The following publication H. -B. Zhang, Y. -L. Chan, C. -H. Fu, S. -H. Tsang and W. -C. Siu, "Quadtree decision for depth intra coding in 3D-HEVC by good feature," 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 2016, pp. 1481-1485 is available at https://doi.org/10.1109/ICASSP.2016.7471923.

QUADTREE DECISION FOR DEPTH INTRA CODING IN 3D-HEVC BY GOOD FEATURE

Hong-Bin Zhang[1,2], *Yui-Lam Chan*[1] *, *Chang-Hong Fu*[2], *Sik-Ho Tsang*[1], *Wan-Chi Siu*[1]

[1]Centre for Signal Processing, Department of Electronic and Information Engineering,
The Hong Kong Polytechnic University, Hong Kong
[2] School of Electronic and Optical Engineering,
Nanjing University of Science and Technology, Nanjing, China

ABSTRACT

3D-HEVC is a good coding solution for multi-view video plus depth data. It achieves good coding performance of synthesized views. However, depth intra coding brings unbearable complexity, which is the most urgent issue to be solved for the practical applications. Typically, depth maps have a good feature of structure or less texture compared with natural videos. Therefore, in this paper, a fast depth intra coding algorithm is proposed to limit the quadtree structure by the good feature-corner point (CP). The proposed algorithm can adaptively extract CPs and pre-allocate the depth level of coding quadtree. The large size of coding units (CUs) can be skipped for blocks, which have higher predicted depth level. On the contrary, the blocks, with lower predicted depth level, do not check the smaller size of CUs. Simulation results show that the proposed algorithm can provide about 41% time reduction while maintaining the BD performance.

Index Terms—Multi-view video plus depth, depth map, 3D-HEVC, corner detection, coding unit decision

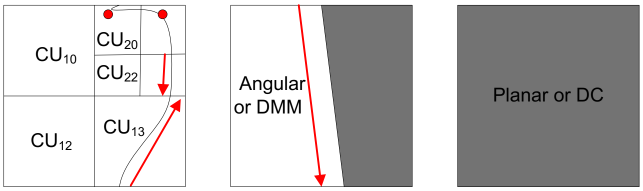
1. INTRODUCTION

After achieving a great success in the film industry, 3D video has gradually came to people's daily life and became an important status in multimedia. In this context, 3D family interactive entertainment is becoming increasingly popularization and has vast commercial potential, which has attracted more and more attention from both of industry and academia. One of the most important challenges in 3D video applications is to find a good compression solution for storing and transmitting the huge data. Therefore, the Joint Collaborative Team (JCT), specifically, established a working group called JCT-3V to charge with the standardization of 3D video [1]. More specifically, this group integrated 3D videos into the current HEVC coding framework from then on, developing the 3D video coding standard, 3D-HEVC [1]. Unlike the other extension standards, for example MV-HEVC (multi-view video coding), 3D-HEVC had a greatly improvement and made a dramatic change compared with the naked HEVC in spite of still inheriting of coding structure [2]. The purpose of these major modifications takes advantage of the new geometric information from the depth map. It is an auxiliary data for rendering virtual views via depth image based rendering technique in 3D-HEVC as compared with the conventional multi-view video format specified in the early 3D-video system [3]-[6]. The special characteristic of depth maps is very different from natural videos, which has very limited texture inside an object and discontinuity sharp edges at object boundaries [7]. All the newly added methods, including depth modelling mode, segment-wise DC coding, single depth mode (SDM), view synthesized optimization [8]-[9], etc., are inspired by the unique characteristic and their purposes are all to provide good quality of synthesized views [2].

However, another issue, which becomes important for depth coding, is to reduce the high complexity due to checking the huge amount of mode candidates and the hierarchy quadtree structure by evaluating RD cost function. Since depth map has a good structure [10], which has widely applications in computer vision, such as face recognition, gesture recognition, object detection, tracking and 3D shape retrieval [11]-[14]. However, in coding aspect, previous works just designed some coding tools to better represent the special signals. To the best of our knowledge, no one has made an attempt to speedup depth intra coding using its excellent structural information. Therefore, in this paper, a quadtree depth limited strategy is proposed by taking advantage of corner point (CP), in which the depth level of quadtree is pre-allocated and depth search range is adaptively reduced according to the predicted depth levels. Consequently, some redundant depth levels can be eliminated and simulation results show the propose algorithm can effectively reduce the complexity on the premise of guaranteeing coding quality.

