Journal of Computers Vol. 30 No. 3, 2019, pp. 205-212
doi:10.3966/199115992019063003015

Efficient Predictive Structure Algorithm for MV-HEVC

Tao Yan[1*], In-Ho Ra[2], Lin-Yun Huang[1], Min-Hang Weng[1]

[1] School of Information Engineering, Putian University, Putian 351100, Fujian, China
{yantao, huanglinyun, wengminhang}@ptu.edu.cn

[2] School of Computer Information and Communication Engineering, Kunsan National University,
Gunsan 54150, South Korea
ihra@kunsan.ac.kr

Received 13 January 2019; Revised 9 March 2019; Accepted 12 March 2019

Abstract. Efficient prediction structure for multi-view high efficiency video coding (MV-HEVC) is a key factor in determining performance such as compression efficiency, random access, and scalability. At present, MV-HEVC reference software provided by the international standard organization JCT-3V adopts a fixed inter-view prediction structure, which makes it difficult to adaptively adjust the prediction structure. Therefore, this paper proposes a novel predictive structure algorithm based on multi-objective optimization for MV-HEVC. This paper, the design of MV-HEVC prediction structure is regarded as a multi-objective optimization problem. The idea of this paper is to first determine the position of the I-view based on the similarity analysis between the viewpoints. Then, considering the coding efficiency and the random access of the user, the multi-objective optimization function is established to solve the coded B-viewpoint between the I-viewpoint and the P-viewpoint. In the end, this paper flexibly adjusts the inter-view prediction structure according to the prediction structure coding parameters to improve the coding performance. Experiments show that compared with MV-HEVC, the proposed algorithm not only improves the coding efficiency but also has better the random access performance.

Keywords: multi-objective optimization, multi-view high efficiency video coding, prediction structure, random access, similarity analysis

1 Introduction

Multiview Video (MV) is a new type of video with three-dimensional stereoscopic and viewpoint interaction functions, which can meet the user's ability to select and operate audiovisual objects from multiple angles, and provides 360-degree scene roaming interactive capability. However, the compression efficiency of existing coding standards is still insufficient for high-definition, ultra-high-definition video applications, which still requires a more efficient coding compression scheme. To this end, ITU-T and Moving Picture Experts Group (MPEG) established the Joint Collaborative Team on Video Coding (JCT-VC). In 2013, the first generation of High Efficiency Video Coding (HEVC) standard was completed [1]. As one of the new 3D standards based on HEVC, MV-HEVC not only has strong stereoscopic effect and flexible interaction ability, but also vividly presents video scenes. At present, it has shown broad application prospects in the fields of 3D television (TV) and video conferencing [2-3], which has become an international research hotspot in the video field.

In the early research of Multiview Video Coding (MVC) standard algorithm, MVC prediction structure is also a very important aspect, and many scholars have conducted in-depth research [4-6]. However, there are relatively few studies on the prediction structure of the new 3D standard MV-HEVC. At present, most algorithms perform single-objective optimization for a certain encoding performance (compression efficiency, encoding complexity, random access performance, etc.) of multi-view video coding. Zhang et al. proposed a fast mode selection algorithm using the correlation between viewpoints

[*] Corresponding Author

[7]. Lei et al. proposed multi-view video coding based on inter-view correlation [8]. This algorithm has very good coding performance, which not only improves peak signal-to-noise ratio (PSNR), but also reduces coding complexity. [9] uses the CU in the spatial and view direction reference frames to predict the rate distortion cost of the current CU, and terminates the mode selection process by setting the threshold. Zhang et al. proposed a merge-skip mode advance decision method to reduce the computational complexity by analyzing the spatial correlation of the prediction mode and the correlation between the viewpoints [10]. Zhang et al. [11] proposed a fast method for merging mode based on the correlation of inter-view coding modes, which can significantly reduce the computational complexity and has been included in the 3D-HEVC draft. However, this method is not sufficient. The use of correlation between viewpoints results in a very limited reduction in B-frame complexity. At the same time, since the method is not applicable to P-frames, the computational complexity of P-frames is still high. Park et al. [12] skip the traversal of the candidate mode by classifying the edge information of the transform unit in the transform domain. Gu et al. [13] prematurely terminate the traversal of the prediction mode by analyzing the characteristics of the rate error function of the traditional intra prediction mode and the depth model mode (DMM). [14] uses the correlation between spatial domain correlation and depth map and texture map to reduce the computational complexity of intra prediction.

