Low-complexity depth map compression in HEVC-based 3D video coding

Qiuwen Zhang[1*], Ming Chen[2], Xinpeng Huang[1], Nana Li[1] and Yong Gan[1]

Abstract

In this paper, a low-complexity algorithm is proposed to reduce the complexity of depth map compression in the high-efficiency video coding (HEVC)-based 3D video coding (3D-HEVC). Since the depth map and the corresponding texture video represent the same scene in a 3D video, there is a high correlation among the coding information from depth map and texture video. An experimental analysis is performed to study depth map and texture video correlation in the coding information such as the motion vector and prediction mode. Based on the correlation, we propose three efficient low-complexity approaches, including early termination mode decision, adaptive search range motion estimation (ME), and fast disparity estimation (DE). Experimental results show that the proposed algorithm can reduce about 66% computational complexity with negligible rate-distortion (RD) performance loss in comparison with the original 3D-HEVC encoder.

Keywords: 3D video; 3D-HEVC; Depth coding; Low complexity

1 Introduction

Three-dimensional video standard has been recently finalized by the Joint Collaborative Team on 3D Video Coding (JCT-3V), and the high-efficiency video coding (HEVC)-based 3D video coding (3D-HEVC) is developed as an extension of HEVC [1-3]. For the efficient compression of 3D video data with multiview texture video and depth map, a number of coding tools are investigated to exploit in 3D-HEVC such as inter-view motion prediction and disparity-compensated prediction [4]. This technique achieves the highest possible coding efficiency in multiview texture video compression, but it results in extremely large encoding time with small increase of depth coding efficiency which obstructs it from 3D-HEVC practical use. Therefore, it is necessary to develop a fast algorithm that can reduce the complexity of multiview depth map compression with minimal loss of coding efficiency in a 3D-HEVC encoder.

Recently, a number of approaches have been made to explore fast algorithms in depth map coding. A motion vector (MV) sharing algorithm is proposed in [5] to reduce the complexity of depth map coding. An early termination algorithm for depth coding is introduced in [6] based on the detection of the differences between the current macroblock (MB) and the co-located MBs in texture video. An intra prediction algorithm for depth coding is presented in [7] to reduce the number of candidate prediction directions for smooth regions. A low-complexity mode decision and motion estimation algorithm is proposed in [8] to take advantage of the texture motion information which may be usefully exploited in the encoding of the corresponding depth map. A novel depth and depth-color codec is proposed in [9] based on a shape-adaptive wavelet transform and an explicit encoding of the locations of major depth edges. A depth map compression algorithm [10] uses the corresponding texture video as side information to improve the coding performance. A fast motion search and mode decision algorithm is proposed in [11] to speed up the motion estimation (ME) stages of the depth coding process, and a fast depth map method is proposed in our previous work [12] based on sharing motion vector and SKIP mode from the texture video to reduce complexity of depth coding. All these algorithms are efficient in reducing computational complexity with acceptable quality degradation in coding performance for previous video coding standards. However, these algorithms are not directly applicable to the new standard 3D-HEVC, where high

* Correspondence: zhangqwen@126.com
[1]College of Computer and Communication Engineering, Zhengzhou University of Light Industry, No. 5 Dongfeng Road, Zhengzhou 450002, China
Full list of author information is available at the end of the article

Zhang *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:2

Page 2 of 14

computational complexity is intrinsically related to the use of new prediction coding structures for the 3D-HEVC encoder.

To this end, several fast algorithms [13-16] have been proposed for the 3D-HEVC encoder to reduce the complexity of depth map coding. A fast mode decision algorithm is proposed in [13] to early terminate the unnecessary prediction modes full rate-distortion (RD) cost calculation in 3D-HEVC. A low-complexity depth map coding algorithm based on the associated texture video is introduced in [14] to reduce the number of wedgelet candidates. A fast wedgelet partitioning algorithm is proposed in [15] to simplify the intra mode decision in 3D-HEVC depth map coding. A content adaptive complexity reduction algorithm is proposed in [16] to reduce the 3D-HEVC coding complexity by utilizing the correlations between the base view and the dependent view. The aforementioned algorithms are well developed for depth map coding achieving significant time savings in 3D-HEVC. However, the coding information correlations between the depth map and the texture video are not fully studied. This situation results in a limited time saving. There is still some room for further reduction of computational complexity of the 3D-HEVC depth map compression.

The depth map represents a 3D scene information, which has the same content with similar characteristic of the texture video. Therefore, there is a high correlation among motion information from depth map and texture video. In this paper, we propose a low-complexity depth compression algorithm using the correlation among motion information from depth map and texture video. The proposed algorithm consists of three approaches: early termination mode decision, adaptive search range ME, and fast disparity estimation (DE) for depth map coding. Experimental results illustrate that the proposed algorithm can significantly reduce the computational complexity of depth map compression while maintaining almost the same coding performance in comparison with the original 3D-HEVC encoder.

The rest of the paper is organized as follows. Section 2 analyzes the property of depth map and the correlation among motion information from depth map and texture video. A low-complexity depth coding algorithm base on adaptive search range ME and fast DE is presented in Section 3. Experimental results and conclusions are given in Sections 4 and 5, respectively.

2 Observations and analysis

In the test model of 3D-HEVC, the variable sizes the ME and DE to exploit both temporal and view correlation within temporally successive pictures and neighboring views. The coding unit (CU) is the basic unit of region splitting used for 3D-HEVC similar to macroblock in H.264/AVC, which has a hierarchical quadtree structure having variable sizes from 64×64 to 8×8. The partition unit (PU) is the basic unit used for 3D-HEVC inter/intra prediction processes. At each treeblock, 3D-HEVC performs ME and DE with different PU sizes including $2N \times 2N$, $2N \times N$, $N \times 2N$, and $N \times N$.

Similar to HEVC for a treeblock, the mode decision process in 3D-HEVC is performed using all the possible prediction modes to find the one with the least RD cost using a Lagrange multiplier. The RD cost function (J) used in 3D-HEVC is defined as follows:

$$J = D + \lambda \cdot \text{SSE} \tag{1}$$

where D specifies the bit cost to be considered for the 3D-HEVC mode decision, SSE is the average difference between the current treeblock and the matching treeblock, and λ is the Lagrange multiplier. However, calculation of the RD cost needs to execute both the ME and DE processes in 3D-HEVC, and these 'try all and select the best' method will result in high computational complexity and limit the use of 3D-HEVC encoders in practical applications. Therefore, low-complexity algorithms, which can reduce the complexity of the ME and DE processes with negligible loss of coding efficiency, are extremely necessary for real-time implementation of 3D-HEVC encoders.

Since the depth map and its associated texture video are both projections of the same scenery from the same viewpoint at the same time instant, the motion characteristics (i.e., block partitioning and corresponding motion vectors) of the depth map and its associated texture video are typically similar. Therefore, a new coding mode motion parameter inheritance (MPI) [4,17], where the data that are already transmitted for the texture video picture can be reused for efficient encoding of the depth map, has been introduced in the 3D-HEVC encoder. This achieves the highest coding efficiency but requires a very high computational complexity. Since the motion vectors of the texture video have quarter-sample accuracy, whereas for the depth map only full-sample accuracy is used, in the inheritance process, the motion vectors are quantized to their nearest full-sample position. In addition, the inherited reference picture shall be the one with the same picture order count (POC) and viewpoint as the reference picture of the co-located block in the texture video picture. If there is no reference picture in the reference lists that satisfies this condition, such a candidate is treated as invalid and it is not inserted to the merge candidate list. However, the coding information correlations between the depth map and texture video are not fully studied. The coding information includes the reference picture, prediction mode, and motion vector.

Therefore, the prediction mode of the depth map tree-block is similar to that of the corresponding texture video treeblock. Meanwhile, the homogeneous regions in the depth map have a strong spatial correlation, and thus, spatially neighboring depth map treeblocks have similar coding information. The relationship among the current depth map treeblock, co-located texture video treeblock, and spatially neighboring treeblock is shown in Figure 1. The reference picture in the co-located texture view has the same POC value as the reference picture of current depth map view.

On the basis of these observations, we propose to analyze the depth intra prediction mode using the coding information from the spatial neighboring depth map and the co-located texture video treeblock. The neighboring depth map and the co-located texture video treeblock are described as in Figure 2. D_c denotes the current depth map treeblock, D_l, D_u, D_{ul}, and D_{ur} denote the neighboring treeblocks in the depth map. C_{col} denotes the co-located treeblock in the texture video and C_l, C_u, C_{ul}, and C_{ur} its left treeblock, up treeblock, upleft treeblock, and upright treeblock, respectively, as shown in Figure 2.

