IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 9, 2019, accepted December 26, 2019, date of publication January 6, 2020, date of current version January 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2964056

Efficient Content Adaptive Plenoptic Video Coding

WANG TU[1], XIN JIN[2], (Senior Member, IEEE), LINGJUN LI[2], CHENGGANG YAN[1], YAOQI SUN[1], MANG XIAO[3], WEIDONG HAN[4], AND JIYONG ZHANG[1]

[1]Intelligent Information Processing Laboratory, Hangzhou Dianzi University, Hangzhou 310018, China
[2]Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China
[3]Department of Otolaryngology Head and Neck Surgery, Sir Run Run Shaw Hospital, Zhejiang University, Hangzhou 310058, China
[4]Department of Medical Oncology, College of Medicine, Sir Run Run Shaw Hospital, Zhejiang University, Hangzhou 310058, China

Corresponding author: Xin Jin (jin.xin@sz.tsinghua.edu.cn)

This work was supported in part by the Zhejiang Province Nature Science Foundation of China under Grant LR17F030006 and Grant Q19F010030, in part by the National Nature Science Foundation of China under Grant 61771275, Grant 61827804, Grant 61931008, Grant 61671196, Grant 61701149, Grant 61801157, Grant 61971268, Grant 61901145, Grant 61901150, and Grant 61972123, in part by the National Natural Science Major Foundation of Research Instrumentation of China under Grant 61427808, and in part by the 111 Project under Grant D17019.

ABSTRACT In this paper, a content adaptive coding method for plenoptic video is proposed to reduce both of the spatial and temporal redundancy. Based on the spatial correlations among subapertures, the plenoptic video is divided into two subaperture groups: the central view videos and the residual videos. Coding methods, multiview coding method based on spatial correlation(S-MVC) and multiview coding method based on temporal correlation(T-MVC), are adopted for residual videos according to correlation and residual energy analysis, while the central view videos are always encoded by S-MVC. Correlation analysis is performed first, and residual energy analysis is further performed if necessary. Coding the central view videos and the residual videos separately reduces bitstream, choosing different encoding methods for residual videos makes full use of the characteristics of different video content. Both the central view videos and the residual videos are compressed by MV-HEVC. The experimental results demonstrate that the proposed method outperforms a pseudo-sequence-based MVC method for plenoptic video by an average of 25.06% bitrate saving.

INDEX TERMS Plenoptic coding, adaptive, content-based, MVC, HEVC, MV-HEVC, subaperture image reordering.

I. INTRODUCTION

Plenoptic video is based on a basic concept, light field. Light field describes the distribution of light rays in free space. A 4D model proposed by Levoy and Hanrahan [1] and Gortler *et al*. [2], which measure the light rays at every possible location(x, y, z) from every possible angle θ, is now widely used to describe light field. Light field has been applied to many fields, such as light field microscope [3], super resolution application [4], lightweight head-mounted display [5], VR [6], etc. In recent years, more and more work focus on light field photography [7]. Portable plenoptic photographic devices have been made, which can record both spatial and angular light radiance in a single shot.

Compared with videos captured by traditional cameras, plenoptic videos captured by plenoptic photographic devices record not only intensity information of light rays from different direction but also the intensity variation in temporal domain. Therefore, plenoptic videos have huge data volume, which brings great difficulty to transmission and storage. In addition, complex subaperture relationship and huge data volume brings great challenges to compression. Thus, an efficient compression method designed for plenoptic video is highly desired.

The existing methods for plenoptic video coding can be classified into two categories: methods based on predictive coding and multiview-based methods. Many different methods based on predictive coding exist in the first category, they take advantage of different characteristics: displacement intra prediction mode [8], disparity-based compensation mode [9] and self-similarity compensated prediction [10]. These methods can be used for both plenoptic video and image coding, however, they cannot fully exploit the optical imaging correlations among micro-images. Some other work focused on the geometric relation among subaperture images and proposed homography transformation based methods [11]–[13].

The associate editor coordinating the review of this manuscript and approving it for publication was Xinfeng Zhang.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

These methods increase computational complexity, although they provide higher efficiency compared with the methods that coding plenoptic content directly using spatial coding tools in a video encoder like HEVC [14].

Many multiview-based methods [15]–[17] for plenoptic video compression in the second category have been proposed. In these methods, plenoptic video is decomposed into several viewpoint sequences, that is, a multiview video. Some methods focused on prediction structures. In a prediction structure [18], IPP or IBP structures are applied to each horizontal line and only the first central column of the views array. The number of available vertical inter-view predictions is very limited in such structures. In another structure [19]–[21], additional vertical inter-view prediction is introduced, views are divided into different types based on its references. Multiview video coding (MVC) can fully exploit both of the spatial correlation and temporal correlation among the subapertures. Thus, it improves the coding efficiency greatly but also increases huge computational complexity at the same time, especially when the number of views is large.