In recent years, some scholars have also studied the new 3D standard MV-HEVC prediction structure. Feng et al. [15] proposed a novel MV-HEVC coding for depth-assisted motion vector prediction, but did not consider factors such as random access performance. In [16], a new 3D HEVC coding scheme is proposed, which improves the traditional view synthesis method, uses optical flow technology to derive dense motion vector field (MVF) from adjacent views, and makes full use of parallax. Information to synthesize views and correlations between viewpoints for MV-HEVC encoding. Reference [17] proposed an HEVC inter-frame prediction optimization algorithm based on the adaptive coding unit access order, which further reduces the complexity of the HEVC coding step. De et al. [18] proposed the MV-HEVC predictive design optimization algorithm, but did not weigh the relationship between coding performance, high complexity, large computational complexity, and no operability in practical applications. Lee et al. [19] proposed an effective inter-view motion vector prediction algorithm, which uses polar line geometry, similarity transformation and natural transformation to derive geometric interrelationship between two adjacent views, and then predictively predict motion vectors (PMV), the algorithm is better than the traditional MV-HEVC, but the coding gain is only about 1.19%.

The above research has not thoroughly considered the relationship between MV-HEVC prediction structure and related coding performance. This paper regards the design of MV-HEVC prediction structure as a multi-objective optimization problem, which needs to be based on the characteristics of video content and the needs of specific applications. Adapting and adjusting the MV-HEVC prediction structure, under the given constraints, the optimal balance between coding performance and coding complexity is achieved. In this paper, considering the factors such as random access of user and coding efficiency, the inter-view prediction structure is flexibly adjusted according to the multi-view video correlation analysis to obtain better coding performance. The experimental results show that compared with the reference software MV-HEVC provided by JVT-3V, the random access performance is better, and the PSNR is improved.

2 Design Predictive Structure for MV-HEVC

Although compression efficiency has always been the primary indicator of MV-HEVC, it is not the only coding performance indicator, and there are many important performance indicators, including random access, coding complexity, backward compatibility, parallel processing, etc. Depends on the prediction structure used. Moreover, these goals are often contradictory, and put forward higher requirements for the design of MV-HEVC prediction structure.

In MV-HEVC research and application, the prediction structure is directly related to MV-HEVC coding performance (including coding complexity, compression efficiency, random access performance, etc.). Therefore, determining the appropriate MV-HEVC prediction structure is particularly critical. In MV-HEVC coding, key pictures of the I, B, and P frame-encoded viewpoints are referred to as I-viewpoints, B-viewpoints, and P-viewpoints. Before determining the MV-HEVC inter-view prediction structure, only two aspects need to be determined: determining the position of the I-viewpoint and the number of inserted B-viewpoints between the P-viewpoint and the I-viewpoint.

The idea of this paper is to first determine the location of the I-view based on the similarity analysis between the viewpoints, and then comprehensively consider the coding efficiency and the random access of the user to establish a multi-objective optimization function to solve the coded B-viewpoint between the I-viewpoint and the P-viewpoint. Flexible adjustment of the inter-view prediction structure can much improve the coding performance.

2.1 Determine the Position of I-viewpoint

The I-viewpoint is called the main viewpoint. As the reference viewpoint of other viewpoints, selecting the I-viewpoint can improve the coding efficiency. Therefore, some scholars have made in-depth research on the selection of the I-viewpoint. Park et al. [6] determined the position of the I-view by calculating the average global disparity of each viewpoint. The calculation method is obtained by equation (1):

$$\overline{D(V_i)} = \frac{1}{N-1}\sum_{j=0,j\neq i}^{N-1} \left|g(V_i,V_j)\right| \tag{1}$$

Where N represents the number of encoded viewpoints, $g(V_i,V_j)$ represents the global disparity between V_i the viewpoint and V_j the viewpoint, and $\overline{D(V_i)}$ represents the average of the global disparity between V_i the viewpoint and the other *N-1* viewpoints.

