characteristics of MVD videos, 3D-HEVC modifies HEVC codecs for coding the dependent views and the depth data, which exploits inter-component (texture-depth) and inter-view redundancies in 3D videos. In the current 3D-HEVC design, new depth intra modes have been introduced for depth map coding. DMM, as a new intra mode, is adopted for the sharp edges mainly characterizing the depth map. When deciding the best intra mode for the depth map, DMM is added to the full RD-list along with the conventional intra prediction modes. Finally, the mode with the lowest RD cost in the list can be selected as the best 3D-HEVC intra mode. Although DMM improves the coding performance of the depth map, it brings unaffordable complexity to compressing the depth map.

In the reference software implementation of 3D-HEVC (HTM), the complex Rate Distortion Optimization (RDO) process is performed for each depth level in order to determine the best CU splitting depth level and the best prediction mode.

$$\begin{cases} P^* = \arg\min P_m, \\ P_m = D(m) + \lambda \cdot B(m). \end{cases} \quad m \in C \tag{1}$$

where C denotes the candidate mode set; P_m is the RD cost function; $D(m)$ is the distortion caused by coding the original CU with the candidate mode m; λ is the Lagrangian multiplier; $B(m)$ represents the number of bits used for encoding the current CU with the candidate mode m. Finally, the coding mode with the least RD cost is selected as the best one in the RDO process.

III. Motivation and Statistical Analyses

As described above, the depth level and prediction mode decision have inevitably increased the computational complexity of HTM encoder. Therefore, the key issue in this paper is how to accurately early terminate CU splitting and skip unnecessary PU candidate modes to predigest the exhausted original method.

Consequently, some experiments are conducted to analyze the coding process. The test conditions are presented in Table I. The main configuration is set as Common Test Condition (CTC) defined by JCT-3V for HTM experiments [41].

Table I. Test Conditions.

Profile	Main
Group size	8
QP pairs	(25,34), (30,39), (35,42), (40,45)
CTU size	64 × 64
Intra Period	24
Coding Structure	Hierarchical B frame structure
Fast Options-QTL	Enabled
Software Version	HTM 13.0

A. Distributions of depth levels and PU modes

Generally, all possible CU splitting depth levels ranging from 0 to 3 are recursively examined in order to decide the best depth level of a CTU in a depth map. However, only a few CUs need to be split into the maximum depth level. The distribution of splitting depth levels of depth map CUs are shown in Table II. It can be observed from the table that the average percentage of CUs selecting the first two depth levels as the best depth level is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2617332, IEEE Transactions on Circuits and Systems for Video Technology

up to 99.04%, while the total average percentage of CUs selecting the last two depth levels is about 0.97%, and the CUs selecting the last depth level only accounts for 0.18%. Obviously, if we early decide the max depth level of current CU and skip the unnecessary larger depth levels, the majority of the coding time will be saved.

Table II. Depth Levels Distribution of CU Splitting in Depth Maps.

Sequences	Depth 0	Depth 1	Depth 2	Depth 3
Lovebird1	96.89%	2.30%	0.64%	0.18%
BookArrival	90.89%	7.33%	1.53%	0.26%
Doorflowers	96.11%	3.00%	0.74%	0.15%
Poznan_CarPark	97.55%	1.75%	0.54%	0.18%
Poznan_Hall1	96.36%	3.03%	0.50%	0.11%
Average	95.56%	3.48%	0.79%	0.18%

In addition, the mode decision performs a complex RDO process by traversing all inter and intra modes to select the best mode with the least RD cost for each depth level, which is a time-consuming computation. It is necessary to analyze the characteristics of mode distribution to skip unnecessary modes. Merge mode is added into the HEVC as a new prediction mode, which derives the motion model parameters from spatial or temporal neighbors. In other words, the CUs that choose the Merge mode as the best mode share motion information with their neighboring CUs. These CUs generally have similar characteristics and belong to the same region or have a similar depth value [8], [42]. Thus, CUs located in the consecutive background and static regions are suitable for encoding in the Merge mode [43]. Table III shows the distribution probability of three category modes for depth map coding, which are Merge, Inter 2N×2N, and other modes. It can be observed from Table III that the Merge mode accounts for 96.07% on average for depth map coding. In HTM, the Merge mode is checked ahead of other modes in the mode decision. Nevertheless, there is no need to traverse extra modes (remaining inter and intra modes) in most circumstances. If the Merge mode can be early determined, the mode decision process can reduce the computational complexity to a great extent. Thus, we can figure out reasonable conditions to skip unnecessary depth level checking and early terminate the mode decision to save encoding time.

Table III. PU Mode Distribution of Depth Maps.

Sequences	Merge	Inter2Nx2N	Other modes
Lovebird1	98.17%	0.13%	1.72%
BookArrival	94.06%	0.25%	5.69%
Doorflowers	97.73%	0.07%	2.20%
Poznan_CarPark	97.09%	0.17%	2.75%
Poznan_Hall1	93.33%	0.20%	6.47%
Average	96.07%	0.16%	3.77%

B. Grayscale similarity of depth map between inter-frames

Depth maps have quite different characteristics from texture videos. The depth map represents the distance between an object and the camera. Most regions within an object have similar depth-level distributions. Thus, the homogeneous regions belonging to backgrounds or the same object mostly select small CU depth levels and simple prediction modes, while the boundary regions with sharp edges choose finer split types and small PUs [44]. Generally, regions within an object have similar distance information reflecting on the grayscale in the depth map. Consequently, it is feasible to exploit the average grayscale to determinate whether the regions have a similar property, such as homogeneous regions or backgrounds. The two regions with approximate average grayscales probably have similar depth levels and prediction modes.

Based on the above analysis and taking account into the strong correlation between the current CU and the co-located CUs in the temporal reference frame [45], [46], the grayscale similarity between the current CU and the co-located CU in the reference frame is investigated. We define several parameters to explicitly represent the investigation process. The flag of the grayscale similarity, S_G, is defined as:

$$S_G = \begin{cases} true, & if \left| G_{cur} - G_{ref} \right| < TH \\ false, & otherwise. \end{cases} \quad (2)$$

where G_{cur} is the average grayscale of the current CU, and G_{ref} is the average grayscale of the co-located CU in the reference frame. The current CU and the co-located CU in reference frame are depicted in Fig. 3. *TH* represents the threshold of the grayscale similarity which is set to 1 through extensive tests by considering the trade-off between encoding time and coding performance.

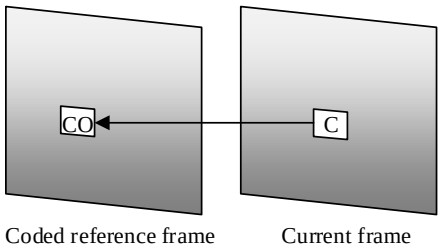


Fig. 3. The current CU and the co-located CU in the reference frame. C represents the current CU; CO is the co-located CU in coded reference frame.

Two percentages, P_m and P_s are defined as:

$$P_m = \frac{N(D_{cur} <= D_{ref} \;\&\&\; S_G = true)}{N(S_G = true)} \quad (3)$$

$$P_s = \frac{N(S_G = true)}{N(all\ CUs)} \quad (4)$$

where D_{cur} represents the max coding depth level of the current CU and D_{ref} is the max coding depth level of the co-located CU in the reference frame. $N(\cdot)$ represents the number of CUs that meet the conditions. P_m represents the percentage of the CUs whose max coding depth level is less than or equal to that of its reference CUs in the CUs where S_G is true. P_s represents the percentage of the CUs where S_G is true in the all encoding CUs. Experimental results are shown in Table IV. It can be seen that P_m accounts for 99.88% on average. P_s appears from 32.41% to 94.15%, 74.94% on average.