In this paper, a content adaptive plenoptic video compression method is proposed. In our method, subaperture projection-based grouping [22] is adopted. According to subaperture projection-based grouping [22], the plenoptic video is separated into two groups: the central view videos and the adjacent videos. The central view videos including a quite few of subaperture videos are considered as the key video and encoded with multiview coding method based on spatial correlation(S-MVC) first. Then, the residual videos are encoded with multiview coding method based on temporal correlation(T-MVC) or S-MVC, determined by correlation and residual energy analysis. The residual videos are generated by the difference between the reconstructed central subaperture videos and the adjacent subaperture videos. Both the central view videos and the residual videos are compressed by the multiview extension of high efficiency video coding (MV-HEVC) [23]. Subaperture projection-based grouping [22] has been proved effective for plenoptic image coding, it is applied to plenoptic video coding by combining with MVC for the first time in this paper. The total bitstream is reduced for the residual videos contain much less content than the central view videos. In our work, the proposed method exploits correlation and residual energy analysis for compression. Based on correlation and residual energy analysis of video containing different motion modes, the more efficient method from S-MVC and T-MVC is precisely chosen. In this method, several plenoptic images, instead of the entire plenoptic video, are selected randomly for correlation and residual energy calculation. The calculation process is very simple but effective. The proposed method fully removes inter-view and inter-frame redundancy with small computational overhead introduced. In pseudo-sequence-based methods, a pseudo-sequence is generated by reordering subaperture images with scanning topologies. Our experimental results demonstrate that the proposed method outperforms a pseudo-sequence-based [28] MVC method for plenoptic video by an average of 25.06% bitrate saving.

The remainder of this paper is organized as follows. Section II introduces the generation process of plenoptic video and two widely used methods for encoding plenoptic video. Content-based analysis of plenoptic video, the workflow and details of proposed plenoptic video compression method are introduced in Section III. In Section IV, we conducted experiment to explore its compression efficiency, we also did some comparative experimental analysis. Section V concludes the paper.

II. PLENOPTIC VIDEO AND EXISTING WORKS

A. PLENOPTIC VIDEO GENERATION

Plenoptic camera inserts a microlens array between the main lens and the sensor, allowing the sensor to capture intensity and angular information of the light rays. After passing through the main lens and microlens, the light rays in different direction from one object point are recorded by the sensor as a group of pixels [24], called macropixel or an elemental image (EI). The number of macro pixels is related to the number of lenses in the microlens array. Subaperture images are extracted from plenoptic image according to Ng's optical analysis [7] and light field decoding [25]. Repeating the extract process for each frame of plenoptic video, a fixed number of subaperture images are generated in each frame. In MVC, a subaperture image represents a view. Thus, a complete multiview video is generated by concatenating every view of the same position in each frame into a sequence respectively. In traditional MVC methods, the complete multiview plenoptic video is used as input for encoding directly. In this paper, several different ways for rearranging plenoptic video are used, more details are described in experimental part.

B. EXISTING PLENOPTIC VIDEO COMPRESSION METHODS

Two plenoptic video compression methods exist at present, we define them as S-MVC(multiview coding method based on spatial correlation) and T-MVC(multiview coding method based on temporal correlation), as depicted in Fig. 1 and

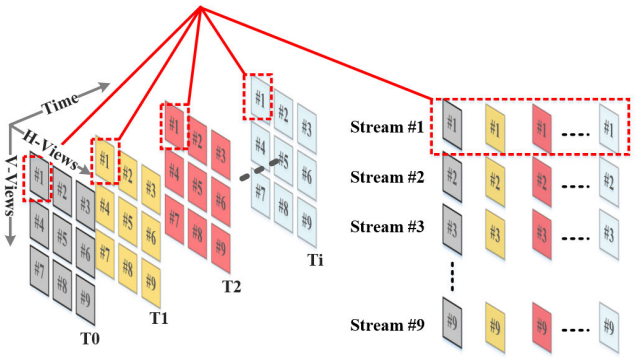


FIGURE 1. Plenoptic video compression method: S-MVC.

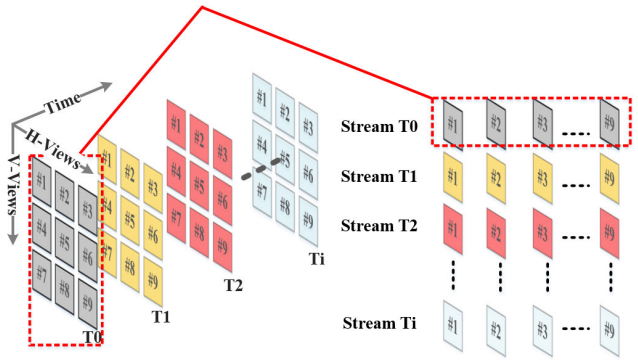


FIGURE 2. Plenoptic video compression method: T-MVC.

Fig. 2. Subaperture images in the same position in each frame constitute a multiview video in S-MVC, while all subaperture images in a frame constitute a multiview video in T-MVC. The multiview video is encoded as input by standard MVC. In terms of spatial domain, a problem exists in T-MVC is that coding efficiency begins to reduce significantly when the video content moves a lot. However, this problem has no effect on S-MVC, because there is only a difference in perspective between subaperture images, and the spatial correlation is relatively stable. It is this problem that makes S-MVC more efficient in most instances.