According to the coding information correlation with the mode maps of encoded frames, we define a set of intra mode predictors (P) for depth map treeblock as follows:

$$P = \{D_l, D_u, D_{ul}, D_{ur}, C_{col}, C_l, C_u, C_{ul}, C_{ur}\} \qquad (2)$$

Based on this predictor set, a mode complexity (C) parameter is defined according to the mode context of

the spatial neighboring depth map and the co-located texture video treeblock, and then, the mode characteristic of a depth map treeblock is estimated. The mode complexity of a depth map treeblock is described as follows:

$$C = \sum_{i \in P} \beta_i \cdot \eta_i \qquad (3)$$

where i is the related treeblock in predictors P, β_i is the treeblock weight factor of each predictor in Equation 2, and η_i is the treeblock mode factor of each predictor. Only the prediction modes of those available neighboring treeblocks in predictors P will be used. In 3D-HEVC, various prediction mode sizes are used in the mode decision process. The mode factor of each predictor η_i can be assigned based on the complexity of each mode as follows: when the predictor i is SKIP mode, merge mode, inter $2N \times 2N$, and intra $2N \times 2N$ mode, η_i is assigned with a small value '1;' when the predictor i is inter $2N \times N$, inter $N \times 2N$ mode, η_i is assigned with a medium value '2;' when the predictor i is small-size inter modes, intra $N \times N$ mode (depth modeling modes (DMM) and region boundary chain (RBC) mode in the neighboring depth map treeblocks), and DE mode, η_i is assigned with a large value '3.' The treeblock weight factors of these nine predictors have an additional property, $\sum_i \beta_i = 1$. β_i is defined according to the effect of related treeblocks on current treeblock. Since treeblocks in the horizontal and vertical directions have a large effect on the current treeblock compared to treeblocks in the diagonal direction, the weight factors β_i for the horizontal
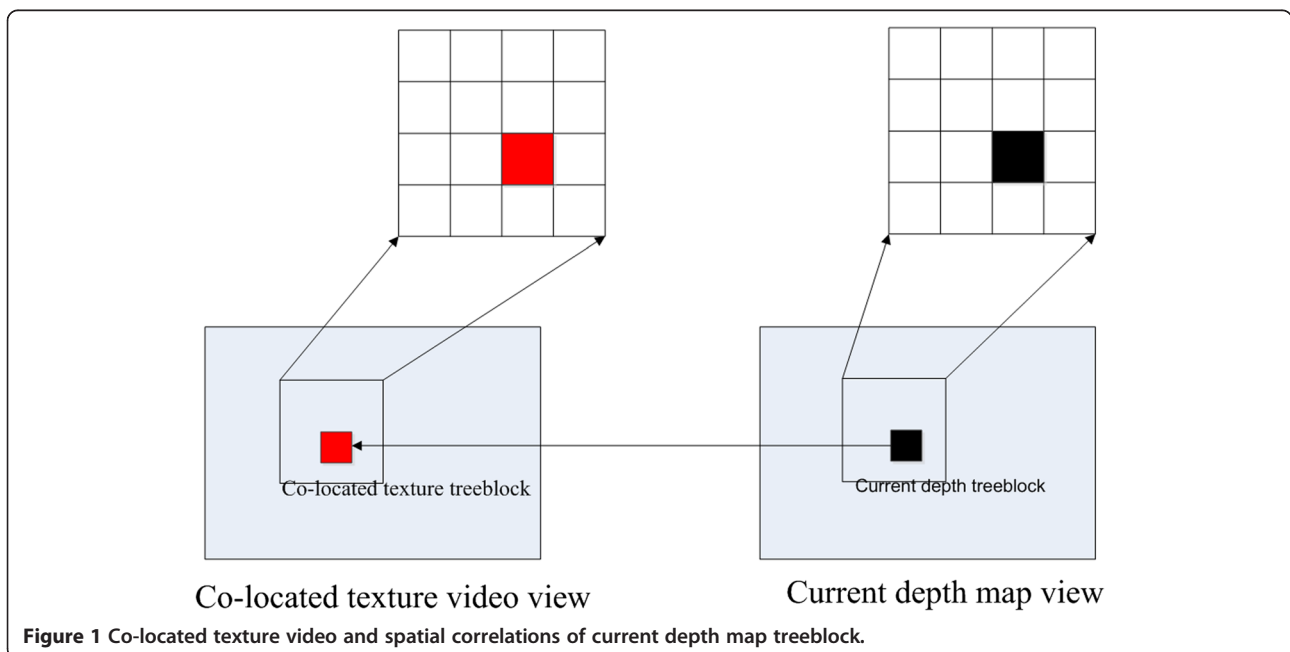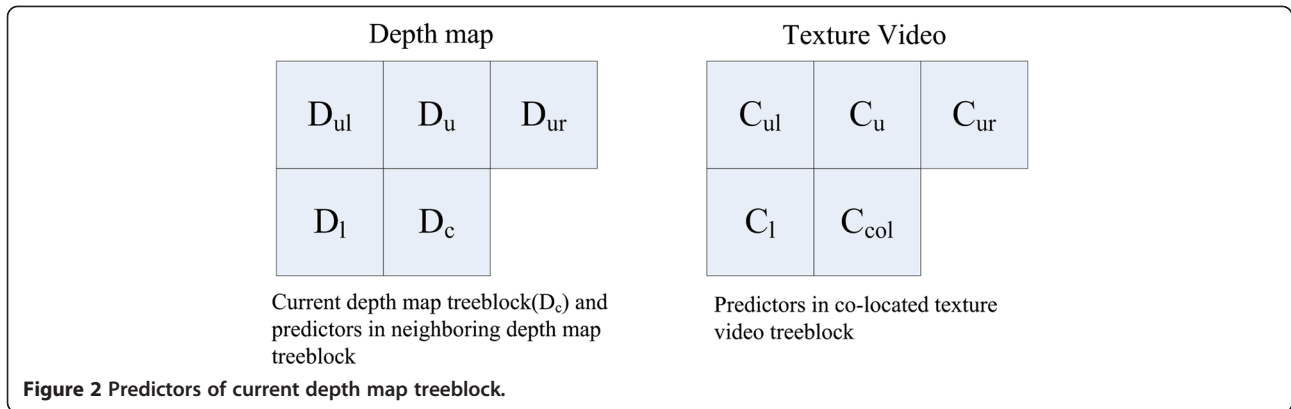

Figure 1 Co-located texture video and spatial correlations of current depth map treeblock.

Co-located texture treeblock

Current depth treeblock

Co-located texture video view

Current depth map view

Zhang *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:2

Page 4 of 14

Figure 2 Predictors of current depth map treeblock.

and vertical treeblocks (D_l, D_u, C_l, and C_u) are set to 0.1, and that of the diagonal direction treeblocks (D_{ul}, D_{ur}, C_{ul}, and C_{ur}) are set to 0.05. In the case of the co-located texture video treeblock, the treeblock weight factor β_{Ccol} is set to 0.4.

Generally, the larger the mode factor, the more complex the treeblock is. According to the value of C, each treeblock can be divided into three types. T_1 and T_2 are set to determine whether a treeblock belongs to the region with the simple mode, normal mode, or complex mode. The criterion is defined as follows:

$$\begin{cases} C \le T_1 & \textbf{Treeblock} \in \textbf{simple mode region} \\ T_1 < C < T_2 & \textbf{Treeblock} \in \textbf{normal mode region} \\ C \ge T_2 & \textbf{Treeblock} \in \textbf{complex mode region} \end{cases}$$

(4)

where T_1 and T_2 are mode-weight factors. Those threshold settings are crucial for effective depth map compression, and it is always a tradeoff between depth map coding quality and computational complexity reduction. From simulations on various test sequences, it can be found that the optimal threshold for each sequence depends on the sequence content. In order to cope with different texture characteristics of test sequences, extensive simulations have been conducted on eight video sequences to analyze the thresholds for three types of treeblocks. Among these test sequences, *Kendo*, *Balloons*, and *Newspaper* are in 1,024 × 768 resolution, while *Undo_Dancer*, *GT_Fly*, *Poznan_Street*, *Poznan_-Hall2*, and *Shark* are in 1,920 × 1,088 resolution, and the 'Shark' and 'Undo_Dancer' sequences are with a large global motion or rich texture, the 'Kendo', 'Balloons', 'Newspaper', and 'Poznan_Street' sequences are with a medium local motion or a smooth texture, and 'Poznan_-Hall2' is a small global motion or a homogeneous texture sequence. The test conditions are as follows: I-B-P view structure; test full-length frames for each sequence; quantization parameter (QP) is chosen with 34, 39, 42, and 45; group of pictures (GOP) size = 8; treeblock size =

64; search range of ME is configured with 64; and context-adaptive binary arithmetic coding (CABAC) is used for entropy coding. Then, we calculated the average thresholds of those eight test sequences.