In addition to the depth level, the prediction mode distribution of the current CU is also tested. First, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2617332, IEEE Transactions on Circuits and Systems for Video Technology

condition R_1 is defined as follows.

$$R_1 = \{ \left| G_{cur} - G_{ref} \right| < TH \ \&\& \ M_r = (Merge \,||\, Inter\ 2N \times 2N) \} \quad (5)$$

where M_r represents the best mode of its co-located CU in the reference frame. It can be observed from Table V that a majority of the CUs select the Merge mode as the best prediction mode when R_1 is satisfied.

Table IV. Experimental Results of P_m and P_s.

Sequences	P_m	P_s
Lovebird1	99.92%	94.15%
BookArrival	99.85%	69.24%
Doorflowers	99.91%	87.46%
Poznan_CarPark	99.91%	91.45%
Poznan_Hall1	99.83%	32.41%
Average	99.88%	74.94%

Table V. Mode Distribution of Depth Maps when R_1 is Satisfied.

Sequences	Merge	Inter2Nx2N	Other modes
Lovebird1	99.41%	0.22%	0.37%
BookArrival	98.02%	0.73%	1.25%
Doorflowers	99.28%	0.27%	0.45%
Poznan_CarPark	99.17%	0.25%	0.59%
Poznan_Hall1	97.20%	0.67%	2.13%
Average	98.62%	0.43%	0.96%

Based on the above analysis, it can be obviously concluded that the correlation of CU depth level and PU mode between inter-frames is strong when S_G is true. If the maximum depth level of CU splitting and PU modes could be early determined, the encoding time of the depth map would be greatly decreased. Accordingly, an efficient fast coding scheme can be developed by utilizing the grayscale similarity of the depth map between inter-frames.

C. The correlation of inter-view for dependent views

Since multiview videos represent the same scene at the same time with only differences in camera angle, there is a lot of redundant information between multiple views. To take full advantage of the inter-view correlation, the statistic of inter-view correlation of the prediction mode is also performed. Different from the statistic in Table III which shows the distribution probability for both independent views and dependent views, this part lists the mode distribution only for dependent views. Table VI gives the prediction mode distribution of current CU in dependent depth views when the best mode of its corresponding CU (located by the disparity vector) in the independent depth view M_b is the Merge mode or Inter 2N×2N. In the table, the condition R_2 is expressed as

$$R_2 = \{ M_b = (Merge \,||\, Inter\ 2N \times 2N) \} \quad (6)$$

It can be observed from the table that 94.41% of CUs choose the Merge mode as the best prediction mode and 1.63% of CUs choose Inter 2N×2N when the condition R_2 is satisfied.

Based on the above analysis, it can be inferred that the mode correlation between views is tight, so the mode information of the coded independent views can be used as a reference to guide the coding of the current view. Thus, the dependent view coding can exploit the coded mode information of the independent depth view to reduce computational complexity.

Table VI. Mode Distribution of Depth Maps for Dependent Views when R_2 is Satisfied.

Sequences	Merge	Inter2Nx2N	Other modes
Lovebird1	95.41%	1.63%	2.96%
BookArrival	90.71%	2.53%	6.76%
Doorflowers	97.08%	0.96%	1.96%
Poznan_CarPark	97.16%	0.78%	2.06%
Poznan_Hall1	91.67%	2.22%	6.11%
Average	94.41%	1.63%	3.97%

IV. Proposed Fast Mode Decision Algorithm

As observed in Section III, there are grayscale similarity between inter-frames for depth map coding and the inter-view correlation for dependent views. In this section, we present a fast decision algorithm for depth map coding, not only utilizing the temporal relevance about CU splitting and prediction mode checking, but also the mode information of the independent depth view for the remaining views. The details of the proposed algorithm are presented as follows.

A. Early CU depth determination based on grayscale similarity

It can be observed from Table IV that when the grayscale similarity S_G is true, there are 99.88% CUs on average whose max coding depth level is less than or equal to the co-located CUs in the coded reference frames. Based on this observation, an early CU depth level termination scheme is proposed in order to simplify the process of quad-tree recursive traversal.

First, G_{cur} and G_{ref} are calculated. Then, the grayscale similarity S_G needs to be judged by the similarity threshold TH. When the difference of G_{cur} and G_{ref} is smaller than TH, the two CUs have a similar average grayscale and S_G is true; otherwise, S_G is false. In other words, if the content of the two CUs have too many differences, then S_G would be false and the current CU would not be accurately predicted when it utilizes the coded information of the co-located CU to simplify the coding process.

The depth level information D_{ref} is extracted as the reference for the current CU. The maximum depth level of the current CU, $D_{\max}$, is determined according to the value of S_G and can be expressed as:

$$D_{\max} = \begin{cases} D_{ref}, & if\ S_G = true \\ D_{org}, & otherwise. \end{cases} \quad (7)$$

where D_{org} is the maximum depth level of original HTM algorithm, and the value is usually set as 3. When $D_{\max}$ is determined, it can be used for limiting the splitting of the current CU to avoid unnecessary RD cost checks for the larger depth levels.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2617332, IEEE Transactions on Circuits and Systems for Video Technology

Generally, CUs covered by simple motion-content or background areas choose a small depth level. In contrast, most of the CUs that have sharp edges and complex motion-content tend to split to larger depth levels. To further simplify the time-consuming recursive traversal process, a minimum depth level limitation condition is added to the early CU depth determination scheme based on grayscale similarity. It utilizes the coded information of spatial adjacent CUs and temporal co-located CU in the forward reference frame. The condition of minimum depth level can be described in detail as follow.

The predicted depth level of the current CU, D_{pre} , is defined as:

$$D_{pre} = \alpha \cdot (D_{left} + D_{above}) + \beta \cdot D_{co} \tag{8}$$

where D_{left} represents the maximum depth level of the adjacent left CU of the current CU, and D_{above} represents the maximum depth level of the adjacent above CU of the current CU. D_{co} is the maximum depth level of the co-located CU in the reference frame of the current CU. α and β are the weight factors. Since this limitation is used under the condition that the grayscale similarity S_G is true, the temporal co-located CU in the reference frame have a larger influence with current CU. Thus, α and β are set to 0.25 and 0.5, respectively.

Then, the minimum depth level of the current CU, $D_{\min}$, is determined based on the value of D_{pre} and S_G, which can be formulated as follows.

$$D_{\min} = \begin{cases} 1, & if\ S_G = true\ \&\&\ 1 < D_{pre} \le 2, \\ 2, & if\ S_G = true\ \&\&\ 2 < D_{pre} \le 3, \\ 0, & otherwise. \end{cases} \tag{9}$$

When S_G is true, $D_{\min}$ has a different value with a different D_{pre} .When S_G is false, $D_{\min}$ would not be used as the depth level limitation for the current CU and the minimum depth level of the current CU would be the original value 0 in HTM. When the reference CU of the current CU is split to a large depth level, the current CU quite possibly splits to a large depth level. The RDO in small depth levels is obviously unnecessary. Based on the $D_{\min}$ calculated using equation (9), a minimum depth level is set for the current CU, and the RDO process for the unnecessary small depth levels is skipped to avoid redundant computation.