This method has two shortcomings. Firstly, it is determined by the global disparity that the I-viewpoint is not accurate. Secondly, the global disparity between the viewpoints is obtained, and the disparity estimation is needed. The calculation amount of this method is very large, especially when the coding viewpoint is large.

In this paper, the bilinear similarity measure algorithm is used to test the correlation between viewpoints, find the viewpoints most relevant to other viewpoints, and use this viewpoint as the I-viewpoint. The bilinear similarity measure algorithm algorithm has been successfully used in the field of image retrieval. This algorithm is superior to the traditional distance measure method, and there are no restrictions such as distance measurement. $S(\boldsymbol{V_j}, \boldsymbol{V_k})$ represents the viewpoint similarity between $\boldsymbol{V_j}$ and $\boldsymbol{V_k}$, $\mathbf{E}_j, \mathbf{E}_k$ are respectively two viewpoint feature vectors, and $\boldsymbol{T}$ is the matrix to be studied in this paper. The degree of similarity between viewpoints is calculated by the formula (2):

$$S(V_j,V_k) = \mathbf{E}_j^T \mathbf{T} \mathbf{E}_k \tag{2}$$

Then the structural similarity of S_V viewpoint and other *N-1* viewpoint S_W (*v, w*) is $\overline{S_v}$ and it is as follows:

$$\overline{S_v} = \frac{\sum_{w=0,w\neq v}^{N-1} S(V_w,V_v)}{N-1} \tag{3}$$

In summary, the position of the I-viewpoint $pos(k)$ is the view with the highest structural similarity:

$$pos(k) = \arg\max\{\overline{S_v},\ v = 0,1,\cdots,N-1\} \tag{4}$$

2.2 Determine the Number of B-viewpoint

In general, B-viewpoint coding has higher coding efficiency than I-viewpoint and P-viewpoint. In order to improve the efficiency of MV-HEVC coding, we should use B-viewpoint coding as much as possible. However, when the number of inserted B-viewpoints increases, the global disparity between the I-viewpoint and the P-viewpoint and the P-viewpoint increases, so that the disparity estimation is very inaccurate when encoding the P-viewpoint. Moreover, as the number of inserted B-viewpoints increases, the ratio of bidirectional predictive coding between views is higher, resulting in a significant increase in coding complexity and a very high random access cost. Therefore, further research is needed to determine the number of inserted B-viewpoints. This paper translates the problem of solving the number of B-viewpoints into solving multi-objective optimization problems.

This paper first determines the inserted B-viewpoint maximum. Through a large number of experimental tests, we can determine the maximum value of the inserted B-viewpoint between P-viewpoints by the global disparity between viewpoints and the inter-view correlation. Suppose $\overline{C}$ is the average of the correlation between two adjacent viewpoints at the same time, the parallax estimation search range is ***R***, the average parallax between the viewpoints is ***d***, and $MaxNub_B$ which is the maximum number of inserted B-viewpoints is obtained by equation (5):

$$MaxNub_B = Max\left\{0, \left\lfloor \log_2(\overline{C} \times {}^{R}\!/_{d}) \right\rfloor \right\} \tag{5}$$

where,

$$\overline{C} = \frac{1}{N-1} \cdot \sum_{j=0, j\neq k}^{N-1} \left| C(V_j, V_k) \right| \tag{6}$$