The rest of this paper is organized as follows. The complexity of depth intra coding in 3D-HEVC and some related works are described in Section 2. This section also presents the review of corner detection and the relationship between intra coding and CP. In Section 3, the proposed quadtree decision algorithm is described. It includes

This work is supported in part by National Natural Science Foundation of China (No. 61301109), and a grant from the Research Grants Council of the HKSAR, China (Grant No. PolyU 5119/13E)



(a) (b) (c)

Fig. 1 (a) Example of LCU partition with corner point, (b) intra prediction using angular mode or DMM, (c) intra prediction using Planar or DC mode.

adaptive C selection, pre-allocated quadtree structure and depth range limited strategy. Finally, simulation results and some conclusions are given in Section 4 and 5, respectively.

2. CORNER POINT AND DEPTH INTRA CODING

The two major reasons for high complexity of depth intra coding is that 3D-HEVC introduces (i) a huge number of intra mode candidates, and (ii) quadtree coding structure. Some pervious works have been proposed to reduce the number of mode candidates, including DMM [15]-[17], HEVC intra mode [18]-[19], and SDC [20] for a given prediction unit. However, if the quadtree coding structure can be determined in advance, all the mode candidates can be skipped. As a result, a lot of encoding time can be saved. Therefore, Mora *et al.* [21] exploited texture-depth redundancies and developed an inter-component coding tools, in which the quadtree structure of depth map is limited to that of texture image. But, the correlation between texture and depth quadtree is very low in I-frames. In addition, Miok *et al.* designed a quadtree pruning algorithm using the variance and estimated distortion of SDM [22]. Tsang *et al.* reduced the complexity of mode decision for smooth regions [23]. However, the time consumption of depth intra coding is still intolerable.

Intrinsically, depth map has the superior feature of structure, steadily and precisely represented by CPs. There are so many corner detection algorithms, and the classical one is Harris [24] and its variants. The basic principle of the corner detection is to find the interest points with relative large energy in two directions around its neighborhood. In Harris, an image I(x,y) can be regarded as multiple random variables and two principle directions can be measured by the classical Principal Component Analysis (PCA) method. The covariance matrix of sample p(x,y) and its neighboring random variance is given by

$$M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \tag{1}$$

where w(x,y) is the weight function, and I_x and I_y are the horizontal and vertical gradients, respectively. Based on the PCA theory, the two eigenvectors of M (quadric form) are the two principal directions and their eigenvalues can reflect the degrees of change in their directions. Since depth intra modes can only predict smooth regions by planar or DC or regions with explicit direction by angular modes or DMM as shown in Fig. 1(b) and Fig. 1(c), respectively. It can be assumed that if one or more CPs are located in a given CU, only a single intra mode cannot describe this CU and it should be partitioned into four equally smaller-size CUs. Therefore, CP is a good feature to find the multi-directions and determinate whether a single mode can give a receivable predicted error and whether the current CU should be decomposed or not.

3. PROPOSED QUADTREE DECISION METHOD

Fig. 2(a) shows the distribution of CPs for view 3 in kendo#1, which are detected by the OpenCV function Goodfeaturestotracking()[10]. For comparison, Fig. 2(b) gives the optimal quadtree structure by original method in 3D-HEVC. Intuitively, the more CPs distributes, the more depth level of the quadtree is, which can verify our assumption that the CPs and the quadtree structure have very high correlation in depth intra coding.

3.1. Selection of Corner Points

Given an image, Goodfeaturestotracking() algorithm is employed to find all CPs by a relative small parameter 0.0001 empirically. Noted that in this OpenCV function, if the minimal eigenvalue is greater than that parameter, this sample should be considered as a CPs. Besides, all the CPs are sorted by the eigenvalues of their covariance matrix M. Then according to quantization parameter (QP), we select the number of CPs for further utilization with maximum eigenvalues by

$$F_{Num} = \begin{cases} T_{Num} & , QP \le 36 \\ T_{Num} \times \dfrac{5-((QP-37)\%3)}{3 \times 2^{([((QP-37)/3)]+1)}} & , \text{Otherwise} \end{cases} \tag{2}$$

where T_{Num} is the total number of CPs detected by the parameter 0.0001, F_{Num} is the number of CPs for further process, the operator % is the modular arithmetic and [.] is the round function. By this way, we can adaptively choose CPs that are applied to quadtree depth limited strategy according to QP.