B. Early PU mode decision based on grayscale similarity

It can be seen from Table V that 99.05% on average CUs select the Merge and Inter 2N × 2N modes as the best prediction mode when S_G is true. When inter prediction performs, Merge and Inter 2N × 2N modes are checked first. If the two modes can be early determined, it is unnecessary to check other modes. In addition to early CU depth level termination, an early PU mode decision is presented as well, which is based on the results of the mode distribution of the depth maps. The method only checks the Merge mode and skips other modes when $\left|G_{cur} - G_{ref}\right| < TH$ and the best mode of the reference CU is Merge mode. If the best mode of the reference CU is Inter 2N×2N mode, the method only check Merge and Inter 2N×2N modes. The modes need to be checked in the RDO process, M_c, can be defined according to the mode context of reference CUs as:

$$M_c = \begin{cases} Merge, & if\ S_G = true\ \&\&\ M_r = Merge, \\ Merge\ and\ Inter\ 2N \times 2N, & if\ S_G = true\ \&\&\ M_r = Inter\ 2N \times 2N, \\ All\ modes, & otherwise. \end{cases} \tag{10}$$

where *All modes* means the modes in the RDO process in the original HTM algorithm.

C. Early PU mode decision based on inter-view information for dependent views

When coding dependent depth views, the independent view has been coded. Therefore, a fast coding method for dependent depth views by utilizing coding information of the coded independent depth view is proposed. Based on the observation in Section III-C, it is reasonable to only check Merge and Inter 2N×2N modes and skip other modes when R_2 is satisfied. When M_b is Merge mode, the method only checks the Merge mode and skips other modes. Similarly, when M_b is the Inter 2N×2N mode, the method checks Merge and Inter 2N×2N modes and then skips other modes.

In summary, the mode that needs to be checked for the current CU in B-frames, M_c, can be determined as:

$$M_c = \begin{cases} Merge, & if\ M_b = Merge, \\ Merge\ and\ Inter\ 2N \times 2N, & if\ M_b = Inter\ 2N \times 2N, \\ All\ modes, & otherwise. \end{cases} \tag{11}$$

Based on the prediction structure of MVD mentioned in Section II, we know that there are P-frames and B-frames when coding dependent depth views. Generally, P-frames are encoded finer than B-frames since P-frames are the reference frame of the B-frames. A stricter scheme should be presented for P-frames to ensure a good coding performance.

As the reference frames of P-frames, I-frames perform intra coding in the base view. Due to no inter mode reference information from the I-frames, the early termination strategy in (11) for Merge and Inter 2N×2N modes is disabled for P frames. Thus, the coded neighbor P frame is used as a reference for the current P frame. In addition to the co-located CU in the coded neighbor P frame, the coded spatial neighbor CUs including the upper CU, left CU, left upper CU, right upper CU of the current CU in P-frames, are taken consideration. Furthermore, the mode information of the parent CU is also used as a reference since the parent CU has similar characteristics with its children CU according to the recursive quad-tree partition in HEVC. All the reference CUs are shown in Fig. 4. The best mode of the current CU can be early determined based on the condition of the neighboring CUs as follows.

$$R_3 = \begin{cases} true, & if\ \sum_{i=1}^{N} \theta_i = 0, \\ false, & otherwise. \end{cases} \tag{12}$$

where *N* is the number of reference CUs, which is equal to 6. θ_i represents the Merge flag of the reference CU. If the best mode of the reference CU is the Merge mode, the flag is equal to 0; otherwise, it is equal to 1.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2617332, IEEE Transactions on Circuits and Systems for Video Technology

To further improve the accuracy of the fast mode decision for P-frames, the Code Block Flag (CBF) is combined with the spatial-temporal correlation as the criterion of fast mode decision for dependent view coding. Once the CBF of the current CU equals to zero, it represents that the current CU would be well predicted by the current mode [42], [47].

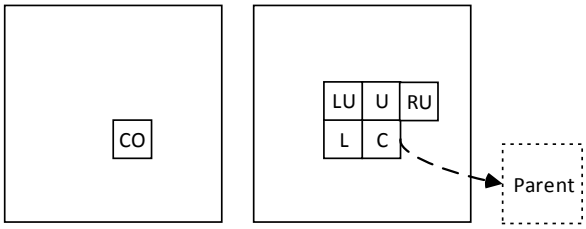


Fig. 4. The reference CUs and current CU. C represents the current CU; CO is the co-located CU in neighbor P frame; L is the left CU of the current CU; U is the upper CU; LU is the left upper CU; RU is the right upper CU; Parent represents the parent CU of the current CU in the HEVC quad-tree structure.

As a result, the fast PU mode decision for P-frames can be presented as follows. When all reference CUs in Fig. 4 select Merge mode as the best mode and meet the condition that CBF is equal to zero, only the Merge mode needs to be checked and the other modes can be skipped in the RDO process. Therefore, the mode needing to be checked for the current depth CU in P-frames, M_c, can be expressed as:

$$M_c = \begin{cases} Merge, \ if \ R_3 = true \,\&\&\, CBF = 0, \\ All\ modes, \qquad\qquad otherwise. \end{cases} \tag{13}$$

D. The overall algorithm

According to the above analysis, the proposed fast algorithm for depth map coding is described in Algorithm 1.

Algorithm 1 Proposed Fast Mode Decision Algorithm

Start: Calculate G_{cur} and G_{ref}
Set $D_{\max}$ and $D_{\min}$ value based on S_G
for Depth = $D_{\min}$ to $D_{\max}$ **do**
if current view id = 0 **then**
  Test Merge mode
  if ($M_r = Merge$ && $|G_{cur} - G_{ref}| < TH$) **then**
    Go to step 28
  else
    Test Inter 2N×2N
    if ($M_r = Inter\ 2N \times 2N$ && $|G_{cur} - G_{ref}| < TH$) **then**
      Go to step 28
    else
      Go to step 27
else
  Test Merge mode
  if current frame = B frame **then**
    if ($M_r = Merge$ && $|G_{cur} - G_{ref}| < TH$) or $M_b = Merge$ **then**
      Go to step 28
    else
      Test Inter 2N×2N
      if ($M_r = Inter\ 2N \times 2N$ && $|G_{cur} - G_{ref}| < TH$) or $M_b = Inter\ 2N \times 2N$ **then**
        Go to step 28
      else
        Go to step 27
  else if (R_3 = true && CBF= 0) **then**
    Go to step 28
  else
    Go to step 27
Test all other inter and intra modes
Determine the best mode with minimal RD cost
end for
Compress the next LCU

V. Experimental Results

A. Experimental setup and parameter setting

In order to evaluate the efficiency of the proposed fast algorithm for 3D-HEVC, we implement the algorithm on the 3D-HEVC reference software HTM 13.0, which has already adopted the state-of-the-art methods [21], [33], [34], [35], [48]. The test conditions based on the CTC [41] are listed in Table I. The proposed algorithm is evaluated with six sequences recommended by JCT-3V with two resolutions 1920×1088 (GT_Fly, Poznan_Hall2, Poznan_Street, Undo_Dancer, and Shark) and 1024×768 (Balloons, Kendo, and Newspaper) formats. We adopt three view texture videos and their corresponding depth maps, which includes an independent (center) view and two dependent (side) views. The center, left, and right views (in coding order) are presented as the following: T_0 D_0, T_1 D_1, T_2 D_2 (where T_i and D_i are the corresponding texture and depth frames in the *i*th view, respectively). The details of the test sequences are shown in Table VII. Six synthesized views are rendered after encoding. In detail, three intermediate views are synthesized between each basic view. The hardware platform is Intel(R) Xeon(R) CPU E5-1620 v2 @ 3.70GHz 3.70 GHz, 16.00GB RAM with Microsoft Windows 7 64-bit operating system.

Table VII. The Multiview Test Sequences with Corresponding Input Views.