Table 1 shows the accuracies of the proposed algorithm using various thresholds. The accuracies here are defined as the ratio of the number of the simple mode, normal mode, and complex mode, which select the same best modes using the 3D-HEVC encoder as well as the proposed algorithm. It can be seen from Table 1 that when the threshold values are $T_1 = 0.8$, $T_2 = 1.2$, the average accuracy of the proposed algorithm achieves more than 93% with a maximum of 97% in the 'Shark' sequence. Based on extensive experiments, T_1 and T_2 are set to 0.8 and 1.2, respectively, which achieve a good and consistent performance on a variety of test sequences with different texture characteristics and motion activities and fixed for each treeblock QP level in 3D-HEVC encoder.

3 Proposed low-complexity depth map compression algorithm

3.1 Early termination mode decision

The depth map is usually not the ground truth because existing depth map estimation methods still have difficulties to generate accurate depths at object edges or in areas with less texture. Distortion may occur during depth map estimation, which will result in a noisy depth map (caused by occlusion and areas of low texture), such that it would be inefficient to spend more bits to achieve an accurate representation of the depth map in 3D-HEVC coding. To overcome this problem, this paper proposes an early termination mode decision for 3D-HEVC, which takes into account the correlations between coding information from texture videos and depth maps to speed up the coding process.

The depth map content is similar with that of texture video, and thus, the coding modes of texture and depth map are similar. By utilizing the information of the corresponding treeblock in the texture video, the coding

Zhang *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:2

Page 5 of 14

Table 1 Statistical analysis of accuracy for proposed algorithm using various thresholds

Sequences	$T_1 = 0.6$, $T_2 = 1.0$ (%)	$T_1 = 0.7$, $T_2 = 1.1$ (%)	$T_1 = 0.8$, $T_2 = 1.2$ (%)	$T_1 = 0.9$, $T_2 = 1.3$ (%)
Kendo	84	86	93	91
Balloons	76	83	91	92
Newspaper	81	85	94	92
Shark	85	89	97	95
Undo_Dancer	82	88	93	91
GT_Fly	80	84	91	87
Poznan_Street	81	85	89	90
Poznan_Hall2	82	86	95	93
Average	81	86	93	91

information of previously encoded texture images at the same view can be effectively shared and reused. Such that we propose a novel early termination mode decision considering a co-located texture video. The merge/skip mode provides good coding performance and requires little complexity in the 3D-HEVC encoder, where the motion vector predictor (MVP) is adopted for the current treeblock to generate a compensated block. Meanwhile, the merge/skip mode is the dominant mode at low bitrates (high QPs) in the 3D-HEVC encoder, and the distribution is similar to that in the previous video coding standard, H.264/AVC. Once the merge/skip mode can be predecided, variable size ME and DE computation for a treeblock can be entirely saved. Usually, the decision to use merge/skip mode is delayed until the RD costs of all other modes (inter-, intra-, and DE-modes) have been calculated and merge/skip mode is found to have the minimum RD cost. Thus, if we can exploit previously encoded texture coding information to determine that those depth map treeblocks are encoded in merge/skip mode (this mode along with CU partition inherited to encode forcefully the depth treeblock without going further in depth quadtree level), we can skip the time-consuming process of computing RD costs on smaller block sizes for a high percentage of treeblocks and, thus, significantly reduce the computation complexity of the 3D-HEVC mode decision process.

Based on this consideration, the proposed algorithm introduces an early termination mode decision to skip checking unnecessary ME and DE by utilizing the co-located texture video prediction mode information. In our approach, we first take advantage of the relations of previously encoded texture images at the same view for early merge/skip mode decision. Since both depth map and texture video are generally captured at the same time, it is likely for each treeblock to have the same motion and block partition information. So when a

treeblock of depth map is encoded, we consider how the corresponding texture video treeblock (C_{col} in Figure 2) was encoded. When the merge/skip mode is selected as the best prediction mode on the texture treeblock in the 3D-HEVC mode decision, it indicates that the current texture treeblock is located in a low-motion or static region. The motion of the texture treeblock can be predicted well using the merge/skip mode, which results in a lower energy residual after motion compensation compared to other prediction modes such as inter $2N \times 2N$, $2N \times N$, $N \times 2N$, and $N \times N$. Thus, no further processing of variable size ME and DE computation is necessary.

However, the proposed early termination mode decision algorithm has a few strong assumptions: depth map content is not always similar to the color content, e.g., in a planar highly textured area, there is a high-color variance but depth is constant. Depth acquisition can be unreliable but the assumption that information can be discarded for this reason is questionable. Finally, if motion estimation on color data is wrong with the proposed approach, errors can propagate to depth data even if the estimation from depth could be correct. Based on this observation, we investigate the effectiveness of the proposed early termination mode decision algorithm. By exploiting the exhaustive mode decision in the 3D-HEVC encoder under the aforementioned test conditions in Section 2, extensive simulations have been conducted on a set of test sequences as listed in Table 2. Table 2 shows the hit rate of the early termination mode decision algorithm. This hit rate is defined as the ratio of the number of depth map treeblocks, which selects the same best prediction mode using the 3D-HEVC encoder as well as the proposed algorithm, to the total number of depth map treeblocks. The average hit rate of the proposed algorithm is larger than 93% with a maximum of 95% in 'QP = 45' and a minimum of 91% in 'QP = 34'. The

Table 2 Hit rate of the proposed early termination mode decision algorithm

Sequences	QP = 34 (%)	QP = 39 (%)	QP = 42 (%)	QP = 45 (%)
Kendo	92	93	94	95
Balloons	86	89	91	93
Newspaper	92	93	95	96
Shark	93	95	96	97
Undo_Dancer	91	92	93	95
GT_Fly	87	90	92	93
Poznan_Street	92	94	95	96
Poznan_Hall2	94	95	97	98
Average	91	93	94	95

Zhang et al. EURASIP Journal on Image and Video Processing (2015) 2015:2

Page 6 of 14

Table 3 Statistical analysis of motion vector distribution for three types of treeblocks

Sequences	Treeblocks in simple mode region			Treeblocks in normal mode region			Treeblocks in complex mode region		
	S1 (%)	S2 (%)	S3 (%)	S1 (%)	S2 (%)	S3 (%)	S1 (%)	S2 (%)	S3 (%)
Kendo	87.3	98.1	98.9	81.8	92.1	97.2	59.4	76.8	84.3
Balloons	85.5	96.3	97.1	79.6	91.3	96.1	47.2	69.2	79.8
Newspaper	86.2	97.4	98.3	78.9	90.8	96.5	49.8	71.3	81.6
Shark	89.2	97.9	99.4	82.1	92.5	97.3	51.5	73.4	86.2
Undo_Dancer	90.1	98.3	99.7	83.8	92.8	98.2	57.2	76.6	87.3
GT_Fly	91.2	99.2	100	84.1	93.2	99.4	59.3	78.2	90.2
Poznan_Street	88.6	96.5	98.2	77.5	90.3	96.5	43.2	69.3	81.4
Poznan_Hall2	86.3	96.8	97.6	76.3	91.1	95.7	39.8	64.7	78.2
Average	88.1	97.6	98.7	80.5	91.8	97.1	50.9	72.4	83.6

'S1', 'S2', and 'S3', respectively, represent the motion search windows of [SR/16 × SR/16], [SR/8 × SR/8], and [SR/4 × SR/4].

simulation results shown in Table 2 indicate that the proposed early termination mode decision algorithm can accurately reduce the unnecessary depth map CU mode by utilizing the information of the corresponding treeblock in texture video.