III. PROPOSED COMPRESSION METHOD

A. CONTENT BASED ANALYSIS

Subaperture projection-based grouping [22] is an efficient coding method for plenoptic image, making full use of the spatial correlation among subaperture images, which can be extended to apply into plenoptic video coding with MVC. Accordingly, the plenoptic video is divided into two subaperture groups: the central view videos and the residual videos. The residual videos contain much less content with lower correlation than the central view videos. Thus, more analyses are needed to determine which encoding method is more efficient for the residual videos. In order to determine the best coding method between S-MVC and T-MVC adaptively, the correlation and residual energy analyses are conducted before compression. The correlation between two subapertures can be calculated by:

$$r = \frac{\sum_m \sum_n (A_{m,n} - \bar{A})(B_{m,n} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{m,n} - \bar{A})^2)(\sum_m \sum_n (B_{m,n} - \bar{B})^2)}}. \quad (1)$$

where A and B denote two subaperture images. $\bar{A}$ and $\bar{B}$ denote the average of A and B respectively. The energy calculate algorithm is as follows:

$$E = (A - B)^2. \quad (2)$$

where E denotes energy between A and B, A and B denote two residual subapertures. We calculate the correlation between views and frames respectively according to (1). Taking a plenoptic video that containing 30 plenoptic images for example, there are 25 central views in each plenoptic image after grouping. Calculating the correlation between all views and frames will introduce a huge amount of computational complexity. In fact, we calculated the correlation between all views and frames of several different materials, finding that correlation between the views in each plenoptic image and correlation between frames in each view are very close. Therefore, selecting several plenoptic images or views randomly can be used as a representative.

The actual calculation process is as follows. First, for correlation calculation between views, 3 plenoptic images are chose randomly. In each plenoptic image, we calculate correlation between each view with its reference views according to the MVC prediction structure used in actual coding. Then, 25 averages are calculated in a single plenoptic image. We get an average of 25 averages for a single plenoptic image and get an average of 3 plenoptic images. Finally, we get an average that represents the correlation between views of 3 plenoptic images. For correlation calculation between frames, we choose 3 views from all 25 central views randomly and calculate correlation between each frame with its reference frames according to the GOP structure [23] used in actual coding. The follow steps are same with correlation calculation between views. Similarly, we get an average that represents the correlation between frames of 3 views. More details will be explained in the experimental part.

In order to analyze whether S-MVC is more efficient than T-MVC in most cases, we calculated the correlations of inter-views and inter-frames for the central view videos of several test materials by using the method mentioned above. All test materials are showed in Fig. 5. The results are showed in Table 1.

TABLE 1. Correlation of the central view videos.

Material/Correlation	Inter-frames	Inter-views
Auto Race	0.59888	**0.99180**
Flowers	0.21974	**0.93540**
Cube	0.63906	**0.97805**
Matryoshka	0.24596	**0.96270**
Toys	0.95775	**0.97626**
Trees	0.42273	**0.91894**

As showed in Table 1, we can notice that correlation of inter-views is generally higher than inter-frames for central view videos. Correlation of inter-views and inter-frames represent spatial correlation and temporal correlation respectively. According to this analysis result, we think S-MVC is more efficient for central view videos. Relevant experiments are implemented to prove in experimental part. In order to find out the better coding method for the residual videos, correlations of inter-views and inter-frames for residual videos are calculated. The results are showed in Table 2.

As showed in Table 2, not all correlations of inter-view are higher than inter-frame in these test materials. We came to realize that for the residual videos the correlation calculation

TABLE 2. Correlation of the residua videos.

Material/Correlation	Inter-frames	Inter-views
Auto Race	**0.35064**	0.52248
Flowers	**0.04890**	0.40318
Cube	0.62429	**0.43117**
Matryoshka	**0.13228**	0.44906
Toys	0.82355	**0.49869**
Trees	**0.08754**	0.54965

TABLE 3. Energy of the residual videos.

Material/Correlation	Inter-frames	Inter-views
Auto Race	**1203075**	1317459
Flowers	4632348	**4604420**
Cube	2658648	**2620865**
Matryoshka	**2013614**	2041065
Toys	**1875234**	1915060
Trees	2517490	**2508016**

may not be accurate. On one hand, the residual matrix contains more zero coefficients or coefficients close to zero. On the other hand, the correlation calculation algorithm starts to deteriorate when the difference between the frames is relatively large. The second problem also exists in central view videos, but it may not cause huge influence for it is the whole video. The inter-frames correlation of *Flowers*, *Matryoshka* and *Trees* are very small. According to Fig. 5, among all test materials, *Flowers*, *Matryoshka* and *Trees* have relatively large difference in temporal domain. The correlation value is not accurate for these test materials under this circumstance. Further analysis is needed to determine which coding method is better for the residual videos.

There is a commonly used method in video coding to measure the residue, which is to calculate the residual energy. As showed in Table 3, we adopt (2) to calculate the energy of residue with the same method described in correlation calculation. Theoretically speaking, the higher the correlation, the smaller the residual energy. The results of Table 2 and Table 3 are perfect match except for *Flowers*, *Cube* and *Trees*. So, we think the correlation of inter-frames can be a sign: when its value is super small, it presents low redundancy in temporal domain. S-MVC is better for residual videos under this circumstance. For test materials that have relatively more redundancy in temporal domain, choosing coding method based on residual energy is more reasonable.

B. PROPOSED COMPRESSION ARCHITECTURE

The encoding architecture is proposed in Fig. 3. As depicted in the architecture, the plenoptic video is divided into the central view videos and the adjacent videos according to the subaperture projection-based grouping [22]. First, the central view videos are encoded by S-MVC. Then, the residual videos are generated by the difference between the reconstructed central subaperture videos and the adjacent subaperture videos. Encoding methods between S-MVC or T-MVC are chose adaptively for the residual videos according to correlation and residual energy calculation. A flag bit is set to record the encoding method selected by the residual videos.