Resolution	Test Sequence	3-view input
1024×768	Balloons	1-3-5
	Kendo	1-3-5
	Newspaper	2-4-6
1920×1088	GT_Fly	9-5-1
	Poznan_Hall2	7-6-5
	Poznan_Street	5-4-3
	Undo_Dancer	1-5-9
	Shark	1-5-9

TH represents the threshold of the grayscale similarity, which controls the calculation of similarity. The higher the value of TH, the more CUs are calculated as grayscale

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2617332, IEEE Transactions on Circuits and Systems for Video Technology

similarities in the current frame. We have tested several values of TH (which range from 0 to 2 with step of 0.5) using four test sequences to evaluate the reasonability. The relationship between TH and the encoding time saving (TS) of depth maps and the BDBR increase of the coded and synthesized views are shown in Fig. 5. It can be observed form the figure that, when TH becomes larger, the encoding time saving of depth maps increases. Meanwhile, the loss of coding efficiency will also increase. The selected value of TH should keep an acceptable coding performance while reducing the complexity observably. As shown in Fig. 5, when TH <= 1, the encoding time saving of depth maps increases considerably with the increase of TH. When TH is set to be 1, a significant encoding time saving of depth maps can be achieved with a lower increase of BDBR. When TH > 1, the encoding time saving of depth maps increases relatively slowly, however it leads to a continuous increase of BDBR. Based on the above analyses, TH is set as 1 by considering the trade-off between the encoding time and performance.

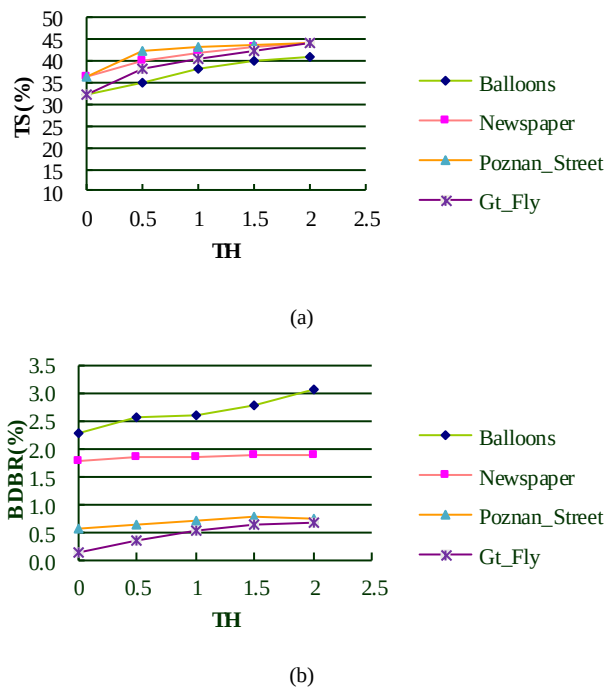


Fig. 5. Relation between TH and the encoding time saving (TS) of depth maps and the BDBR increase of coded and synthesized views for four test sequences. (a) Relation between TH and TS. (b) Relation between TH and the BDBR increase.

B. Experimental results

The coding efficiency of the proposed algorithm is measured by the encoding time saving of depth maps, BDBR and BDPSNR [49]. The experimental results are summarized in Table VIII. In the table, BDBR and BDPSNR columns show the coding results of depth maps. The bitrate and PSNR of the independent view is the depth map coding result of D_0. For dependent views, the bitrate and PSNR considered the average bitrate and PSNR of two dependent views, namely, D_1 and D_2. *TS* denotes the encoding time saving of depth maps compared to the original HTM13.0 encoder, which is computed as

$$TS = \frac{ET_f - ET_o}{ET_o} \times 100\% \quad (14)$$

where ET_f is the encoding time of the depth map encoded by the proposed fast coding method; ET_o represents the encoding time of the depth map encoded by the original HTM13.0 encoder.

It can be observed from Table VIII that the proposed method can effectively achieve time saving for depth map coding and can hold a similar performance for all the test sequences. The proposed method reduces the encoding time from 13.16% to 36.43%, on average 25.56% for independent view of the depth map. Meanwhile, the proposed method presents an average BDPSNR increase of 0.01dB. The above results indicate that the fast method based on the grayscale similarity can avoid unnecessary CU size and PU mode check processes for the independent view. For dependent views, it is obvious that the proposed method achieves more encoding time saving because of the addition of the inter-view correlation. From 42.12% to 55.55%, on average 49.49%, encoding time is saved for dependent views of the depth map with BDBR saving of 0.97% and BDPSNR increase of 0.06dB, on average.

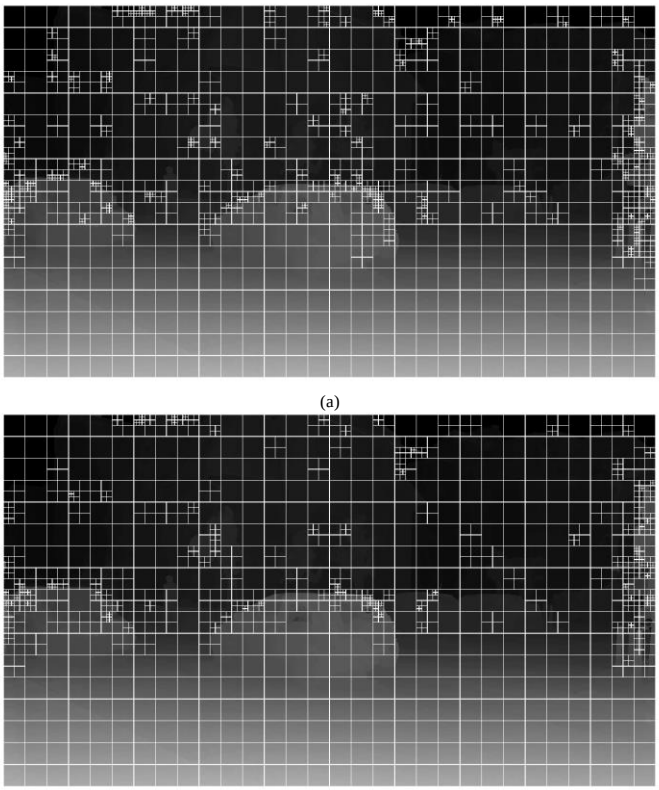
Table VIII. Coding Results for Depth Maps of Independent View and Dependent Views.

Sequences	Independent view			Dependent views		
	TS (%)	BDBR (%)	BDPSNR (dB)	*TS* (%)	BDBR (%)	BDPSNR (dB)
Balloons	-25.25%	0.63	-0.02	-46.98%	3.02	-0.11
Kendo	-13.16%	0.67	-0.02	-44.24%	12.14	-0.38
Newspaper	-32.42%	1.40	-0.07	-48.28%	-1.14	0.04
GT_Fly	-24.21%	-2.64	0.13	-55.55%	-11.70	0.79
Poznan_Hall2	-24.10%	-4.94	0.21	-52.95%	-0.13	0.00
Poznan_Street	-33.24%	0.12	0.00	-51.50%	-4.82	0.14
Undo_Dancer	-36.43%	1.15	-0.07	-54.30%	0.74	-0.06
Shark	-15.63%	3.71	-0.08	-42.12%	-5.85	0.09
1024×768	-23.61%	0.90	-0.04	-46.50%	4.67	-0.15
1920×1088	-26.72%	-0.52	0.04	-51.28%	-4.35	0.19
Average	**-25.56%**	**0.01**	**0.01**	**-49.49%**	**-0.97**	**0.06**

It is worthwhile to mention that BD-rate gains are obtained for some depth sequences. There are maybe two reasons for the BD-rate gains. 1) In 3D-HEVC, the mode decision process for depth maps considers the synthesized view distortion by view synthesis optimization (VSO). The VSO aims to optimize the overall quality on synthesized views, which is probably inconsistent with the quality of the depth map. 2) The depth maps have many regions with static backgrounds or simple shapes, so that lots of CUs are encoded using simple prediction modes, and the complex prediction modes (such as DMM) have been skipped [32]. It can be observed from Table VIII that the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2617332, IEEE Transactions on Circuits and Systems for Video Technology

BD-rate gains are mainly obtained from three sequences (GT_Fly, Poznan_Hall2, and Poznan_Street). Generally, the CUs encoded in our proposed method probably select larger CU size, which may be forced to be split in the original HTM algorithm. Thus, the encoder does not need to send the split flag or partition size for the smaller CUs, and the header information reduction is further achieved. In addition, the quality loss of depth map coding is less than the header information reduction, which results in BD-Rate gains of depth map coding. The depth levels distributions of CU splitting in depth maps coding for Poznan_Street, when respectively encoded in our proposed method and the original HTM, are shown in Fig. 6. It can be observed that some CUs encoded with our proposed method select a larger CU size than the ones encoded with original HTM. Moreover, the number of CUs selecting the last two depth levels is lower when the proposed method is used.