3.2 Pre-allocation of quadtree structure

After obtaining CPs, all the LCUs are assigned the quadtree structure according to the locations of CPs as shown in Fig. 2(c). More specifically, an instance with the red rectangle in Fig. 2 is further illustrated in Fig. 3. Since this LCU has a

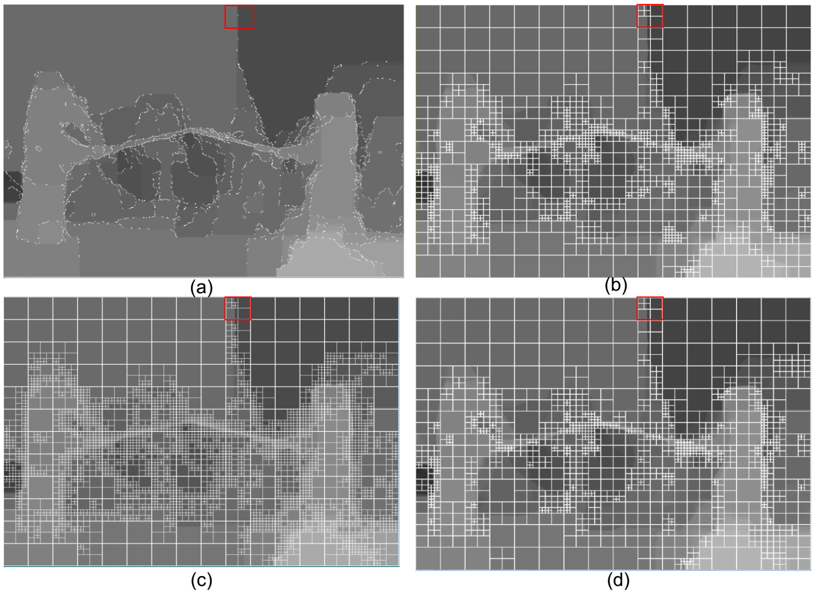


Fig. 2 (a) Corner points (white points) in view 3 of Kendo#1, (b) coding quadtree structure by 3D-HEVC, (c) pre-allocation coding quadtree by corner points, and (d) coding quadtree structure by the proposed algorithm.

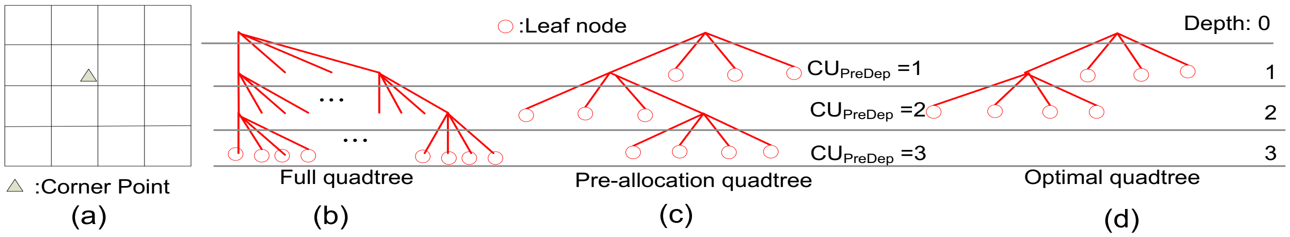


Fig.3 Instance of pre-allocation coding quadtree according to corner point. (a) LCU with one corner point, (b) full quadtree in original method, (c) pre-allocation coding quadtree, (d) the optimal quadtree structure by full search.

CP, it should be split into four CUs, labelled as CU_{10}, CU_{11}, CU_{12}, and CU_{13}. Here, CU_{ij} denotes j^{th} CU in depth level i by the zig-zag scanning order. In the next level, the CP is only located at the CU_{10}. Therefore, only CU_{10} needs to be partitioned continually, and the others should be regarded as the leaves nodes of the quadtree and their pre-allocated depth levels (CU_{PreDep}) are equal to 1 as shown in Fig. 3(c). In the depth level 2 of quadtree, CU_{20}, CU_{21}, CU_{22} and CU_{23} can be obtained from the segmentation of CU_{10}. Similar as the depth level 0 and 1, only CU_{23} with the CPs should be divided into CU_{312}, CU_{313}, CU_{314} and CU_{315} and their CU_{PreDep} are 3. The decomposition of the others CU_{2j} should be stopped and their CU_{PreDep} equals to 2. Finally, we obtain the pre-allocation quadtree (or CU_{PreDep}). On the other hand, if there is no CP in a given LCU, the CU_{PreDep} of this LCU is equal to 0. Here, we define CU_{PreMax} is the maximum depth level of the pre-allocation quadtree. CU_{PreMax} is equal to 0 for N-CP LCU and 3 for CP LCU.