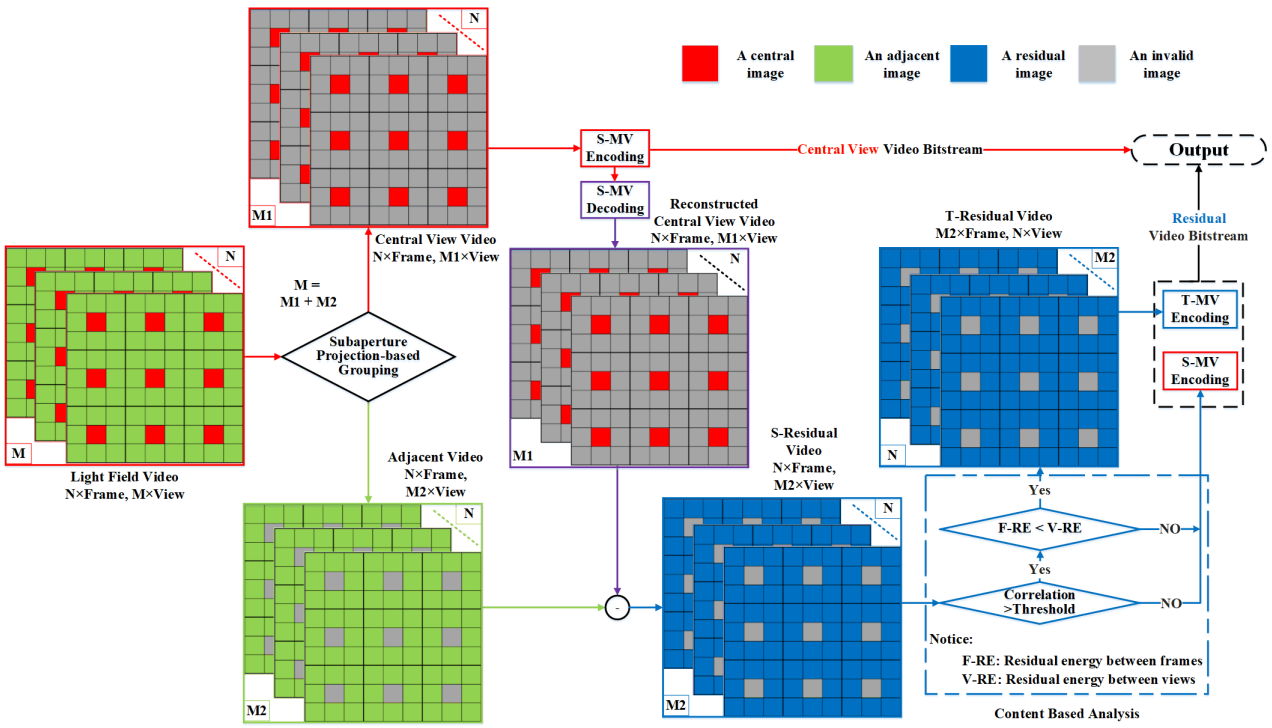


FIGURE 3. Proposed encoding architecture.

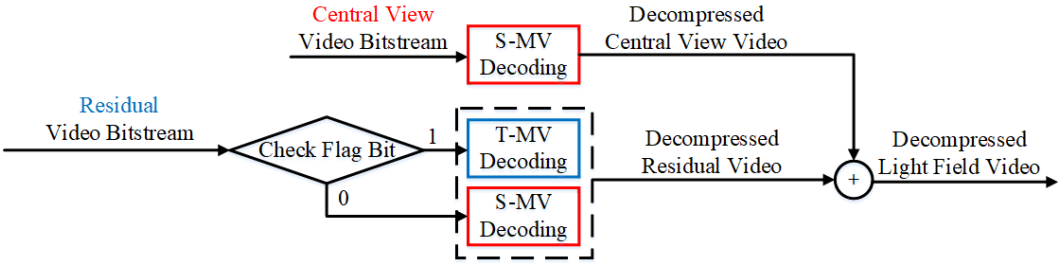


FIGURE 4. Proposed decoding workflow.