Based on this statistical tendency, the proposed depth map early termination algorithm checks the prediction modes from the co-located texture video: if texture treeblock (C_{col}) has no motion, corresponding depth map treeblock (D_c) has motion due to unreliable depth estimation; therefore, the motion in the depth map can be ignored. When the texture video treeblock selects merge/skip as the best mode, it indicates that the motion can be efficiently represented using the current depth

map treeblock, and the variable size ME and DE computation for a depth map treeblock can be skipped in the 3D-HEVC mode decision.

3.2 Adaptive search range motion estimation

ME is the most computationally expensive task in the 3D-HEVC encoder, which is defined as the search of the best matched treeblock within a predefined region in the reference frame. The larger ME search range produces higher computational load, and a very small ME search range may reduce the coding performance due to poor matching results. A suitable ME search range can reduce the computational complexity of 3D-HEVC and also maintain the good RD performance.
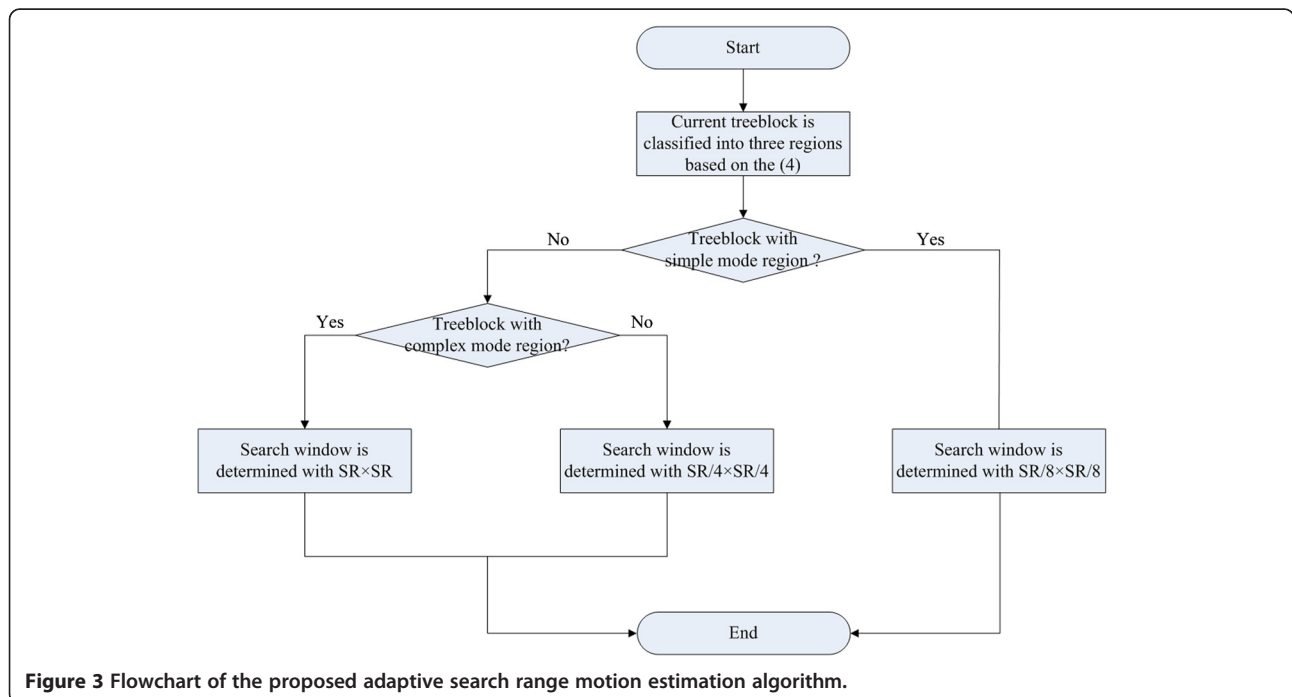

Figure 3 Flowchart of the proposed adaptive search range motion estimation algorithm.

Zhang *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:2

Page 7 of 14

Table 4 Analysis of view prediction and temporal prediction distributions for three treeblock types

Sequences	Treeblocks in simple mode region		Treeblocks in normal mode region		Treeblocks in complex mode region	
	T (%)	V (%)	T (%)	V (%)	T (%)	V (%)
Kendo	98.9	1.1	88.2	11.8	60.2	39.7
Balloons	97.5	2.5	84.1	15.9	68.6	31.4
Newspaper	99.7	0.3	92.3	7.8	71.3	29.7
Shark	98.2	1.8	85.2	14.8	64.2	35.9
Undo_Dancer	97.7	2.2	92.7	7.3	58.9	41.1
GT_Fly	95.8	4.2	89.9	10.1	54.7	45.3
Poznan_Street	97.9	2.1	88.4	11.6	61.5	38.5
Poznan_Hall2	96.2	3.7	91.7	8.3	70.1	30.0
Average	97.7	2.2	89.1	11.0	63.7	36.4

'T' and 'V' represent temporal prediction and view prediction, respectively.

In 3D video coding, since both depth map and texture video represent the same scene, it is likely for the depth map and texture video treeblock to have similar motion information. Based on this observation, we propose to use the mode complexity parameter (C) defined in Equation 3 to speed up the procedure of depth map coding DE search range computational complexity. According to the motion complexity of a depth map treeblock (based on Equation 3), we first classify the depth map treeblock to different categories in terms of DE search range as follows:

$$\text{Search range}_{\text{depth}} = \begin{cases} SR/8 \times SR/8 & \text{Treeblock} \in \text{simple mode region} \\ SR/4 \times SR/4 & \text{Treeblock} \in \text{normal mode region} \\ SR \times SR & \text{Treeblock} \in \text{complex mode region} \end{cases}$$

$$(5)$$

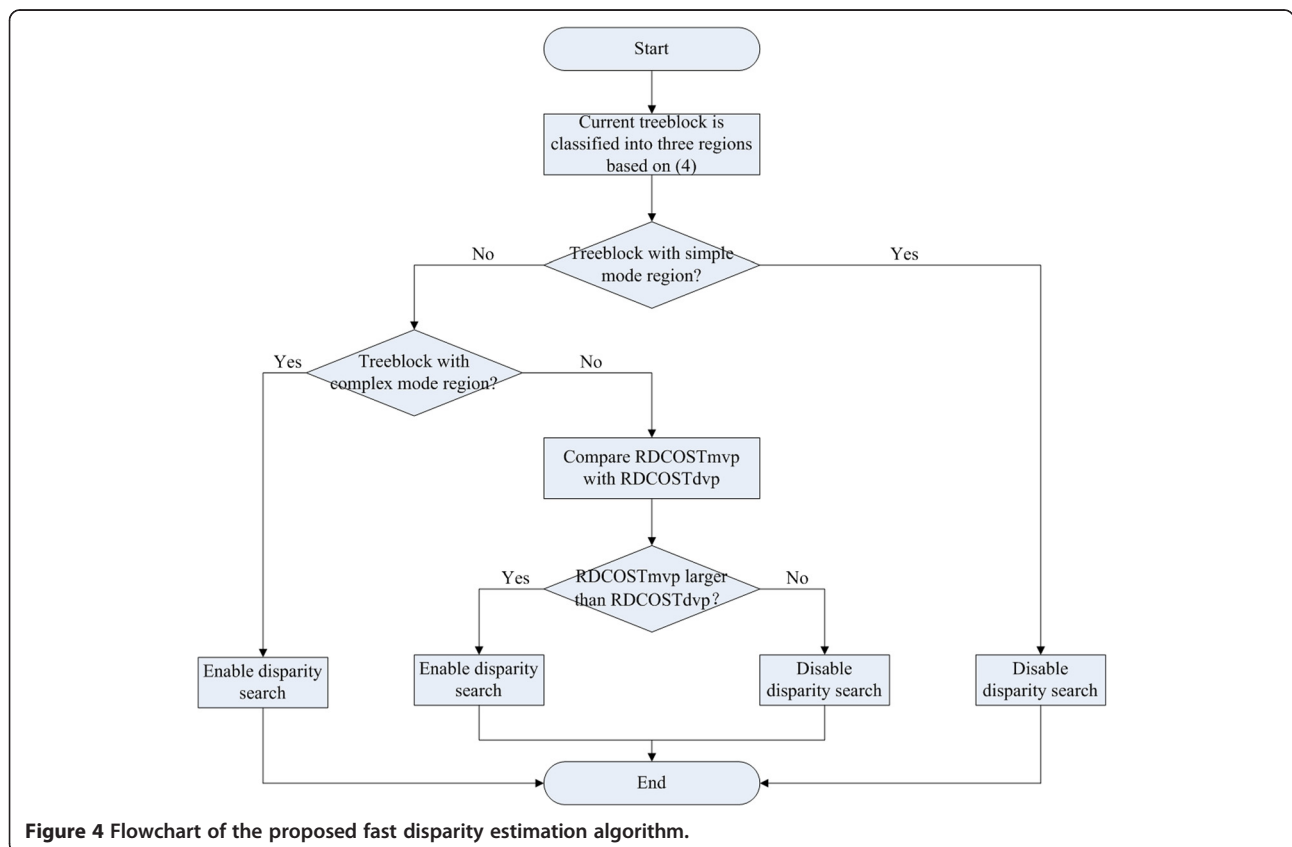where SR represents the search range defined in the configuration file of the 3D-HEVC, and Search

Figure 4 Flowchart of the proposed fast disparity estimation algorithm.