3.3 Proposed quadtree depth limited algorithm

In this section, we firstly exploit the relationship between CU_{PreDep} and the optimal coding quadtree (CU_{FinDep}) as tabulated in Table 1, which is generated from *Kendo* under QP=34. It can be seen that the LCUs with CU_{PreMax} "0" have about 99.98% to be encoded by 64x64 and 32x32 size CUs in the N-CP case. In addition, when CU_{PreDep} is "3" or "2", mostly (94.23% or 89.14%) LCUs select depth level "1", "2" and "3" to be the optimal depth level. Since the CU_{FinDep} is unevenly distributed according to CU_{PreDep}, some depth levels of quadtree can be skipped for time reduction and the proposed quadtree depth limited strategy is summarized as the following pseudo-code in **Algorithm**.

Given a LCU and its CU_{PreDep}, if CU_{PreMax} is equal to 0 (N-CP), the depth levels 2 and 3 for rich texture will be skipped. Otherwise, if CU_{PreMax} is equal to 3 (CP), we skip the depth levels "2" and "3" for CUs with CU_{PreDep} "1" and the depth level "0" for CUs with CU_{PreDep} "2" or "3". However, since the CUs with CU_{PreDep} "3" have about 6% to be coded by depth level "0" and this percentage is 10% for CU_{PreDep} "2", the prediction of depth level is not always exact and a remedy is proposed to check planar, DC, and SDM for depth level "0" when the condition, **c1,** is satisfied. Here, **c1** is that the CU_{FinDep} is 1 and all the four 32x32 CUs select SDM, planar, or DC to be the optimal mode. Based on the statistical analysis of the predicted CU level and optimal

Table 1 Statistical analysis of predicted CU level for CP and non-CP cases for kendo sequence

LCU without corner point (CU_{PreMax}=0) (%)				
[1]P\F	0(64x64)	1(32x32)	2(16x16)	3(8x8)
0	57.16	42.82	0.02	0
1,2,3	-	-	-	-
LCU with corner point (CU_{PreMax}=3) (%)				
[1]P\F	0(64x64)	1(32x32)	2(16x16)	3(8x8)
0	-	-	-	-
1	21.99	75.28	2.47	0.24
2	10.86	47.55	40.16	1.42
3	5.77	31.22	35.45	27.55

[1] Noted that P is CU_{PreDep} and F denotes CU_{FinDep}

Algorithm Quadtree Depth Limited Strategy

```
Input: LCU and its $CU_{PreDep}$ and $CU_{PreMax}$
for depth = 0 to 3 do
    if($CU_{PreMax}$==0&&depth level == 2,3)
    skip the mode decision process of current depth level;
    depth ++;
    else if($CU_{PreMax}$==3&& $CU_{PreDep}$ ==1&&depth level == 2,3)
    Skip the mode decision process of current depth;
    depth++;
    else if($CU_{PreMax}$==3&& $CU_{PreDep}$ ==2,3&&depth level == 0)
    Skip the mode decision process of current depth level;
    depth++;
    else
    Check all the intra modes of current depth level ;
    end if
end for
Remedy:  for corner point case to check depth level of 0
if ($CU_{PreMax}$==3&&optimal depth level ==1&&c1)
   Check SDM, planar and DC of depth level 0;
end if
Output: The optimal quadtree structure
```

CU level, the proposed algorithm can adaptively select the search range of depth level. Compared with the final quadtree from the original and proposed methods in Fig. 2(b) and (d), our method can exactly predict the depth level for most CUs.

4. SIMULATION RESULTS

For evaluation and validation of the proposed quadtree depth limited strategy, we integrated it into the 3D-HEVC test model [25] of HTM-13.0 [26]. We followed the Common test conditions (CTC) to set the quantization parameters (QP (Texture, depth) {(25, 34), (30, 39), (35, 42), {40, 45}}). The all intra configuration were used and view synthesized optimization was enabled. Seven various sequences VGA (1024x768) and 1080i (1920x1088) recommended by CTC were encoded. For each sequence, 200~250 frames were encoded. The simulations were conducted on the platform of which specifications are the CPU of Intel Xeon(R) E3-1230 @3.3GHz and RAM 16.0GB. The performance is evaluated based on the difference of coding time (ΔT), the bitrate difference of the synthesized view (BDBR), which is calculated by the overall bitrate and PSNR of synthesized views.