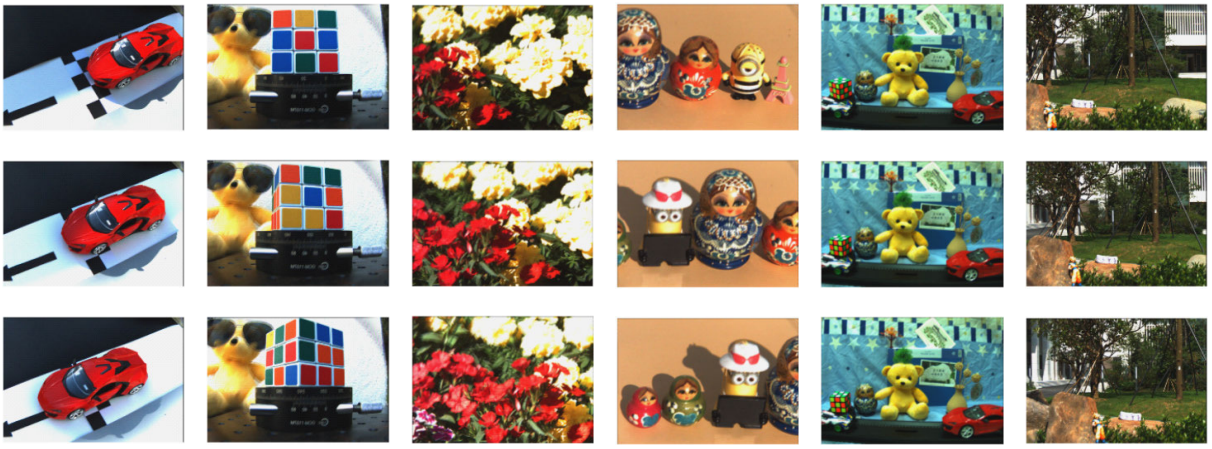


FIGURE 5. Test materials: row 1, 2 and 3 represent the 1st, 15th and 30th frame of each video, respectively.

The adaptive selection process is as follows: Correlation analysis is performed first, S-MVC is adopted when inter-frame correlation is smaller than an empirical threshold. Otherwise, residual energy analysis is implemented. Then, residual energy among frames and views are compared, T-MVC is adopted when inter-frame energy is smaller and S-MVC is adopted in the opposite case. In addition, ''S-Residual'' video is transformed to ''T-Residual'' video by exchanging the spatial domain and temporal domain before encoded by T-MVC. Finally, a central view videos bitstream and a residual videos bitstream are obtained.

In H.265, the NAL header contains three fixed length bit regions: NALU(T), NAL-REFERENCE-IDC(R) and hidden bits (F). The NALU type uses 5 bits to represent the 32 different types of NALU. Types 1-12 are defined by H.265, types 24-31 are used outside of H.265. Other values are reserved by H.265, which means types 13-23 are available. Therefore, we take the type 13 as the flag bit with no extra overheads introduced.

C. PROPOSED DECODING WORKFLOW

The decoding process of our proposed method is shown in Fig. 4. The central view videos bitstream is decoded by S-MVC. For the residual videos bitstream, the flag bit is checked first. S-MVC or T-MVC is chose adaptively for the residual videos bitstream based on the flag bit. By plus the decompressed central view videos to the decompressed residual videos, the complete decompressed plenoptic video is obtained.

IV. EXPERIMENT

A. TEST MATERIALS

All test materials are captured by Lytro camera. As depicted in Fig. 5, there are 30 frames in each plenoptic video. Row 1, 2 and 3 represent the 1st, 15th and 30th frame of each video, respectively. The motion modes of test materials are very different: The content of *Flowers* moves down in the both vertical direction and horizontal direction; The content of *Auto Race* is a toy car that moves diagonally; The *Cube* captures a rotating cube; The content of *Matryoshka* is a doll that moves down in the horizontal vertical direction, same with the *Trees*. The content of *Toys* is a combination of *Auto Race* and *Cube*. Different contents and different change mode of content ensure the sufficiency and rationality of the experiment.

B. EXPERIMENTAL SETUP

Subaperture images are extracted from plenoptic image using light field decoding toolbox [26]. The color space is converted from RGB to YCbCr 4:4:4 in the subaperture extraction process. According to the standard recommended by MPEG [27], 165 views in s single plenoptic image are used for compression. 25 subaperture images are classified as central view images and 140 subaperture images are

TABLE 4. Test methods and implementation details.

Method	Prediction Structure	Encoding Standard	Encoding method
Anchor	"PSB"	MV-HEVC	"S-MVC"
T-Anchor	"PSB"	MV-HEVC	"T-MVC"
LF-MVC	"LF-MVC"	MV-HEVC	"S-MVC"
HSS	"S-Shape"	HEVC	"HEVC"
CTRS	"LF-MVC" for central; "PSB" for residue	MV-HEVC	"T-MVC" for central; "S-MVC" for residue
CSRS	"LF-MVC" for central; "PSB" for residue	MV-HEVC	"S-MVC" for central; "S-MVC" for residue
CTRT	"LF-MVC" for central; "LF-MVC" for residue	MV-HEVC	"T-MVC" for central; "T-MVC" for residue
CSRT	"LF-MVC" for central; "LF-MVC" for residue	MV-HEVC	"S-MVC" for central; "T-MVC" for residue
CSRS(LF-MVC)	"LF-MVC" for central; "LF-MVC" for residue	MV-HEVC	"S-MVC" for central; "S-MVC" for residue
CHRT	"Spiral" for central; "PSB" for residue	HEVC+MV-HEVC	"HEVC" for central; "T-MVC" for residue
CHRS	"Spiral" for central; "PSB" for residue	HEVC+MV-HEVC	"HEVC" for central; "S-MVC" for residue
Proposed	**"PSB"**	**MV-HEVC**	**Content adaptive**

classified as adjacent images according to subaperture projection-based grouping [22]. Accordingly, 140 residual subaperture images are obtained in a plenoptic image. Two MVC prediction structures, pseudo-sequence-based (PSB) [28] and "LF-MVC" [29], are adopted in this paper.