(a)

(b)

Fig.6. Depth levels distributions of CU splitting in depth maps for Poznan_Street. (a) Encoded with the original HTM. (b) Encoded with the proposed method.

In order to further evaluate the performance of our proposed method, two fast mode decision methods [43], [45] are used for an objective comparison of coding performance. The early Merge mode decision algorithm in [43] and the effective CU size decision method in [45] both utilize the spatial-temporal correlation information based on the HEVC quad-tree structure to early determine the prediction mode or CU splitting. The texture and depth videos are all processed in the experiments of the two methods. The proposed method in this paper refers to the CU depth level and mode decision based on the spatial-temporal and inter-view correlations. The comparison results are presented in Table IX. The encoding time and synthesized view performance are used as a criteria to give a rational comparison. The processed data *Enc Time* is calculated as

$$Enc\ Time = \frac{ET_f}{ET_o} \times 100\% \tag{15}$$

The columns INDE and DE represent the *Enc Time* of independent views and dependent views, respectively. The BDBR and BDPSNR in synthesized columns in Table IX are calculated using the average PSNR of synthesized views versus the total bitrate. The bitrate is the total bitrate of three coded texture and depth views. The PSNR is the average PSNR calculated by the six synthesized views. The BDBR and BDPSNR in the coded+synthesized columns are calculated using the average PSNR of the three coded and the six synthesized views versus the total bitrate. The PSNR is the average PSNR calculated by the three coded and the six synthesized views. The PSNR of the synthesized views are measured with regard to the views synthesized by the uncompressed original views.

For independent views, the *Enc Time* of the proposed strategies is 74.45% on average with the maximum of 86.84% and the minimum of 63.57%. For dependent views, the *Enc Time* of the proposed strategies is 50.51% with the maximum of 57.88% and the minimum of 44.45%. It can been observed from Table IX that the encoding time of the proposed strategies is much less than that of the original HTM13.0 encoders. On the other hand, the coding loss of the proposed method is acceptable. As shown in Table IX, the BDPSNR loss of the synthesized views in the synthesized column is 0.07dB on average for all sequences, and the BDBR increase is 2.06%. For coded+synthesized views, the BDPSNR loss is 0.01-0.10 dB, on average 0.04dB, and the BDBR increase is 0.14-2.59%, on average 1.22%. In addition to the HTM encoder, better performance on time saving are achieved compared with Pan's method. The proposed method further reduced 18.24% and 32.28% encoding time for independent views and dependent views, respectively. Meanwhile, the BDBR and BDPSNR performance of synthesized views are almost kept at the same level. The BDBR and BDPSNR performance of coded+synthesized views are better than Pan's method, with less than 0.81% BDBR increase and 0.02 BDPSNR loss. Similarly, the proposed algorithm also achieves 5.57% and 22.69% encoding time saving for independent view and dependent views, respectively, with compared with Shen's method, and have a better RD performance. The above experimental results indicate the proposed scheme is efficient for depth map coding in 3D-HEVC.

In order to give an intuitive presentation, the comparison of time saving between the three methods is depicted in Fig. 7. Obviously, the proposed fast mode decision method based on grayscale similarity and inter-view correlation reduces considerable encoding time as compared with the other two methods. Especially, it can be clearly seen that the proposed method achieves more time saving for dependent views as shown in Fig. 7. Due to the strong inter-view correlation, the coded mode information of the independent view, which is used as reference, accelerates the encoding process of dependent views.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2617332, IEEE Transactions on Circuits and Systems for Video Technology

Table IX. *Enc Time*(%), BDBR(%) and BDPSNR(dB) Performance Comparison of Different Methods.

Sequences	Pan [43] vs HTM13.0				Shen [45] vs HTM13.0				Proposed vs HTM13.0			
	Enc Time		Synth.	Coded + Synth.	*Enc Time*		Synth.	Coded + Synth.	*Enc Time*		Synth.	Coded + Synth.
	INDE	DE	BDBR / BDPSNR	BDBR / BDPSNR	INDE	DE	BDBR / BDPSNR	BDBR / BDPSNR	INDE	DE	BDBR / BDPSNR	BDBR / BDPSNR
Balloons	93.08	85.66	3.07 / -0.12	2.38 / -0.10	85.07	77.03	3.49 / -0.13	2.43 / -0.09	74.75	53.02	4.32 / -0.16	2.59 / -0.10
Kendo	88.38	87.21	2.66 / -0.10	2.53 / -0.10	81.01	74.84	4.25 / -0.17	4.00 / -0.16	86.84	55.76	1.20 / -0.05	0.87 / -0.03
Newspaper	93.16	90.9	2.95 / -0.10	2.40 / -0.08	85.45	76.15	3.34 / -0.11	1.96 / -0.07	67.58	51.72	3.43 / -0.12	1.83 / -0.06
GT_Fly	92.68	78.38	0.79 / -0.02	0.83 / -0.02	80.67	70.96	1.67 / -0.04	1.58 / -0.04	75.79	44.45	0.87 / -0.02	0.53 / -0.01
Poznan_Hall2	98.31	86.03	1.86 / -0.04	1.70 / -0.04	71.54	65.94	1.22 / -0.03	0.89 / -0.02	75.90	47.05	3.31 / -0.08	2.14 / -0.05
Poznan_Street	93.84	77.84	2.11 / -0.05	1.94 / -0.05	81.91	74.97	2.50 / -0.06	2.08 / -0.05	66.76	48.50	1.22 / -0.03	0.69 / -0.02
Undo_Dancer	91.63	76.1	2.65 / -0.08	2.64 / -0.08	79.96	73.95	3.24 / -0.09	2.59 / -0.07	63.57	45.70	1.67 / -0.05	0.97 / -0.03
Shark	90.46	80.17	1.93 / -0.07	1.78 / -0.07	75.98	71.76	3.42 / -0.13	2.97 / -0.11	84.37	57.88	0.48 / -0.02	0.14 / -0.01
1024×768	91.54	87.92	2.89 / -0.11	2.44 / -0.09	83.84	76.01	3.69 / -0.14	2.80 / -0.10	76.39	53.50	2.98 / -0.11	1.77 / -0.07
1920×1088	93.384	79.70	1.87 / -0.05	1.78 / -0.05	78.01	71.52	2.41 / -0.07	2.02 / -0.06	73.28	48.72	1.51 / -0.04	0.90 / -0.02
Average	**92.69**	**82.79**	**2.25 / -0.07**	**2.03 / -0.06**	**80.20**	**73.20**	**2.89 / -0.10**	**2.31 / -0.08**	**74.45**	**50.51**	**2.06 / -0.07**	**1.22 / -0.04**