Here $C(V_j, V_k)$ represents the dependence of the viewpoint V_j and V_k. We use the correlation analysis function to test the correlation of the two viewpoints. Let P_1 and P_2 respectively represent two viewpoints V_j and V_k, two images at the same time. $P_1(i,j)$, $P_2(i,j)$ respectively represent the pixel values of the P_1, P_2 image at position (i,j), and I and J respectively represent the height and width of the image P_1 or P_2. $\overline{P_1}$, $\overline{P_2}$ represent the average pixel value of the P_1, P_2 image, respectively. The correlation between two viewpoints V_j and V_k can be obtained from the correlation of two images P_1, P_2, calculated by (7)

$$C(V_i, V_j) = C(P_1, P_2) = \left| \frac{\sum_{i=0}^{I-1}\sum_{j=0}^{J-1} \left|P_1(i,j) - \overline{P_1}\right| \times \left|P_2(i,j) - \overline{P_2}\right|}{\sqrt{\sum_{i=0}^{I-1}\sum_{j=0}^{J-1} \left|P_1(i,j) - \overline{P_1}\right|^2} \times \sqrt{\sum_{i=0}^{I-1}\sum_{j=0}^{J-1} \left|P_2(i,j) - \overline{P_2}\right|^2}} \right| \tag{7}$$

However, the obtained $MaxNub_B$ is the maximum value that can be inserted into the B-viewpoint, and is not the optimum value for inserting the B-viewpoint. To this end, this paper considers the coding efficiency and random access of user factors, and finds the best value of the inserted B-viewpoint in the value range $0, \cdots, MaxNub_B$.

This paper first explores the relationship between MV-HEVC prediction structure and compression efficiency, random access performance, etc., and then constructs mathematical models between each other. Finally, comprehensive random access of user and coding efficiency are considered to establish multi-objective optimization model. Since the inserted B-viewpoint generally does not exceed 5, this paper uses the exhaustive method to pre-code several group of pictures (GOPs) to find an optimal value for inserting the B-viewpoint in the value range $0, \cdots, MaxNub_B$.

We consider the factors such as coding efficiency and random access by users to determine the number of B-viewpoints. Suppose that the multi-view video used for encoding has N viewpoints, and each viewpoint has M frames. The number of encoded frames $S = N \times M$ is limited, Let $V_n (n = 0,1,\cdots,N-1)$ be the nth viewpoint, $P(V_n)$ is the probability that the randomly accesses of user of the nth viewpoint, $V_{n,m}(n = 0,1,\cdots,n-1, m = 1,2,\cdots,m)$ represents the mth frame of the nth viewpoint, $X_{n,m}$ represents the number of frames needed before decoding the $V_{n,m}$ frame, $P(V_{n,m})$ is the probability of random access of user to frame $V_{n,m}$, $P(V_{n,m}|V_n)$ is the conditional probability that the user randomly accesses the mth frame of the nth viewpoint $V_{n,m}$. The mathematical expectation of random access cost is as follows:

$$E(X) = \sum_{n=1}^{N}\sum_{m=1}^{M} P(V_{n,m}) X_{n,m} = \sum_{n=1}^{N}\sum_{m=1}^{M} P(V_{n,m} \mid V_n) P(V_n) X_{n,m} \tag{8}$$

In order to obtain superior coding performance, the number of inserted B-viewpoints N_B should satisfy (9) and (10).

$$\min\ E(X) = f(N_B) \tag{9}$$

$$\max\ PSNR = \eta(N_B) \tag{10}$$

Where E(X) is the random access cost function and PSNR is the rate distortion function, which are all related to NB, and then the optimization problem model is:

$$\min\ Y = \lambda f(N_B) - \eta(N_B) \tag{11}$$

Where λ is the adjustment parameter. N_B meets the conditions (12)

$$0 \le N_B \le MaxNub_B \tag{12}$$

Fig. 1 shows the flow chart of the proposed algorithm for finding the best N_B. N_B is from 0 to $MaxNub_B$. For each given N_B, the N-frame image between the inter-viewpoints can get Y, and the number of inserted B-viewpoints which is N_B is determined by comparing the size of Y.