Table 2 Coding performance of proposed quadtree depth limited algorithm.

Sequence	Miok's (%) [22]		Proposed (%)	
	BDBR	ΔT	BDBR	ΔT
Kendo	0.09	-24	0.29	-41
Balloons	0.05	-25	0.29	-39
Newspaper	0.10	-20	-0.23	-36
1024x768	0.08	-23	0.12	-39
Fly	0.10	-23	0.20	-45
Hall2	0.40	-47	0.33	-48
Street	0.28	-25	1.16	-40
Dancer	0.27	-37	1.01	-39
1092x1088	0.26	-33	0.68	-43
Average	0.18	-29	0.44	-41

Table 2 tabulates the coding performance of Miok's algorithm [22] and the proposed algorithm in terms of ΔT and BDBR compared with HTM-13.0. The algorithm in [22] provides a quadtree pruning strategy using a threshold method by jointly consideration of variance and estimated distortion of SDM. However, this method also should check the quadtree depth by depth similar as full search in Fig. 3(b). Some CUs with consistent orientation can be described by the angular mode, but they also have a very large variances. Therefore, it is very difficult to determinate the value of threshold in Miok's method, since variance cannot provide any directional information. Our method can provide the limited quadtree structure of both smooth or complexity regions adaptively according to CPs. As a result, the time saving of the proposed algorithm is 41% while that is only 29% in [22]. Besides, both of two algorithms can provide a reasonable BD performance. But, it is also demonstrated that the BD performance of *Dancer* and *Street* are of a little high in the proposed algorithm. The reason for this is that the depth level of scene is so high that the difference of depth values in distant scene is different to be detected by CPs in split of rich texture in these regions.

5. CONCLUSION

In this paper, we proposed a fast quadtree decision for depth intra coding by corner point. The proposed algorithm can provide a significant time reduction while maintain the BD performance. The proposed algorithm studies the correlations between intra modes and CPs to speed up intra mode decision in a given CU level. Considering the spatial discontinuity by CPs, our algorithm can preserve the sharp edges for key scenario or perceptible objects by using small CUs. The main contribution of this paper is that the good structure for depth map in computer version is brought into video coding. Actually, this has widespread applications prospect where further research works can be inspired by our algorithm, such as adaptive bit-allocation, region-of-interesting coding, and 3D shape coding for special applications.

6. REFERENCES

[1] "Call for Proposals on 3D Video Coding Technology," ISO/IEC/SC29/WG11/MPEG2011/N12036, Geneva, Switzerland, March, 2011.

[2] K. Muller, H. Schwarz, D. Marpe and C. Bartnik, "3D High-efficiency video coding for multi-view video and depth data," IEEE Transactions on Image Processing, Vol. 22, no. 9, pp. 3366-3378, Sep. 2013.

[3] A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard,"Proc. of the IEEE, vol. 99, pp. 626-656, Apr. 2011.

[4] Z.-P. Deng, Y.-L. Chan, K.-B. Jia, C.-H. Fu, and W.-C. Siu, "Fast Motion and Disparity Estimation with Adaptive Search Range Adjustment in Stereoscopic Video Coding," IEEE Trans. on Broadcasting, vol. 58, no. 1, pp. 24-33, Mar. 2012.

[5] Z.-P. Deng, Y.-L. Chan, K.-B. Jia, C.-H. Fu, and W.-C. Siu, "Fast Iterative Motion and Disparity Estimation Algorithm for Multiview Video Coding," in Proc. of 3DTVCON, Tampere, Finland, Jun. 2010, pp. 1-4.

[6] Z.-P. Deng, Y.-L. Chan, K.-B. Jia, C.-H. Fu, and W.- C. Siu, "Iterative Search Strategy with Selective Bidirectional Prediction for Low Complexity Multiview Video Coding," J. of Visual Commu. and Image Repres.,vol. 23, no. 3, pp. 522-534, Apr. 2012.

[7] K. Müller, P. Merkle, and T. Wiegand, "3D video formats and coding methods,"17th IEEE international conference on Image Processing, Hong Kong, Step. 2010, pp. 2389-2392.

[8] E.-H. Dou, Y.-L. Chan, K.-B. Jia, and W.-C. Siu, "View Synthesis Optimization based on Texture Smoothness for 3D-HEVC," in Proc. of the IEEE Int. Conf. on ICASSP, Brisbane, Australia, Apr. 19-24, 2015, pp. 1443-1447.