Zhang *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:2

Page 8 of 14

Table 5 Test sequence information

Sequence	Resolution	Frames	Two-view case	Three-view case
Kendo	1,024 × 768	300	1-3	1-3-5
Balloons	1,024 × 768	300	1-3	1-3-5
Newspaper	1,024 × 768	300	2-4	2-4-6
Shark	1,920 × 1,088	300	1-5	1-5-9
Undo_Dancer	1,920 × 1,088	250	1-5	1-5-9
GT_Fly	1,920 × 1,088	250	9-5	9-5-1
Poznan_Street	1,920 × 1,088	250	5-4	5-4-3
Poznan_Hall2	1,920 × 1,088	200	7-6	7-6-5

range$_{depth}$ is the adjusted search range of the corresponding treeblock in the depth map.

To verify legitimacy of the proposed adaptive search range motion estimation algorithm, extensive simulations have been conducted on eight video sequences to analyze the motion vector distribution for these three types of treeblocks. By exploiting the exhaustive mode decision in 3D-HEVC under the aforementioned test conditions, we investigate the motion vector distribution for these three types of treeblocks.

Table 3 shows the motion vector distribution for each type of treeblocks. It can be seen from Table 3 that for treeblocks in the simple mode region, more than 97% of all motion vectors lie in the [SR/8 × SR/8] window. In other words, if the maximum search range is set to SR/8, it will most likely cover about 97% of all motion vectors. For the treeblocks in normal mode region, about 97% of all motion vectors lie in the [SR/4 × SR/4] window. If the maximum search range is set to be SR/4, it will most likely cover about 97% of motion vectors. For the treeblocks in the complex mode region, the percentage of all motion vectors that lie in the [SR/16 × SR/16], [SR/8 × SR/8], and [SR/4 × SR/4] windows are relatively low, only about 51%, 72%, and 84%, respectively, and thus, 3D-HEVC motion vector search range cannot be reduced. The results shown in Table 3 demonstrate that the proposed adaptive search range motion estimation algorithm can accurately reduce the unnecessary ME search range in 3D-HEVC. A flowchart of the proposed adaptive search range motion estimation algorithm is given in Figure 3.

3.3 Fast disparity estimation for depth map coding

In the test model of 3D-HEVC, when coding the dependent views, the HEVC codec is modified by including some high-level syntax changes and the disparity-compensated prediction (DCP) techniques, similar to the inter-view prediction in the MVC extension of H.264/AVC [4]. In addition, different from coding dependent texture view, depth map is characterized by sharp edges and large regions with nearly constant values. The eight-tap interpolation filters that are used for ME interpolation in HEVC can produce ringing artifacts at sharp edges in the depth map, which are visible as disturbing components in synthesized intermediate views. For avoiding this issue and for decreasing the encoder and decoder complexity, the ME as well as the DE has been modified in a way that no interpolation is used. That means, for depth map, the inter-picture prediction is always performed with full-sample accuracy. For the actual DE, a block of samples in the reference picture is directly used as the prediction signal without interpolating any intermediate samples. In order to avoid the transmission of motion and disparity vectors with an unnecessary accuracy, full-sample accurate motion and disparity vectors are used for coding the depth map. The transmitted motion vector differences are coded using full-sample instead of quarter-sample precision. This modified technique achieves the highest possible depth map coding efficiency, but it results in extremely large encoding time which obstructs 3D-HEVC from practical application. In this paper, a fast DE algorithm for depth map coding is proposed to reduce 3D-HEVC computational complexity.

Table 6 Results of each individual algorithm compared to 3D-HEVC encoder in two-view case

Sequences	ETMD			ASRME			FDE		
	BDBR (%)	BDPSNR (dB)	Dtime (%)	BDBR (%)	BDPSNR (dB)	Dtime (%)	BDBR (%)	BDPSNR (dB)	Dtime (%)
Kendo	0.02	−0.01	−20.1	0.13	−0.03	−32.4	0.02	−0.00	−18.2
Balloons	0.01	−0.01	−17.3	0.09	−0.02	−36.8	0.02	−0.00	−22.7
Newspaper	0.00	−0.02	−26.8	0.21	−0.02	−42.1	0.04	−0.01	−15.6
Shark	−0.07	0.00	−9.8	0.07	−0.01	−61.2	0.75	−0.06	−30.8
Undo_Dancer	−0.12	0.01	−12.1	0.05	−0.01	−55.8	0.63	−0.05	−38.2
GT_Fly	−0.08	0.00	−7.6	0.08	−0.01	−63.2	0.82	−0.07	−36.9
Poznan_Street	0.01	−0.01	−26.5	0.21	−0.03	−43.9	0.06	−0.00	−16.5
Poznan_Hall2	0.02	−0.02	−31.2	0.16	−0.02	−42.6	0.09	−0.01	−24.3
Average	−0.03	−0.01	−18.9	0.13	−0.02	−47.3	0.30	−0.03	−25.4

Zhang *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:2

Page 9 of 14

Table 7 Results of each individual algorithm compared to 3D-HEVC encoder in three-view case

Sequences	ETMD			ASRME			FDE		
	BDBR (%)	BDPSNR (dB)	Dtime (%)	BDBR (%)	BDPSNR (dB)	Dtime (%)	BDBR (%)	BDPSNR (dB)	Dtime (%)
Kendo	0.01	−0.01	−20.9	0.10	−0.02	−33.1	0.02	−0.00	−18.7
Balloons	0.01	−0.00	−18.1	0.07	−0.02	−37.3	0.02	−0.00	−23.6
Newspaper	0.00	−0.02	−27.8	0.19	−0.02	−42.7	0.03	−0.01	−15.9
Shark	−0.05	0.00	−10.4	0.07	−0.01	−63.4	0.68	−0.05	−32.3
Undo_Dancer	−0.09	0.01	−12.7	0.05	−0.01	−56.7	0.57	−0.04	−39.7
GT_Fly	−0.07	0.00	−7.9	0.07	−0.01	−64.5	0.71	−0.05	−38.5
Poznan_Street	0.01	−0.01	−28.3	0.18	−0.03	−44.1	0.05	−0.00	−17.1
Poznan_Hall2	0.02	−0.01	−32.7	0.15	−0.02	−43.6	0.07	−0.01	−25.4
Average	−0.02	−0.01	−19.9	0.11	−0.02	−48.2	0.27	−0.02	−26.4

As mentioned in the above, disparity prediction is used to search the best matched block in frames from neighbor views. Although temporal prediction is generally the most efficient prediction mode in 3D-HEVC, it is sometimes necessary to use both DE and ME rather than only use ME to achieve better predictions. In general, temporal motion cannot be characterized adequately, especially for regions with non-rigid motion and regions with motion boundaries. For the former, ME based on simple translation movement usually fails and, thus, produces a poor prediction. For the latter, regions with motion boundaries are usually predicted using small mode sizes with a larger magnitude of motion vectors and higher residual energy [18]. Thus, the treeblocks with a simple mode region are more likely to choose temporal prediction (ME), and treeblocks with a complex mode region are more likely to choose inter-view prediction (DE).