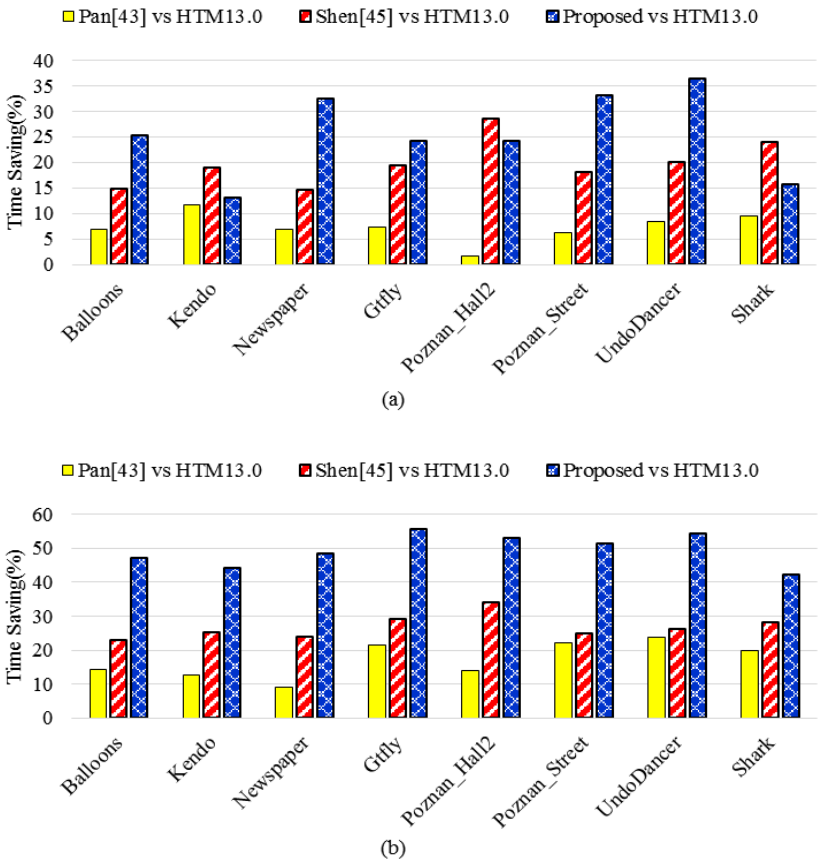


Fig. 7. Comparison of time saving between different methods. (a) Comparison results of all sequences for independent views. (b) Comparison results of all sequences for dependent views.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2617332, IEEE Transactions on Circuits and Systems for Video Technology

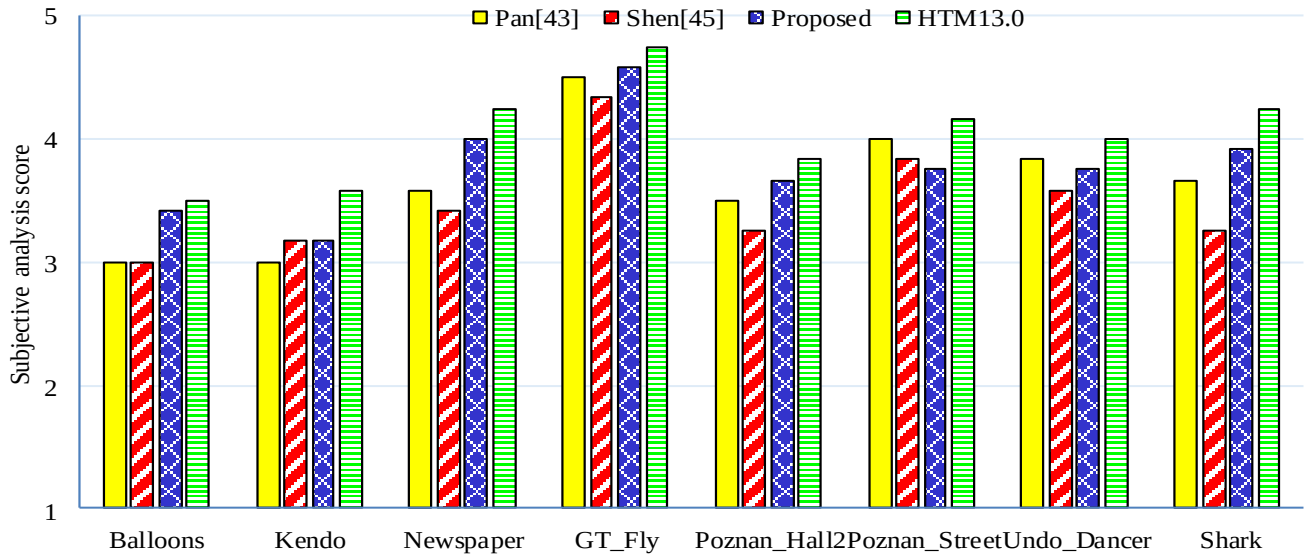


Fig. 8. Comparison of subjective analysis score between different methods.

In order to evaluate the subjective quality of synthesized views, the degradation category rating (DCR) [50], [51] is used. In DCR, the video synthesized by original sequences is presented, then the video with the same view synthesized by sequences encoded with the encoding algorithm is presented, followed by a 2-s interval. After viewing both videos, the observers attribute grades. There are five grades ranging from 1 to 5, where 1 represents the most quality degradation and 5 represents no degradation perceived. 20 participants without a video coding background were invited to participate in the subjective assessment. We first compared the video synthesized by the original sequences against the video with the same view synthesized by sequences encoded with the original HTM13.0 algorithm. Then, we compared the original video against the video encoded with the proposed fast coding scheme. Similarly, the methods in [43] and [45] are also evaluated using subjective analysis. It can be observed from Fig. 8 that the proposed method introduces less subjective quality degradation in synthesized views than that of the methods in [43] and [45].

VI. Conclusion

In this paper, we proposed a fast mode decision method for reducing the computational complexity of 3D-HEVC encoder. The grayscale similarity between the reference frame and current frame is utilized to determine the depth level of CU splitting of the depth maps. Then, the grayscale similarity and the inter-view correlation for the dependent views were combined to skip unnecessary mode checking within the mode decision process. The proposed algorithm can effectively limit the maximum and minimum depth levels of CU splitting and skip unnecessary prediction modes for the current CU. Moreover, in order to guarantee the coding performance of key frames for dependent views, a stricter early termination method was adopted by exploiting the coding information from spatial-temporal neighboring CUs and the parent CUs. In addition, the CBF was also considered to keep the coding accuracy of the proposed algorithm. Experimental results demonstrated that the proposed method can gain a considerable time saving while maintaining an acceptable coding performance.

References

[1] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Multi-view video plus depth representation and coding," in Proc. *IEEE Int. Conf. Image Process*., pp. I201-I204, Otc. 2007.

[2] B. O. Ozkalayc, A. A. Alatan, "3D planar representation of stereo depth images for 3DTV applications," *IEEE Trans. Image Process*., vol. 3, no. 12, pp. 5222-5232, Dec. 2014.

[3] P. Merkle, K. MüLler, and T. Wiegand, "3D video: acquisition, coding, and display," *IEEE Trans. Consumer Electronics*, vol. 56, no. 2, pp. 946-950, May. 2010.

[4] Y. Chen, M. M. Hannuksela, T. Suzuki, S. Hattori. "Overview of the MVC + D 3D video coding standard," *J. Vis. Commun. Image. R.*, vol. 25, no. 4, pp. 679-688, May. 2014.

[5] M. Domanski, O. Stankiewicz, K. Wegner, *et al*. "High efficiency 3D video coding using new tools based on view synthesis," *IEEE Trans. Image Process*., vol. 22, no. 9, pp. 3517-3527, Sep. 2013.