The plenoptic videos in experiment are encoded by HEVC or MV-HEVC using reference software HM-16.9+ SCM8.0 [30] and HTM-16.2 [23]. All plenoptic videos are captured by *Lytro* Illum camera with spatial of 7728 × 5368. The REext main profile with low delay configuration is applied into the HEVC test, and the MV-HEVC encoding uses the low delay configuration with prediction structure corresponding to the proposed and reference methods. The initial residual value varying from -255 to 255 is adjusted as 0 to 510 by adding 255 before compression. Thus, the bit depth of residual videos is set as 10-bit. The PSNR is calculated by the original plenoptic video and the reconstructed plenoptic video. The PSNR of Image I (in dB) is defined as:

$$PSNR(dB) = 10\log_{10}(\frac{MAX_I^2}{MSE}). \tag{3}$$

where, MAX_I is the maximum possible pixel value of the Image I. The rendering method was proposed by [25] and also recommended by plenoptic compression evaluation method in [27]. BD-bitrate defined in [31] is used to measure the compression efficiency.

In order to verify the efficiency of the proposed method, 11 different coding methods for 6 test materials are conducted. All methods are defined and the specific implementation details are described in Table 4. In Table 4, what need to be explained in more detail are HSS, CHRT and CHRS. In HSS, we actually adopt the "S-Shape" [22] and reorder 165 × 30 subaperture images into one video and coding it with HEVC. We use "Normal S-Shape" (upper left corner to bottom right corner) for odd frames and "Reverse S-Shape" (bottom right corner to upper left corner) for even frames to form a continuous sequence. In CHRT and CHRS, we adopt the "Spiral", as depicted in Fig. 6, to reorder 25 × 30 central subaperture images into one video and coding it with HEVC. The residual videos in CHRT and CHRS are encoded with MVC.

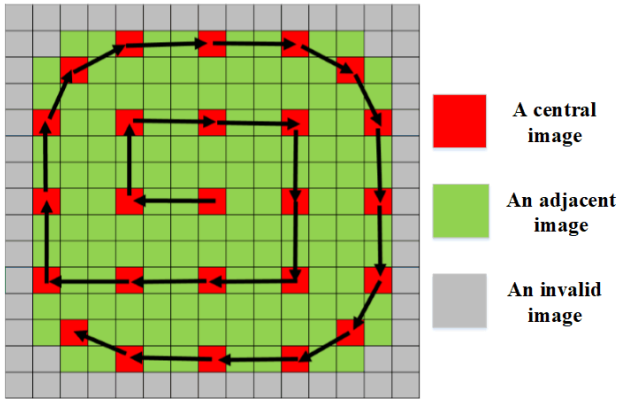


FIGURE 6. Reordering method: "Spiral".

C. EXPERIMENTAL RESULTS

To demonstrate the efficiency of the proposed compression method, several sets of experiments have been conducted in this section. First, the efficiency of S-MVC and T-MVC, PSB and LF-MVC are compared. Then, the efficiency of a special method, which is combination of HEVC and MVC method using subaperture projection-based grouping [22], is explored. Last, the efficiency of proposed method is proved by comparing all possible S-MVC and T-MVC combinations.

Several groups of methods are compared in Table 5. First, as shown in the column HSS, MVC method is quite more efficient for HSS presents an average of 80.04% increase in bitrate. Then, S-MVC is more efficient when coding a complete plenoptic video directly according to the column T-Anchor. However, things are different in CHRS and CHRT because of the similar average results, 17.35% and 17.74%. In fact, T-MVC is more efficient in test materials (c) and (e). In addition, LF-MVC is more efficient than PSB(Anchor) in 5 of 6 test materials.

All possible S-MVC and T-MVC combinations are compared in Table 5. First, a conclusion can be drawn that S-MVC is more efficient for the central view videos by comparing the CSRS and CTRS, CSRT and CTRT. CSRS and CSRT cannot be compared because different prediction structures (PSB for CSRS, LF-MVC for CSRT) are adopted for their residual videos. In order to solve this problem, an extra

TABLE 5. BD-rate performance of reference and proposed methods vs. Anchor.

Video	HSS	T-Anchor	LF-MVC	CHRS	CHRT	CSRS	CTRS	CSRT	CTRT	CSRS(LF-MVC)	**Proposed**
(a)	24.81%	52.53%	-9.70%	-16.73%	-18.11%	-34.01%	-26.88%	-35.08%	-33.13%	-34.58%	-35.08%
(b)	73.69%	38.23%	-6.17%	16.41%	18.30%	-14.74%	-12.71%	-16.41%	-16.01%	-17.91%	-17.91%
(c)	57.57%	73.46%	-11.28%	17.24%	15.31%	-20.24%	-3.81%	-18.82%	-4.48%	-21.93%	-21.93%
(d)	38.54%	66.45%	-10.49%	-0.19%	1.77%	-28.95%	-6.13%	-26.63%	-2.89%	-30.33%	-30.33%
(e)	226.81%	45.72%	10.83%	82.06%	79.64%	-16.93%	-9.62%	-20.73%	-16.71%	-17.47%	-20.73%
(f)	58.84%	36.43%	-10.82%	5.29%	9.51%	-22.95%	-8.40%	-19.79%	-8.85%	-24.37%	-24.37%
Avg.	**80.04%**	**52.14%**	**-6.27%**	**17.35%**	**17.74%**	**-22.97%**	**-11.26%**	**-22.91%**	**-13.68%**	**-24.43%**	**-25.06%**

method, CSRS(LF-MVC), is added. Thus, CSRS(LF-MVC), CSRT and the proposed method share the same prediction structure for both central view videos and residual videos. By comparing the results of these three methods, it can be found that the proposed method always chooses the more efficient method from S-MVC and T-MVC adaptively in all test materials, which proved the efficiency of the proposed method. Overall, the proposed method outperforms PSB by an average of 25.06% bitrate saving.