By exploiting the exhaustive mode decision in the 3D-HEVC encoder under the aforementioned test experimental conditions in Section 3.2, we investigate the probabilities of choosing inter-view prediction and temporal prediction for each type of treeblocks in Table 4. For treeblocks with a simple mode region, the average

probabilities of choosing temporal prediction and inter-view prediction are 97.7% and 2.2%, respectively. For treeblocks with a normal mode region, they are 89.1% and 11.0%, respectively. For treeblocks with a complex mode region, the probabilities are 63.7% and 36.4%, respectively. We can see from Table 4 that treeblocks with a simple mode region are much more likely to choose temporal prediction. Thus, for a simple mode region, the procedure of the inter-view prediction can be skipped with only a very low miss detection ratio by using the optimal prediction mode chosen by the full inter-view and temporal prediction modes. But for complex mode region treeblocks and treeblocks with a normal mode region, the average probabilities of choosing inter-view prediction are 36.4% and 11.0%, respectively. Although the test sequences such as 'Poznan_Hall2' and 'Newspaper' contain a large area of the homogeneous textures and low-activity motion, which are more likely to be encoded with temporal prediction, the probability of inter-view prediction for a treeblock with a normal mode region and complex mode region is still highest. Thus, if we disable inter-view prediction in the normal mode region and complex mode region, the coding efficiency loss is not negligible.

Table 8 Comparing the proposed overall algorithm compared with 3D-HEVC encoder

Sequences	Two-view case			Three-view case		
	BDBR (%)	BDPSNR (dB)	Dtime (%)	BDBR (%)	BDPSNR (dB)	Dtime (%)
Kendo	0.19	−0.03	−47.3	0.17	−0.03	−49.8
Balloons	0.13	−0.04	−51.9	0.11	−0.03	−53.2
Newspaper	0.24	−0.05	−54.2	0.19	−0.05	−56.8
Shark	0.79	−0.06	−83.4	0.72	−0.05	−84.9
Undo_Dancer	0.68	−0.07	−80.7	0.63	−0.06	−82.1
GT_Fly	0.83	−0.05	−86.3	0.79	−0.05	−87.5
Poznan_Street	0.31	−0.04	−53.1	0.26	−0.04	−56.7
Poznan_Hall2	0.27	−0.04	−57.6	0.23	−0.03	−59.3
Average	0.43	−0.05	−64.3	0.39	−0.04	−66.3

Zhang *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:2

Page 10 of 14

Table 9 Comparing the proposed overall algorithm compared with CACRS algorithm [16]

Sequences	Two-view case			Three-view case		
	BDBR (%)	BDPSNR (dB)	Dtime (%)	BDBR (%)	BDPSNR (dB)	Dtime (%)
Kendo	−1.52	0.08	−14.7	−1.87	0.09	−15.1
Balloons	−1.21	0.06	−21.4	−1.58	0.07	−22.8
Newspaper	−0.81	0.03	−19.3	−0.92	0.04	−19.7
Shark	−0.21	0.01	−33.9	−0.31	0.02	−35.2
Undo_Dancer	−0.47	0.02	−29.2	−0.56	0.02	−31.1
GT_Fly	−0.28	0.01	−36.2	−0.32	0.01	−38.6
Poznan_Street	−0.96	0.07	−19.8	−1.26	0.08	−20.2
Poznan_Hall2	−2.12	0.11	−21.8	−2.87	0.13	−22.6
Average	−0.95	0.05	−24.5	−1.21	0.06	−25.7

Based on the aforementioned analysis, we propose a fast disparity estimation algorithm in which a disparity search is selectively enabled. For treeblocks with a simple mode region, disparity search is skipped (only the RD cost of the MVP is used); while for treeblocks with a normal mode region, the RD cost of the MVP is compared with that of the disparity vector predictor (DVP). If the RD cost of MVP is larger than that of DVP, the disparity search is enabled; otherwise, it is disabled. For treeblocks with a complex mode region, disparity search is enabled (all the RD cost of MVP and DVP are used). A flowchart of the scheme is given in Figure 4.

3.4 Overall algorithm
Based on the aforementioned analysis, including the approaches of early termination mode decision, adaptive search range ME and fast DE for depth map coding, we propose a low-complexity depth map compression algorithm for 3D-HEVC as follows.

Step 1: start mode decision for a depth map treeblock.
Step 2: locate the spatial neighboring depth map treeblock and its co-located texture video treeblocks (shown in Figure 2) at the previously coded data. Derive the coding information from predictors in the depth map and texture video.
Step 3: derive the prediction mode of the co-located texture video treeblocks; if texture treeblock has no motion, perform early merge/skip mode decision and go to Step 7, else go to Step 4.
Step 4: compute C based on Equation 3 and T_1 and T_2 based on Equation 4; classify the current depth map treeblock into the simple mode region, normal mode region, and complex mode region.
Step 5: perform adaptive search range ME determination: for the treeblocks in a simple mode region, the search range window is reconfigured with [SR/8 × SR/8]; for the treeblock in a normal mode

region, the search range window is with [SR/4 × SR/4]; otherwise, the search range window is unchanged.
Step 6: perform variable size DE: for treeblocks with a simple mode region, disparity search is skipped, while for treeblocks with a complex mode region, disparity search is enabled. For treeblocks with a normal mode region, the RD cost of the MVP is compared with that of the DVP.
Step 7: determine the best prediction mode. Go to step 1 and proceed with next depth map treeblock.

4 Experimental results
In order to confirm the performance of the proposed low-complexity depth map compression algorithm, which is implemented on the recent 3D-HEVC Test Model (HTM ver.5.1), we show the results obtained in the test on eight sequences released by the JCT-3V Group. The detailed information of the test sequences is provided in Table 5. All the experiments are defined under the common test conditions (CTC) [19] required by JCT-3V. The encoder configuration is as follows: two-view case (coding order: left-right) and three-view case (coding order: center-left-right); GOP length 8 with an intra period of 24; HEVC codecs are configured with 8-bit internal processing; the coding treeblock has a fixed size of 64 × 64 pixels and a maximum CU depth level of 4, resulting in a minimum CU size of 8 × 8 pixels; search range of the ME is configured with 64, inter-view motion prediction mode on, P-I-P inter-view prediction; and CABAC is used as the entropy coder. The proposed algorithm is evaluated with QP combinations for texture video and depth map (25, 34), (30, 39), (35, 42), and (40, 45) (the first number is the texture video QP, the second number gives the depth map QP). The experiments test full-length frames for each sequence. After encoding, the intermediate rendered views were synthesized between each view. The intermediate rendered views are generated at the receiver using view
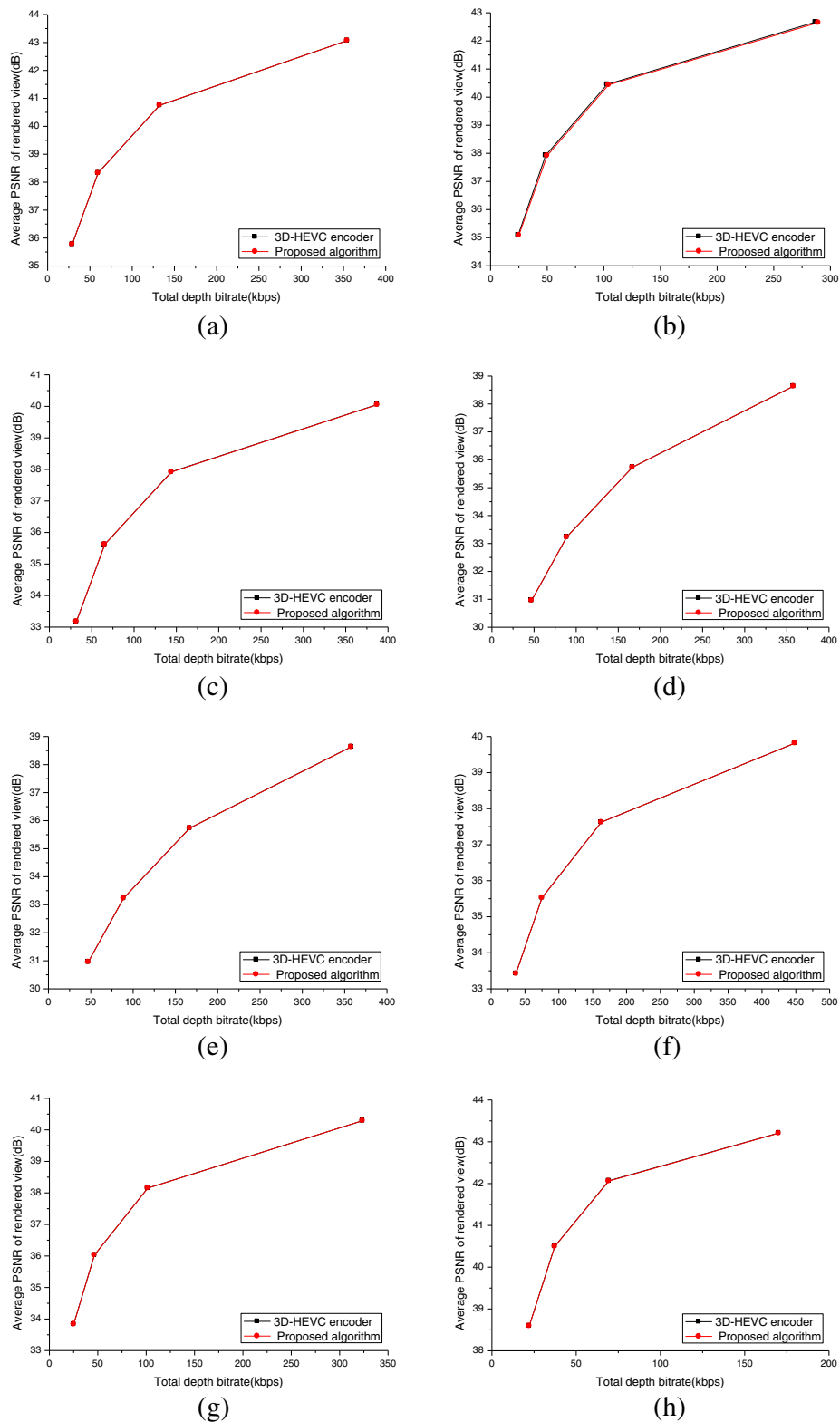
Zhang *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:2