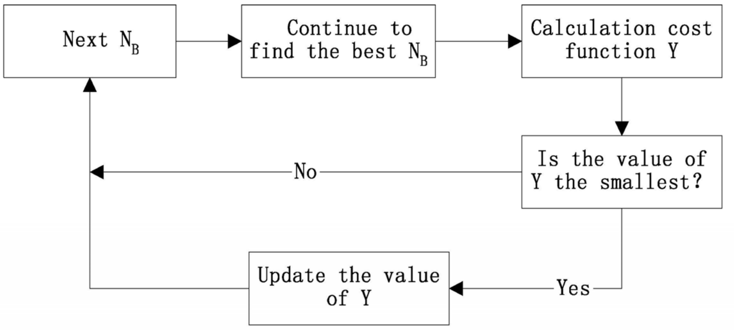
Next N_B
Continue to find the best N_B
Calculation cost function Y
Is the value of Y the smallest?
No
Yes
Update the value of Y

Fig. 1. Flow chart of the proposed algorithm for finding the best N_B

3 Experiment Results

In order to verify that the inter-view prediction structure algorithm proposed in this paper has good coding efficiency and random access performance, we use the test sequences provided by KDDI, MERL, Nagoya University/Tanimoto Lab and other research institutes GT_Fly, Ballroom, Undo_Dancer, Flamenco2, Band06, Exit for experimental analysis. Among them, the Exit sequence camera has a relatively large spacing, and the video content movement is relatively intense; while the GT_Fly sequence camera has a relatively small spacing, and the video content movement is relatively slow.

3.1 Rate Distortion Performance Comparison

In this paper, we compare five multi-view video coding schemes. The five coding methods are (1) Simulcast coding to represent independent coding between viewpoints, and there is no prediction between viewpoints; (2) Park et al. propose improved multi-view video coding of inter-view prediction structure; (3) Virtualization based on Zou et al. MV-HEVC coding for viewpoint synthesis prediction; (4) In 2018, Lee et al. based on polar line geometry, similar transformation and natural transformation MV-HEVC coding; (5) proposed based on inter-view similarity analysis and random access Inter-view prediction structure MV-HEVC coding.

Fig. 2 shows the average PSNR gain for the other four multi-view video coding methods over the Simulcast multi-view video coding method. Compared with the Simulcast multi-view video coding method, the other four methods have higher coding efficiency, and the PSNR gain is 0.18-1.6dB. The algorithm proposed by Zou et al. does not use the inter-view correlation to encode, and the average PSNR gain is 0.93 dB. Compared with the Simulcast multi-view video coding method, the PSNR gain of the prosed algorithm is 0.6-1.8dB. Compared with the coding algorithm proposed by Lei, the average PSNR gain of the prosed algorithm is 0.12dB. The main reason is that the proposed algorithm not only utilizes the correlation between viewpoints, but also focuses on the position of the I-viewpoint. The other reason is that for the slower motion sequence, the parallax is smaller, and thus more B-viewpoints can be used for encoding to improve the coding efficiency.

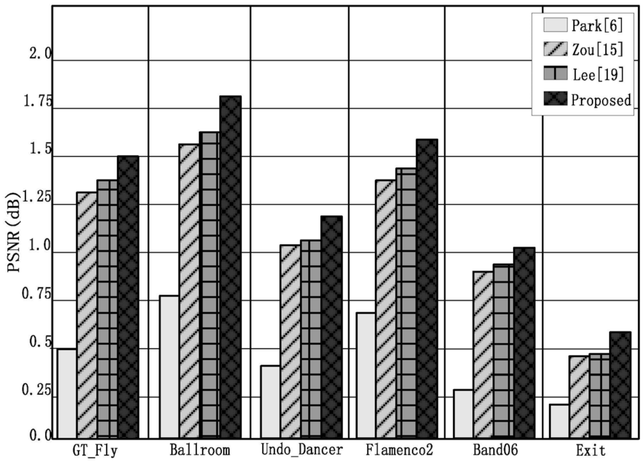

Fig. 2. PSNR experimental results

3.2 Random access Evaluation

Zou et al. [15] proposed the MV-HEVC coding algorithm based on virtual view synthesis prediction. Since there is no method for calculating the random access cost between virtual views, this paper only compares the random access cost between the above four multi-view video coding viewpoints. The random access cost of the predictive structure coding scheme is calculated using equation (10). Here, it is assumed that each coding scheme user randomly accesses each viewpoint and the probability of each frame is the same, so that the cost of randomly accessing each viewpoint is evaluated by the number of reference viewpoints required before the viewpoint decoding.