[6] C. Fehn. "A 3D-TV approach using depth-image-based rendering (DIBR)," in *Proc. VIIP*, vol. 3, no. 3, pp. 482-487, 2003.

[7] A. I. Purica, E. G. Mora, B. Pesquet-Popescu, *et al*. "Multiview plus depth video coding with temporal prediction view synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 2, pp. 360-374, Feb. 2016.

[8] G. J. Sullivan, J. Ohm, W. Han, *et al.* "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[9] J. Ohm, G. J. Sullivan, H. Schwarz, *et al*. " Comparison of the coding efficiency of video coding standards – including high efficiency video coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol.22, no. 12, pp. 1669-1684, Dec. 2012.

[10] J. Lei, S. Li, C. Zhu, *et al*. "Depth coding based on depth-texture motion and structure similarities," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 2, pp. 275-286, Feb. 2015.

[11] M. M. Hannuksela, D. Rusanovskyy, W. Su, *et al*. "Multiview-video-plus-depth coding based on the advanced video coding standard," *IEEE Trans. Image Process*., vol. 22, no. 9, pp. 3449-3458, Sep. 2013.

[12] Y. Chen, *et al*. "Test model 9 of 3D-HEVC and MV-HEVC," ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCT3V-I1003, Sapporo, Japan, Jul. 2014.

[13] K. Muller, H. Schwarz, D. Marpe, *et al.* "3D high-efficiency video coding for multi-view video and depth data," *IEEE Trans. Image Process*., vol. 22, no. 9, pp. 3366-3378, Sep. 2013.

[14] C. Bal, T. Q. Nguyen, "Multiview video plus depth coding with depth-based prediction mode," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 6, pp. 995-1005, Jun. 2014.

[15] E. G. Mora, J. Jung, M. Cagnazzo, *et al.* "Modification of the merge candidate list for dependent views in 3D-HEVC," *IEEE International Conference on Image Processing*, pp. 1709-1713, Sep. 2013.

[16] I. Daribo, C. Tillie, and B. Pesquet-Popescu, "Adaptive wavelet coding of the depth map for stereoscopic view synthesis," in Proc. *IEEE 10th Workshop on Multimedia Signal Processing (MMSP)*, pp.413-417, Oct. 2008.

[17] X. Zhao, Y. Chen, L. Zhang, M. Karczewicz, "3D-CE6.h related: Depth Modeling Mode (DMM) 3 simplification for HTM," ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCT3V-A0098, Stockholm, Sweden, Jul. 2012.

[18] Y. Chen, J. Lin, Y. Huang, S. Lei, "3D-CE3.h results on removal of parsing dependency and picture buffers for motion parameter inheritance," ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCT3V-C0137, Geneva, Holland, Jan. 2013.

[19] L. Shen, Z. Zhang, Z. Liu, "Adaptive inter-mode decision for HEVC jointly utilizing Inter-level and spatiotemporal correlations," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 10, pp. 1709-1722, Oct. 2014.

[20] S. Ma, S. Wang, W. Gao, "Low complexity adaptive view synthesis optimization in HEVC based 3D video coding," *IEEE Trans. on Multimedia.*, vol 16, no. 1, pp. 266-271, Jan. 2014.

[21] G. Chi, X. Jin, Q. Dai, "A quad-tree and statistics based fast CU depth decision algorithm for 3D-HEVC," in Proc. *IEEE Int. Conf. on Multimedia and Expo Workshops (ICMEW)*, pp. 1-5, Sep. 2014.

[22] E. G. Mora, J. Jung, M. Cagnazzo, *et al.* "Initialization, limitation, and predictive coding of the depth and texture quadtree in 3D-HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 9, pp. 1554-1565, Sep. 2014.

[23] H. R. Tohidypour, M. T. Pourazad, P. Nasiopoulos, *et al.* "A content adaptive complexity reduction scheme for HEVC-based 3D video coding," in Proc. *18th Int. Conf. on Digital Signal Processing(DSP)*, pp. 1-5, 2013.

[24] H. R. Tohidypour, M. T. Pourazad, P. Nasiopoulos, "A low complexity mode decision approach for HEVC-based 3D video coding using a Bayesian method," *IEEE Int. Conf. on Acoustics, Speech and Signal Processing(ICASSP)*, pp. 895-899, May. 2014.

[25] H. R. Tohidypour, M. T. Pourazad, P. Nasiopoulos, "Online learning-based complexity reduction scheme for 3D-HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, Doi: 10.1109/TCSVT.2015.2477955, pp. 1-14, 2015.

[26] L. Shen, P. An, Z. Zhang, *et al.* "A 3D-HEVC fast mode decision algorithm for real-time applications," *ACM Trans. on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 11, no. 3, Jan. 2015.

[27] N. Zhang, D. Zhao, Y. Chen, *et al.* "Fast encoder decision for texture coding in 3D-HEVC," *Signal Processing: Image Commun.*, vol. 29, no. 9, pp. 951-61, Oct. 2014.

[28] Q. Zhang, Q. Wu, X. Wang, *et al.* "Early SKIP mode decision for three-dimensional high efficiency video coding using spatial and interview correlations," *Journal of Electronic Imaging*, vol. 23, no. 5, Sep. 2014.

[29] Q. Zhang, H. Chang, Q. Wu, *et al.* "Fast motion and disparity estimation for HEVC based 3D video coding," *Multidimensional Systems & Signal Processing*, vol. 27, pp. 743-761, Jul. 2014.

[30] Q. Zhang, N. Li, Y. Gan, "Low complexity mode decision for 3D-HEVC," *Scientific World Journal*, Doi: 10.1155/2014/392505 pp. 1-12, 2014.

[31] M. Zhang, S. Qiu, H. Bai. "Fast Intra prediction based BCIM for depth-map in 3D-HEVC," in Proc. *IEEE Data Compression Conf. (DCC)*, pp. 438-438, Mar. 2014.

[32] G. Sanchez, M. Saldanha, G. Balota, *et al.* "Complexity reduction for 3D-HEVC depth maps intra-frame prediction using simplified edge detector algorithm," in Proc. *IEEE Int. Conf. Image Process.*, pp. 3209-3213, Jan. 2015.

[33] Z. Gu, J. Zheng, N. Ling, *et al.* "Fast Intra SDC coding for 3D-HEVC Intra Coding," ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCT3V- I0123, Sapporo, Japan, Jul. 2014.

[34] P. Merkle, K. Müller, X. Zhao, *et al.* "Simplified wedgelet search for DMM Modes 1 and 3," ITU-T SG16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Doc. JCT3V-B0039, Shanghai, China, Oct. 2012.

[35] Z. Gu, J. Zheng, N. Ling, P. Zhang, "Fast Depth Modeling Mode selection for 3D HEVC depth intra coding," in Proc. *IEEE Int. Conf.on Multimedia and Expo Workshops (ICMEW)*, pp. 1-4, Jul. 2013.

[36] Z. Gu, J. Zheng, N. Ling, P. Zhang, "Fast bi-partition mode selection for 3D HEVC depth intra coding," in Proc. *IEEE Int. Conf.on Multimedia and Expo (ICME)*, pp. 1-6, Sep. 2014.

[37] Q. Zhang, M. Chen, X. Huang, *et al.* "Low-complexity depth map compression in HEVC-based 3D video coding," *Eurasip J. Image Video Process.*, 2015.

[38] M. C. Li, Y. Lin, Y. Lin, *et al.* "A 3D HEVC Fast mode decision algorithm based on the depth information guided maximum coding level," in Proc. *IEEE Data Compression Conf. (DCC)*, pp. 413-413, Mar. 2014.