Page 11 of 14

Figure 5 RD curves of proposed overall algorithm and 3D-HEVC under different QP combinations. RD curves of proposed overall algorithm and 3D-HEVC under different QP combinations for texture video and depth map (25, 34), (30, 39), (35, 42), and (40, 45). **(a)** Kendo, **(b)** Balloons, **(c)** Newspaper, **(d)** Shark, **(e)** Undo_Dancer, **(f)** GT_Fly, **(g)** Poznan_Street, **(h)** Poznan_Hall2.
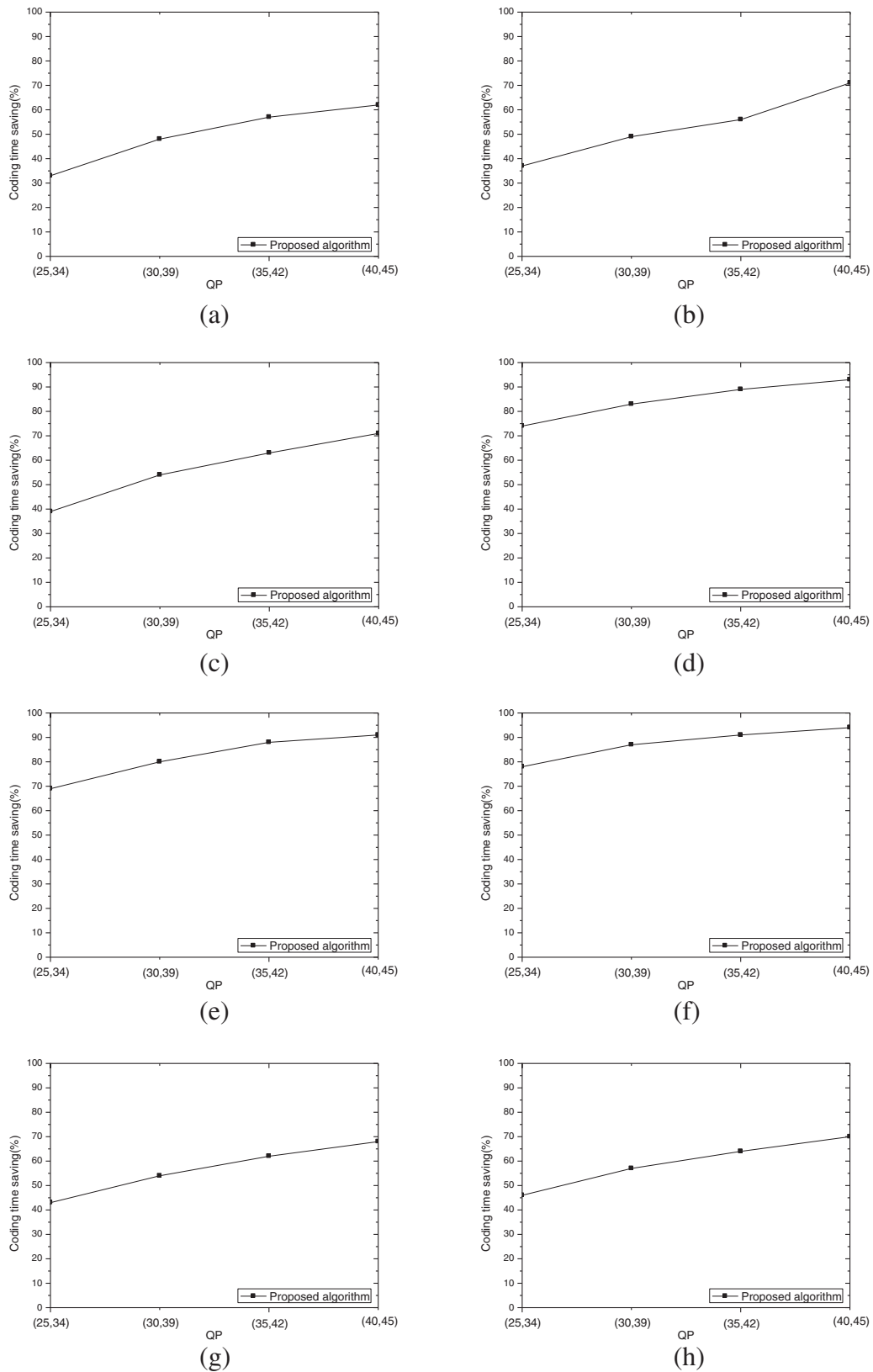
Zhang *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:2

Page 12 of 14

Figure 6 (See legend on next page.)

Zhang *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:2

Page 13 of 14

(See figure on previous page.)

Figure 6 Time-saving curves of proposed overall algorithm compared to 3D-HEVC under different QP combinations. Time-saving curves of proposed overall algorithm compared to 3D-HEVC under different QP combinations for texture video and depth map (25, 34), (30, 39), (35, 42), and (40, 45). **(a)** Kendo, **(b)** Balloons, **(c)** Newspaper, **(d)** Shark, **(e)** Undo_Dancer, **(f)** GT_Fly, **(g)** Poznan_Street, **(h)** Poznan_Hall2.

synthesis reference software (VSRS) algorithm provided by MPEG [20]. Since the depth map sequences are used for rendering instead of being viewed directly, we only compute the peak signal-to-noise ratio (PSNR) between the synthesized views using compressed depth map sequences and the synthesized views using uncompressed depth map sequences. The experimental results are presented in Tables 6, 7, 8, and 9 in which coding efficiency is measured with rendered PSNR and total bitrate (depth map), and computational complexity is measured with the consumed coding time. Since the proposed approaches affect only depth map intra coding, results for texture video coding are identical; thus, the texture video results are not included in the table. The Bjontegaard Delta PSNR (BDPSNR) [21] represents the average texture quality for synthesized views PSNR gain, Bjontegaard Delta Bitrate (BDBR) represents the improvement of total bitrates for depth map coding, and 'Dtime (%)' represents the entire depth map coding time change in percentage.

4.1 Individual performance results of the proposed algorithms

Tables 6 and 7 give individual evaluation results of the proposed algorithms compared with the original 3D-HEVC encoder (Table 6 in two-view case, Table 7 in three-view case), i.e., early termination mode decision (ETMD), adaptive search range motion estimation (ASRME), and fast disparity estimation (FDE), when they are applied alone. The proposed three algorithms can greatly reduce the encoding time with similar encoding efficiency for all sequences. For the ETMD algorithm, about 18.9% and 19.9% coding time has been reduced in the two-view and three-view conditions, respectively, with the highest gain of 32.7% in 'Poznan_Hall2' (two-view case) and the lowest gain of 7.6% in 'GT_Fly' (three-view case). It can be also observed that a consistent gain is obtained over all sequences under both conditions. The average PSNR drop for all the test sequences is 0.01 dB, which is negligible, and bitrate has been reduced by 0.02% to 0.03% on average, which indicates that the proposed ETMD algorithm can improve the bitrate performance for depth map compression in 3D-HEVC. For the proposed ASRME algorithm, 32.4% encoding time has been reduced with the maximum of 64.5% in two-view and three-view conditions. The coding efficiency loss is very negligible with a 0.02-dB PSNR drop or 0.11% to 0.13% bitrate increase. This result indicates that ASRME can efficiently skip unnecessary ME search

range computation in 3D-HEVC depth map coding. As far as the FDE algorithm, 25.4% and 26.4% coding time has been reduced in the two-view and three-view conditions, respectively; the average PSNR drop for all the test sequences is 0.02 to 0.03 dB, and the average increase of bitrate is 0.27% to 0.3%, which is negligible. The foregoing result analysis indicates that the FDE algorithm can efficiently reduce unnecessary DE computation time while maintaining nearly similar coding efficiency as the original 3D-HEVC encoder.