D. EXPERIMENTAL ANALYSIS

According to the content-based analysis conducted in Section III, we think S-MVC is more efficient when coding a plenoptic video directly. Thus, S-MVC is always better than T-MVC for central view videos, this argument has been proved by the experimental results of this section. For the residual videos, S-MVC or T-MVC is selected adaptively by calculating the correlation and residual energy, similarly, the efficiency has been proved. Therefore, the proposed method selects best method for both central view videos and residual videos, which make the proposed method outperforms traditional methods.

V. CONCLUSION

In this paper, an efficient content adaptive plenoptic video coding method is proposed. Based on the spatial correlations among subapertures, the plenoptic video is divided into two subaperture groups: the central view videos and the residual videos. The central view videos are encoded by S-MVC and the residual videos is encoded by S-MVC or T-MVC adaptively based on correlation and residual energy analysis. Experimental results demonstrate its superior improvement in compression efficiency relative to traditional MVC methods. The future work will focus on decreasing the number of views in the central view videos to further improve the compression performance.

REFERENCES

[1] M. Levoy and P. Hanrahan, "Plenoptic rendering," in *Proc. 23rd Annu. Conf. Comput. Graph. Interact. Techn.*, 1996, pp. 31–42.

[2] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," *SIGGRAPH*, vol. 96, no. 30, pp. 43–54, 1996.

[3] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 924–934, 2006.

[4] T. E. Bishop, S. Zanetti, and P. Favaro, "Light field superresolutio," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, Apr. 2009, pp. 1–9.

[5] D. Lanman and D. Luebke, "Near-eye light field displays," *ACM Trans. Graph.*, vol. 32, no. 6, pp. 1–10, Nov. 2013.

[6] F.-C. Huang, K. Chen, and G. Wetzstein, "The light field stereoscope," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 60:1–60:12, Jul. 2015.

[7] R. Ng, "Digital light field photography," Ph.D. dissertation, Stanford Univ., Stanford, CA, USA, 2006.

[8] Y. Li, M. Sjostrom, R. Olsson, and U. Jennehag, "Coding of focused plenoptic contents by displacement intra prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 7, pp. 1308–1319, Jul. 2016.

[9] C. Conti, P. T. Kovacs, T. Balogh, P. Nunes, and L. D. Soares, "Light-field video coding using geometry-based disparity compensation," in *Proc. 3DTV-Conf., True Vis.-Capture, Transmiss. Display 3D Video (3DTV-CON)*, Jul. 2014, pp. 1–4.

[10] C. Conti, L. D. Soares, and P. Nunes, "HEVC-based 3D holoscopic video coding using self-similarity compensated prediction," *Signal Process., Image Commun.*, vol. 42, pp. 59–78, Mar. 2016.

[11] S. Kundu, "Light field compression using homography and 2D warping," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 1349–1352.

[12] X. Jiang, M. Le Pendu, R. A. Farrugia, S. S. Hemami, and C. Guillemot, "Homography-based low rank approximation of light fields for compression," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 1313–1317.

[13] X. Jiang, M. Le Pendu, R. A. Farrugia, and C. Guillemot, "Light field compression with homography-based low-rank approximation," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 1132–1145, Oct. 2017.

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

[15] S. Shi, P. Gioia, and G. Madec, "Efficient compression method for integral images using multi-view video coding," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 137–140.

[16] S. Adedoyin, W. Fernando, A. Aggoun, and K. Kondoz, "Motion and disparity estimation with self adapted evolutionary strategy in 3D video coding," *IEEE Trans. Consum. Electron.*, vol. 53, no. 4, pp. 1768–1775, Nov. 2007.

[17] J. Dick, H. Almeida, L. D. Soares, and P. Nunes, "3D holoscopic video coding using MVC," in *Proc. IEEE Int. Conf. Comput. Tool (EUROCON)*, Apr. 2011, pp. 1–4.

[18] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient prediction structures for multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1461–1473, Nov. 2007.

[19] T. Chung, K. Song, and C.-S. Kim, "Compression of 2-D wide multi-view video sequences using view interpolation," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 2440–2443.

[20] T.-Y. Chung, I.-L. Jung, K. Song, and C.-S. Kim, "Virtual view interpolation and prediction structure for full parallax multi-view video," in *Advances in Multimedia Information Processing*, vol. 5879. Berlin, Germany: Springer, 2009, pp. 543–550.

[21] T.-Y. Chung, I.-L. Jung, K. Song, and C.-S. Kim, "Multi-view video coding with view interpolation prediction for 2D camera arrays," *J. Vis. Commun. Image Represent.*, vol. 21, nos. 5–6, pp. 474–486, Jul. 2010.

[22] H. Han, J. Xin, and Q. Dai, "Plenoptic image compression via simplified subaperture projection," in *Proc. Pacific Rim Conf. Multimedia*. Cham, Switzerland: Springer, 2018, pp. 274–284.