Fig. 3 is an experimental result of the average viewpoint random access cost for each scheme. As shown in Fig. 3, compared with Lee's proposed coding scheme, the proposed algorithm has better view random access performance, and the average view access cost is only about 70% of Lee's proposed algorithm. The main reason is that the proposed algorithm inter-view prediction structure I-viewpoint is close to the middle position, and the number of reference viewpoints required for random access to other viewpoints is small.

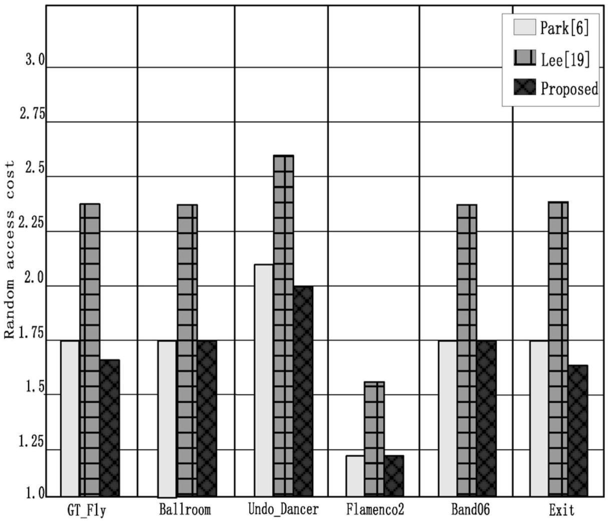

Fig. 3. Experimental result of the average viewpoint random access cost for each scheme

The experimental results of the above two aspects show that compared with the coding algorithm proposed by Lee, the new inter-view prediction structure algorithm proposed in this paper has higher coding efficiency and better random access performance.

4 Conclusion

In the research of video coding algorithms, prediction structure is always the most important aspect. In this paper, a new MV-HEVC inter-view prediction structure is proposed based on factors such as random access of user and coding efficiency. According to the characteristics of video content and the requirements of specific applications, this paper adaptively adjusts the MV-HEVC prediction structure. This paper first explores the relationship between MV-HEVC prediction structure and compression efficiency, random access performance, etc., and builds mathematical models between each other. Then the paper establishes a multi-objective optimization model by integrating random access and considering coding efficiency. In MV-HEVC research and application, the prediction structure is directly related to MV-HEVC coding performance (including coding complexity, compression efficiency, random access performance, etc.). For different multi-view video sequences, this paper adaptively adjusts the MV-HEVC prediction structure according to the inter-view similarity analysis. Considering the factors such as random access of user and coding efficiency, a new MV-HEVC inter-view prediction structure is proposed. The inter-view prediction structure can be flexibly adjusted according to different situations, and the random access performance is improved and the coding efficiency is improved. This paper does not consider coding complexity for the time being, and future work will be further studied in combination with other coding performance.

Acknowledgements

This work was supported by Natural Science Foundation of China (No. 61741111, Yan Tao). This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2016R1A2B4013002). This work was supported by Program for New Century Excellent Talents in Fujian Province University (2018, Yan Tao); in part by Natural Science Foundation of Jiangxi (No. 20181BAB202011, No. 20161BAB212031); in part by Natural Science Foundation of Fujian (No. 2019J01060722, Yan Tao).

References

[1] G.J. Sullivan, J.M. Boyce, Y. Chen, J.R. Ohm, Standardized extensions of high efficiency video coding (HEVC), IEEE Journal of Selected Topics in Signal Processing 7(6)(2013) 1001-1016.