4.2 Combined results

In the following, we will analyze the experimental result of the proposed overall algorithm, which incorporates ETMD, ASRME, and FDE. The comparison result of the overall algorithm is shown in Table 8. The proposed overall algorithm reduces 64.3% and 66.3% encoding time on average under the two-view and three-view case, respectively, and achieves the better gain in coding speed among all the test sequence approaches compared to 3D-HEVC. Also shown is a consistent gain in coding speed for depth map compression with a minimum of 47.3% in 'Kendo' (two-view case) and the maximum of 87.5% in 'GT_Fly' (two-view case). For sequences with ground-truth depth map like 'Shark,' 'Undo_Dancer', and 'GT_Fly', the proposed algorithm saves more than 80% coding time. The computation reduction is particularly high because the variable size ME and DE decision process of a significant number of depth map treeblocks are reasonably skipped. Meanwhile, the coding efficiency loss is negligible, where the average PSNR drop for the entire test sequences is 0.04 to 0.05 dB or the average increase of bitrate is 0.39% to 0.43%. Therefore, the proposed overall algorithm can reduce more than 64% depth map coding time with the same RD performance of the original 3D-HEVC encoder.

In addition to the 3D-HEVC encoder, we also compare the proposed overall algorithm with a state-of-the-art fast algorithm for 3D-HEVC (content adaptive complexity reduction scheme (CACRS) [16]) in Table 9. Compared with CACRS, the proposed overall algorithm performs better on all the sequences and achieves more than 24.5% coding time saving, with a minimum of 14.7% in 'Kendo' (two-view case) and the maximum of 38.6% in 'GT_Fly' (three-view case). Meanwhile, the proposed overall algorithm achieves a better depth map coding performance, with 0.05- to 0.06-dB PSNR increase or 0.95% to 1.21% bitrate decrease for all the test sequences compared to CACRS.

Zhang *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:2

Page 14 of 14

Figures 5 and 6 gives more detailed experiment results (RD and time-saving curves) of the proposed overall algorithm compared to 3D-HEVC in the three-view case. As shown in Figures 5 and 6, the proposed low-complexity depth map compression algorithm can achieve a consistent time saving over a large bitrate range with almost negligible loss in PSNR and increment in bitrate.

5 Conclusions

This paper presents a low-complexity depth map compression algorithm to reduce the computational complexity of the 3D-HEVC encoder by exploiting three fast approaches, i.e., early termination mode decision, adaptive search range motion estimation, and fast disparity estimation for depth map coding. The recent 3D-HEVC test model is applied to evaluate the proposed algorithm. The experimental results show that the proposed algorithm can significantly reduce the computational complexity of depth map compression and maintain almost the same RD performances as the 3D-HEVC encoder.

Author details

[1]College of Computer and Communication Engineering, Zhengzhou University of Light Industry, No. 5 Dongfeng Road, Zhengzhou 450002, China. [2]Software Engineering College, Zhengzhou University of Light Industry, Zhengzhou 450002, China.

References

1. GJ Sullivan, JM Boyce, C Ying, J-R Ohm, CA Segall, A Vetro, Standardized extensions of high efficiency video coding (HEVC). IEEE J. Sel. Top. Sign. Proces. **7**(6), 1001–1016 (2013)
2. GJ Sullivan, J-R Ohm, W-J Han, T Wiegand, Overview of the high efficiency video coding (HEVC) standard. IEEE Trans. Circuits Syst. Video Technol. **22**(12), 1649–1668 (2012)
3. K Müller, H Schwarz, D Marpe, C Bartnik, S Bosse, H Brust, T Hinz, H Lakshman, P Merkle, H Rhee, G Tech, M Winken, T Wiegand, 3D high efficiency video coding for multi-view video and depth data. IEEE Trans. Image Process. **22**(9), 3366–3378 (2013)
4. L Zhang, G Tech, K Wegner, S Yea, in *Joint Collaborative Team on 3D Video Coding Extensions (JCT-3V) Document JCT3V-E1005, 5th Meeting*, 3D-HEVC test model 5 (Vienna, Austria 2013).
5. H Oh, YS Ho, in *Proc. The Pacific-Rim Symposium on Image and Video Technology*, H.264-based depth map sequence coding using motion information of corresponding texture video (LNCS 2006), **4319**, 898–907
6. M Wang, X Jin, S Goto, in *Proc. 28th Picture Coding Symp.*, Difference detection based early mode termination for depth map coding in MVC (2010), pp. 502–505
7. S Tsang, Y Chan, W Siu, Efficient intra prediction algorithm for smooth regions in depth coding. Electron. Lett. **48**(18), 1117–1119 (2012)
8. G Cernigliaro, F Jaureguizar, J Cabrera, N García, Low complexity mode decision and motion estimation for H.264/AVC based depth maps encoding in free viewpoint video. IEEE Trans. Circuits Syst. Video Techn. **23**(5), 769–783 (2013)
9. M Maitrea, MN Do, Depth and depth–color coding using shape-adaptive wavelets. J. Vis. Commun. Image **21**(5–6), 513–522 (2010)
10. S Milani, P Zanuttigh, M Zamarin, S Forchhammer, in *Proc. IEEE Int. Conf. Multimedia and Expo(ICME)*, Efficient depth map compression exploiting segmented color data (Barcelona, 2011), pp.1–6
11. L Shen, P An, Z Liu, Z Zhang, Low complexity depth coding assisted by coding information from color video. IEEE Trans. Broadcasting **60**(1), 128–133 (2014)
12. Q Zhang, P An, Y Zhang, L Shen, Z Zhang, Low complexity multiview video plus depth coding. IEEE Trans. Consumer Electron. **57**(4), 1857–1865 (2011)
13. Z Gu, J Zheng, N Ling P Zhang, in *Proc. 2013 IEEE International Conference on Multimedia and Expo Workshops*, Fast depth modeling mode selection for 3D HEVC depth intra coding (2013), pp. 1–4
14. Y Song, Y Ho, in *Proc. IEEE 11th IVMSP Workshop*, Simplified inter-component depth modeling in 3D-HEVC, (2013), pp. 1–4
15. M Zhang, C Zhao, J Xu, H Bai, in *Proc. IEEE International Symposium on Circuits and Systems(ISCAS)*, A fast depth-map wedgelet partitioning scheme for intra prediction in 3D video coding (2013), pp.2852–2855
16. HRTohidypour, MT Pourazad, P Nasiopoulos, V Leung, in *Proc. 18th International Conference on Digital Signal Processing (DSP 2013)*, A Content Adaptive Complexity Reduction Scheme for HEVC-Based 3D Video Coding (2013), pp.1–5
17. M Winken, H Schwarz, T Wiegand, in *Proc. Picture Coding Symp.*, Motion vector inheritance for high efficiency 3D video plus depth coding (Krakow, Poland 2012), pp. 53–56
18. L Shen, Z Liu, T Yan, Z Zhang, P An, View-adaptive motion estimation and disparity estimation for low complexity multiview video coding. IEEE Trans. Circuits Syst. Video Technol. **20**(6), 925–930 (2010)
19. D Rusanovskyy, K Mueller, A Vetro, *Common test conditions of 3DV core experiments* (Joint Collaborative Team on 3D Video Coding Extensions (JCT-3V) document JCT3V-E1100, 5th Meeting, Vienna, AT, 2013)
20. M Tanimoto, T Fujii, K Suzuki, View Synthesis Algorithm in View Synthesis Reference Software 2.0 (VSRS 2.0), (Lausanne, Switzerland, ISO/IEC JTC1/ SC29/WG11 M16090, Feb. 2008).
21. G Bjontegaard, *Calculation of average PSNR difference between RD-curves* (13th VCEG-M33 Meeting, Austin, TX, 2001)