[23] *HTM-16.2*. Accessed: Mar. 2019. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftw-are/tags/HTM-16.2/

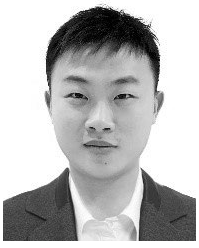
[24] E. Y. Lam, "Computational photography with plenoptic camera and light field capture: Tutorial," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 32, no. 11, p. 2021, Nov. 2015.
[25] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1027–1034.
[26] *Light Field Toolbox*. Accessed: Mar. 2019. [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4
[27] X. Sun, J. Xin, T. Zhong, and P. Wang, *[MPEG-I Visual] A New Self-Designed Focused Plenoptic Camera and the Captured Video Sequence Boys*, document ISO/IEC JTC1/SC29/WG11 MPEG2018/M46259, Marrakesh, Morocco, 2019.
[28] D. Liu, L. Wang, L. Li, Z. Xiong, F. Wu, and W. Zeng, "Pseudo-sequence-based light field image compression," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2016, pp. 1–4.
[29] G. Wang, W. Xiang, M. Pickering, and C. W. Chen, "Light field multi-view video coding with two-directional parallel inter-view prediction," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5104–5117, Nov. 2016.
[30] [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svnHEVCSoftware/tags/HM-16.9+SCM-8.0/
[31] G. Bjontegaard, *Calculation of Average PSNR Differences Between RD-Curves* document ITU-T VCEG-M33, 2001.

WANG TU received the B.S. degree from the Department of Automation, Hangzhou Dianzi University, Zhejiang, China, in 2014, where he is currently pursuing the M.E. degree with the Department of Automation. His current research interest includes light-field image/video compression.

XIN JIN (Senior Member, IEEE) received the M.S. degree in communication and information systems and the Ph.D. degree in information and communication engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2002 and 2005, respectively.

From 2006 to 2008, she was a Postdoctoral Fellow of The Chinese University of Hong Kong. From 2008 to 2012, she was a Visiting Lecturer with Waseda University, Fukuoka, Japan. Since March 2012, she has been with Tsinghua Shenzhen International Graduate School, China, where she is currently a Professor. She has published over 140 conference and journal articles. Her current research interests include computational imaging, and power-constrained video processing and compression.

Dr. Jin is a member of the SPIE and the ACM. She is the Distinguished Professor of Pengcheng Scholar. She received the ISOCC Best Paper Award, in 2010, the First Prize of the Guangdong Science and Technology Award, in 2015, and the Second Prize of the National Science and Technology Progress Award, in 2016.



LINGJUN LI received the B.S. degree from the Department of Automation, Beijing Jiaotong University, Beijing, China, in 2014. He is currently pursuing the M.E. degree with the Department of Automation, Tsinghua University, Beijing. His current research interests include light-field video compression and image/video coding. He has published several articles, such as the PCM, the ICASSP, and the VCIP.

CHENGGANG YAN received the B.S. degree in computer science from Shandong University, in 2008, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, in 2013. He is currently a Professor with Hangzhou Dianzi University. Before that, he was an Assistant Research Fellow of Tsinghua University. His research interests include intelligent information processing, machine learning, image processing, computational biology, and computational photography. He has authored or coauthored over 30 refereed journal and conference papers. As the coauthor, he received the Best Paper Candidate from the International Conference on Multimedia and Expo 2011, and the Best Paper Award from the International Conference on Game Theory for Networks 2014 and the SPIE/COS Photonics Asia Conference 2014.

YAOQI SUN received the B.S. degree from the Zhejiang University of Science and Technology, Zhejiang, China, in 2012. He is currently pursuing the M.E. degree with Hangzhou Dianzi University, Zhejiang. He is an Assistant Research Fellow with the Department of Automation, Hangzhou Dianzi University. His research interests include intelligent information processing, machine learning, and pattern recognition.

MANG XIAO received the B.S. degree in clinical medicine from Nanchang University, Nanchang, China, in 1999, and the Ph.D. degree in otolaryngology head and neck surgery from Zhejiang University, Hangzhou, China, in 2010. He is currently a Professor with the Department of Otolaryngology Head and Neck Surgery, Sir Run Run Shaw Hospital, Zhejiang University. His research interests include the etiology of head and neck tumors, and microvascular reconstruction of head and neck defect.

WEIDONG HAN received the B.S. degree in basic medicine from Zhejiang University, Hangzhou, Zhejiang, China, in 2003, and the Ph.D. degree in oncology from Zhejiang University, in 2008. He is currently an Attending Physician and an Associate Professor with the Department of Medical Oncolgy, Sir Run Run Shaw Hospital, Zhejiang University. He has published more than 60 research articles with more than 3000 citations. His current research interests include molecular mechanism of cancer drug resistance, computer-aided diagnosis, and the treatment of cancer.

JIYONG ZHANG received the B.S. and M.S. degrees in computer science from Tsinghua University, in 1999 and 2001, respectively, and the Ph.D. degree in computer science from the Swiss Federal Institute of Technology (EPFL), Lausanne, in 2008. He is currently a Distinguished Professor with Hangzhou Dianzi University. His research interests include intelligent information processing, machine learning, data sciences, and recommender systems.