[2] G. Tech, Y. Chen, K. Müller, Overview of the multiview and 3D extensions of high efficiency video coding, IEEE Transactions on Circuits and Systems Video Technology 26(1)(2016) 35-49.

[3] C.Y. Lin, Y. Zhao, J.M. Xiao, Region-based multiple description coding for multiview video plus depth video, IEEE Trans. on Multimedia 20(5)(2018) 1209-1223.

[4] P. Merkle, A. Smolic, K. Muller, Efficient prediction structures for multi-view video coding, IEEE Transaction On Circuits and Systems Video Technology 17(11)(2007) 1461-1673.

[5] H.Q. Zeng, X.L. Wang, C.H. Cai, J. Chen, Y. Zhang, Fast multiview video coding using adaptive prediction structure and hierarchical mode decision, IEEE Transactions on Circuits and Systems for Video Technology 24(9)(2014) 1566-1578.

[6] P.K. Park, K.J. Oh, Y.S. Ho, Efficient view-temporal prediction structures for multi-view video coding, Electronics Letters 44(2)(2008) 102-103.

[7] Q. Zhang, H. Chang, X. Huang, Adaptive early termination mode decision for MV-HEVC using inter-view and spatio-temporal correlations, AEU-International Journal of Electronics and Communications 70(5)(2016) 727-737.

[8] J.J. Lei, J.H. Duan, F. Wu, Fast mode decision based on grayscale similarity and Inter-view correlation for Depth map coding in 3D-HEVC, IEEE Trans. on Circuits and Systems Video Technology 28(3)(2018) 706-717.

[9] H.R. Tohidypour, M.T. Pourazad, P. Nasiopoulos, A content adaptive complexity reduction scheme for HEVC-based 3D video coding, in: Proc. 2013 International Conference on Digital Signal Processing, 2013.

[10] Q. Zhang, Q. Wu, X. Wang, Early SKIP mode decision for three- dimensional high efficiency video coding using spatial and interview correlations, SPIE Journal of Electronic Imaging 23(5)(2014) 53017-8.

[11] N. Zhang, D.B. Zhao, Y.W. Chen, J.L. Lin, W. Gao, Fast encoder decision for texture coding in 3D-HEVC, Signal Processing: Image Communication 29(9)(2014) 951-961.

[12] C. Park, Edge-based intra-mode selection for depth-map coding in 3D-HEVC, IEEE Transactions on Image Processing 24(1)(2014) 155-162.

[13] Z.Y. Gu, J.H. Zheng, N. Ling, P. Zhang, Fast bi-partition mode selection for 3D-HEVC depth intra coding, in: Proc. 2014 International Conference on Multimedia and Expo, 2014.

[14] Q. Zhang, N. Li, L. Huang, Effective early termination algorithm for depth map intra coding in 3D-HEVC, Electronics Letters 50(14)(2014) 994- 996.

[15] F. Zou, D. Tian, A. Vetro, View synthesis prediction in the 3D video coding extensions of AVC and HEVC, IEEE Transactions on Circuits and Systems Video Technology 24(10)(2014) 1696-1708.

[16] A.I. Purica, E.G. Mora, B.P. Popescu, M. Cagnazzo, B. Ionescu, Multiview plus depth video coding with temporal prediction view synthesis, IEEE Transactions on Circuits and Systems Video Technology 26(2)(2016) 360-374.

[17] Z. Ivan, G.B. Saverio, P. Eduardo, Inter-prediction optimizations for video coding using adaptive coding unit visiting order, IEEE Transactions on Multimedia 18(9)(2016) 1677-1690.

[18] A.A. De, P. Frossard, F. Pereira, Optimizing multiview video plus depth prediction structures for interactive multiview video streaming, IEEE Journal on Selected Topics in Signal Processing 9(3)(2015) 487-500.

[19] J.Y. Lee, J.K. Han, J.G. Kim, T.Q. Nguyen, Efficient inter-view motion vector prediction in multi-view HEVC, IEEE Transactions on Broadcasting 64(3)(2018) 666-680.