[39] A. D. Abreu, P. Frossard, and F. Pereira. "Optimizing multiview video plus depth prediction structures for interactive multiview video streaming," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 3, pp. 487-500, Apr. 2015.

[40] Sullivan, J. Gary, Boyce, *et al.* "Standardized extensions of High Efficiency Video Coding (HEVC)," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 1001-1016, Dec. 2013.

[41] D. Rusanovsky, K. Muller, and A. Vetro, "Common Test Conditions of 3DV Core Experiments," ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCT3V-A1100, Stockholm, Sweden, Jul. 2012.

[42] P. Helle, S. Oudin, B. Bross, et *al.* "Block merging for quadtree-based partitioning in HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1720-1731, Dec. 2012.

[43] Z. Pan, S. Kwong, M. Sun, *et al.* "Early MERGE mode decision based on motion estimation and hierarchical depth correlation for HEVC," *IEEE Trans. Broadcast.*, vol. 60, no. 2, pp. 405-412, Jun. 2014.

[44] M. Kang, Y. Ho, "Depth video coding using adaptive geometry based Intra prediction for 3D video systems," *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 121-128, Feb. 2012.

[45] L. Shen, Z. Liu, X. Zhang, *et al*. "An effective CU size decision method for HEVC encoders," *IEEE Trans. Multimedia*, vol. 15, no. 2, pp. 465-470, Feb. 2013.

[46] S. h. Jung, H. Park, "A fast mode decision method in HEVC using adaptive ordering of modes," *IEEE Trans. Circuits Syst. Video Technol.*, Doi: 10.1109/TCSVT.2015.2473303, pp. 1-13, 2015.

[47] B. Y. Chen, S. H. Yang, "Using H. 264 coded block patterns for fast inter-mode selection," in Proc. *IEEE Int. Conf. Multimedia and Expo(ICME)*, pp. 721-724, Apr. 2008.

[48] P. Merkle, K. Muller, and T. Wiegand, "Coding of depth signals for 3D video using wedgelet block segmentation with residual adaptation," in Proc. *IEEE Int. Conf. Multimedia and Expo(ICME)*, pp. 1-6, San Jose, USA, Jul. 2013.

[49] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," Doc. VCEG-M33, Thirteenth Meeting, Austin, Texas, USA , 2001.

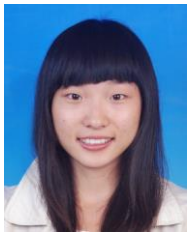
[50] Rec. ITU-T P.910, "Subjective video quality assessment methods for multimedia applications," International Telecommunication Union Recommendation, Apr. 2008.

[51] G. Sanchez, M. Saldanha, G. Balota, *et al.* "DMMFast: a complexity reduction scheme for three-dimensional high-efficiency video coding intraframe depth map coding," *Journal of Electronic Imaging*, vol. 24, no. 2, pp. 1-15, Apr. 2015.

Jianjun Lei (M'11) received the Ph.D. degree in signal and information processing from Beijing University of Posts and Telecommunications, Beijing, China, in 2007.

He was a visiting researcher at the Department of Electrical Engineering, University of Washington, Seattle, WA, from August 2012 to August 2013. He is currently a Professor with the School of Electronic Information Engineering, Tianjin University, Tianjin, China. His research interests include 3D video processing, 3D display, and computer vision.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2617332, IEEE Transactions on Circuits and Systems for Video Technology



Jinhui Duan received the B.S. degree in telecommunication engineering from Tianjin University, Tianjin, China, in 2014.

Currently, she is pursuing the M.S. degree at the School of Electronic Information Engineering, Tianjin University. Her research interests include 3-D video processing and video coding.

Feng Wu (M'99–SM'06–F'13) received the B.S. degree in Electrical Engineering from XIDIAN University in 1992. He received the M.S. and Ph.D. degrees in Computer Science from Harbin Institute of Technology in 1996 and 1999, respectively. Now he is a professor in University of Science and Technology of China and the dean of School of Information Science and Technology. Before that, he was principle researcher and research manager with Microsoft Research Asia.

His research interests include image and video compression, media communication, and media analysis and synthesis. He has authored or co-authored over 200 high quality papers (including several dozens of IEEE Transaction papers) and top conference papers on MOBICOM, SIGIR, CVPR and ACM MM. He has 77 granted US patents. His 15 techniques have been adopted into international video coding standards. As a co-author, he got the best paper award in IEEE T-CSVT 2009, PCM 2008 and SPIE VCIP 2007. Wu has been a Fellow of IEEE. He serves as an associate editor in IEEE Transactions on Circuits and System for Video Technology, IEEE Transactions on Multimedia and several other International journals. He got IEEE Circuits and Systems Society 2012 Best Associate Editor Award. He also serves as TPC chair in MMSP 2011, VCIP 2010 and PCM 2009, and Special sessions chair in ICME 2010 and ISCAS 2013.

Nam Ling (S'88–M'90–SM'99–F'08) received the B.Eng. degree from the National University of Singapore, Singapore, in 1981, and the M.S. and Ph.D. degrees from the University of Louisiana, Lafayette, LA, USA, in 1985 and 1989, respectively.

From 2002 to 2010, he was an Associate Dean with the School of Engineering, Santa Clara University, Santa Clara, CA, USA. He is currently the Sanfilippo Family Chair Professor and the Chair for the Department of Computer Engineering, Santa Clara University. He is also a Consulting Professor with the National University of Singapore, a Guest Professor for Tianjin University, Tianjin, China, a Guest Professor for Shanghai Jiao Tong University, Shanghai, China, a Cuiying Chair Professor for Lanzhou University, Gansu, China, and a Distinguished Professor for Xi'an University of Posts and Telecommunications, Shaanxi, China. He has authored or coauthored over 180 publications and standard contributions, including two books in the fields of video coding and systolic arrays. He has filed/granted over 20 U.S. patents.

Dr. Ling is an IEEE Fellow due to his contributions to video coding algorithms and architectures. He is also an IET Fellow. He was named as an IEEE Distinguished Lecturer twice and was also an APSIPA Distinguished Lecturer. He received the IEEE ICCE Best Paper Award (First Place) and the IEEE Umedia Best Paper Award. He was a recipient of six awards from Santa Clara University, four at the University level (Outstanding Achievement, Recent Achievement in Scholarship, President's Recognition, and Sustained Excellence in Scholarship) and two at the School/College level (Researcher of the Year and Teaching Excellence). He was a Keynote Speaker for IEEE APCCAS, VCVP (twice), JCPC, IEEE ICAST, IEEE ICIEA, IET FC Umedia, IEEE Umedia, ICCIT, and Workshop at XUPT (twice), as well as a Distinguished Speaker for IEEE ICIEA. He has served as a General Chair/Co-Chair for IEEE Hot Chips, VCVP (twice), IEEE ICME, IEEE Umedia (four times), and IEEE SiPS. He has also served as a Technical Program Co-Chair for IEEE ISCAS, APSIPA ASC, IEEE APCCAS, IEEE SiPS (twice), DCV, and IEEE VCIP. He was a Technical Committee Chair for IEEE CASCOM TC and IEEE TCMM, and has served as a Guest Editor or Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: REGULAR PAPERS, the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, Springer JSPS, Springer MSSP, and more.

Chunping Hou received the M.Eng. and Ph.D. degrees, both in electronic engineering, from Tianjin University, Tianjin, China, in 1986 and 1998, respectively.

Since 1986, she has been the faculty of the School of Electronic and Information Engineering, Tianjin University, where she is currently a Full Professor and the Director of the Broadband Wireless Communications and 3D Imaging Institute. Her current research interests include 3D image processing, 3D display, wireless communication, and the design and applications of communication systems.