[9] H.-Dou, Y.-L. Chan, K.-B. Jia, and W.-C. Siu, "An adaptive segment-based view synthesis optimization method for 3D-HEVC," in proceedings of 2015APSIPA ASC, Hong Kong, Dec. 16-19, 2015, pp.

[10] Shi, J., and Tomasi. C, "Good features to track," in the proceedings of Computer Vision and Pattern Recognition (CVPR), Seattle, WA, pp. 593-600, June 1994.

[11] C. H. Xu, Y. H. Wang, T. N. Tan, and L. Quan, "Depth vs. intensity: which is more important for face recognition," in the proceedings of ICPR, pp. 342-345, Aug. 2004.

[12] W. Q. Li, Z. Y. Zhang, and Z. C. Liu, "Action recognition based on a bag of 3D points," in the proceeding of Computer Vision and Pattern Recognition Workshops (CVPRW), San Francisco, CA, pp. 9-14, June 2010.

[13] X. Bai, Z. Zhu, and L. Latecki, " 3D Shape Matching via Two Layer Coding," IEEE Transaction on Pattern Analysis and Matching Intelligence, Vol. pp, no. 99, pp.1, April 2015.

[14] Helten, T., M. Muller, H.-P. Seidel, and C. Theobalt, "Real-time Body Tracking with one Depth Camera and Inertial Sensors," in the proceedings of international conference on Computer vision (ICCV), Sydney, NSW, pp. 1105-1112, Dec. 2013.

[15] Z. Y. Gu, J. H. Zheng, N. Ling and P. Zhang, "Fast Depth Modeling Mode Selection For 3D HEVC Depth Intra Coding", IEEE international conference on ICMEW, San Jose, CA, July 2013, pp. 1-4.

[16] C.-H. Fu, H.-B. Zhang, W.-M. Su, S.-H. Tsang, and Y.-L. Chan, "Fast Wedgelet Pattern Decision for DMM in 3D-HEVC," 2015 IEEE international conference on Digital Signal Processing, Singapore, pp. 447-481

[17] Park. C, "Edge-Based Intra Mode Selection for Depth-Map Coding in 3D-HEVC," IEEE Transaction on Image Processing, Vol. 24, no. 1, Nov, 2014.

[18] H.-B. Zhang, C.-H. Fu, and W.-M. Su, S.-H. Tsang, Y.-L. Chan, "Adaptive fast intra mode decision of depth map coding by low complexity RD-Cost in 3D-HEVC," 2015 IEEE international conference on Digital Signal Processing, Singapore, pp. 487-491.

[19] H.-B. Zhang, C.-H. Fu, and W.-M. Su, S.-H. Tsang, W.-C. Siu, "Efficient Depth Intra Mode Decision by Reference Pixels Classification in 3D-HEVC," IEEE international conference on image processing, Quebec, Canada, Sept. 2015, pp.961-965.

[20] H.-B. Zhang, S.-H. Tsang, Y.-L. Chan, C.-H. Fu, and W.-M. Su, "Early Determination of Intra Mode and Segment-wise DC Coding for Depth Map Based on Hierarchical Coding Structure in 3D-HEVC," in Proceedings of 2015APSIPA ASC， Hong Kong, Dec. 16-19, 2015, pp.396-400.

[21] More. E. G, Jung. J, and *et al*.," Initialization, Limitation, and Predictive Coding of the Depth and Texture Quadtree in 3D-HEVC," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 24, No.9, pp 1554-1565, Sept. 2014.

[22] Miok K., Nam L., and Li S.: 'Fast single depth intra mode decision for depth map coding in 3D-HEVC', IEEE International Conference on ICMEW, 2015, p. 1-6.

[23] S. H. Tsang, Y. L. Chan, and W. C. Siu, "Efficient intra prediction algorithm for smooth regions in depth coding," Elect. Letters, Vol. 48, no. 18, pp. 1117-1119, Aug. 2012.

[24] Harris C., and Satephens M. J., "A combined corner and edge detector," in the proceeding of Alvey Vision Conference, Manchester, 1988, pp. 147-152.

[25] Y. Chen, G. Tech, K. Wegner, S. Yea, "Test Model 10 of 3D-HEVC and MV-HEVC," in ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, document JCT3V-J1003, 10th Meeting, Strasbourg, FR, 18–24 Oct. 2014.

[26] 3D-HEVC software HTM-13.0. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn 3DVCSoftware /tags/HTM-13.0.