A low complexity Wyner-Ziv coding solution for Light Field image transmission and storage

Huy Phi Cong[1,2,3], Stuart Perry[4], Xiem HoangVan[1]

[1] VNU-University of Engineering and Technology
[2] JTIRC, VNU University of Engineering and Technology, Hanoi, Vietnam
[3] School of Electrical and Data Engineering, University of Technology
Sydney, Australia
[4] University of Technology Sydney
17028025@vnu.edu.vn, stuart.perry@uts.edu.au, xiemhoang@vnu.edu.vn

Abstract— **Compressing Light Field (LF) imaging data is a challenging but very important task for both LF image transmission and storage applications. In this paper, we propose a novel coding solution for LF images using the well-known Wyner-Ziv (WZ) information theorem. First, the LF image is decomposed into a fourth-dimensional LF (4D-LF) data format. Using a spiral scanning procedure, a pseudo-sequence of 4D-LF is generated. This sequence is then compressed in a distributed coding manner as specified in the WZ theorem. Secondly, a novel adaptive frame skipping algorithm is introduced to further explore the high correlation between 4D-LF pseudo-sequences. Experimental results show that the proposed LF image compression solution is able to achieve a significant performance improvement with respect to the standard, notably around 54% bitrate saving when compared with the standard High Efficiency Video Coding (HEVC) Intra benchmark while requiring less computational complexity.**

Keywords— **Light field coding, distributed video coding, Wyner-Ziv coding, Signal processing**

I. INTRODUCTION

A. Context and motivations

LF is a popular form of image-based rendering (IBR) [1]. LF data captures information on the angle of incidence of light rays on an image sensor together with traditional spatial and intensity information. It can be presented as still or moving pictures. In particular, many cameras have been developed to capture LF data, for instance the Lytro LF, Illum [2] and Raytrix [3]. These cameras offer access to the amazing features of LF data such as changing perspective and viewpoints, digital refocusing, three-dimensional (3D) data extraction, depth estimation and modifiable post-capture [4].

However, deploying LF data are also facing to two main challenges, i.e. the storage and the transmission of the enormous size of data, which can be easily exceed ten Gigabytes in an uncompressed form [5]. This type of data requires highly efficient compression techniques. For instance, the work in [6] proposed a new context-adaptive encoding solution developed on the top of the HEVC inter-frames encoder structure [7] while in [8] a sparse set of LF views is encoded by an on-developing hybrid video encoder specified in the Joint Exploration Model (JEM) [9]. Likewise, data arrangement in [10,11] is also a prospective approach by generating the most suitable pseudo-sequence then compressing it using recent compression standards.

B. Contributions

WZ coding [12], a well-known source coding paradigm, provides a low encoding complexity capability by shifting the motion estimation part from the encoder to the decoder. This coding approach has successfully been applied to many different forms of video, e.g., natural images and hyperspectral images [13]. Several approaches for distributed compression of multi-view images which are similar in concept to LF images, have also been proposed in [14, 15].

In this paper, to achieve a LF compression solution with low encoding complexity capability while providing a good compression performance, we propose a WZ based LF image compression solution. In the proposed WZ based LF compression solution, the LF image is firstly decomposed into a pseudo-sequence of 4D-LF data. After that, the 4D-LF data is separated into 2 sub-sequences in which the WZ coding approach is employed for one part while the standard HEVC approach is used for the remaining part. In addition, to further explore the high temporal correlation between LF data, an adaptive frame skipping mechanism is also introduced. The contributions of this paper can be summarized as:

- A novel LF compression solution based on the combination between the WZ coding and a conventional video coding approach specified in HEVC standard
- An adaptive frame skipping mechanism for improving the proposed LF coding performance

The remainder of this paper is organized as follows. Section 2 briefly describes the background work on LF image and distributed video coding in general whereas the details of proposed architecture with the distributed video coding (DVC) approach are listed in Section 3. Section 4 mainly analyzes the experimental results for each test case while Section 5 gives some conclusions and future work.

II. BACKGROUND WORKS ON LIGHT FIELD IMAGE AND WYNER-ZIV CODING

A. Light Field image coding

LF data describes the set of light rays traveling at every angle at every point in 3D space [16], thus it includes information such as location (x, y, z), angle $(\theta, \emptyset)$, and wavelength γ, and the capture time t for light rays in the scene. This explains the huge amount of data stored in each LF image, as a LF image can include seven-dimensional information $(L(x, y, z, \theta, \emptyset, \gamma, t))$ [16].

Due to the complexity of LF information, it is common practice to introduce a set of constraints on the plenoptic function wherein it is reduced to a still extensive 4D function as in Eq. (1)

$$P_{LF} = L(u, d, x, y) \qquad (1)$$

Here, the light intensity P_{LF} is combined by (u, d) and (x, y) which denotes the angles and the set of viewpoints stored in each LF, respectively. Following [17], a set micro-

image (MI) which is generated by each micro-lens and represents as a set of views/perspective usually called sub-aperture images (SAI).

B. Wyner-Ziv Coding

WZ coding is the lossy case of the distributed source coding [18]. WZ theorem mainly states that separate encoding and joint decoding of two correlated sources, X and Y, can be as efficient as joint encoding and decoding. It refers to the lossy compression of X with side information (SI), Y available at the decoder [18]. Since Y is independently encoded and decoded while X is independently encoded but conditionally decoded, it is also known as asymmetric coding. For lossy coding, a rate loss is incurred when the SI is not available at the decoder. Thus, the rate-distortion (RD) function $R_{X/Y}^*(D)$ is established when the side SI is available at decoder only, with a given distortion D as shown below:

$$R_{X/Y}(D) \leq R_{X/Y}^*(D) \leq R_X(D) \qquad (2)$$

Where, $R_{X/Y}(D)$ is the RD function and Y is available at both encoder and decoder.

III. PROPOSED APPROACH

A. Observations

In the proposed LF coding solution, the LF image is firstly converted into 4D-LF. To form a pseudo-sequence, the set of 2D sub-aperture images (views) is scanned in a particular order. Several scanning order methods have been presented [10, 11]. It is observed that adjacent views in both horizontally and vertically of 4D-LF exhibit higher similarity with each other. Specifically, the similarity is between the views around the center compared to the views near the border. Thus, a spiral scanning order of the SAIs, is used to generate 4D-LF pseudo-sequences as shown in Fig. 3.
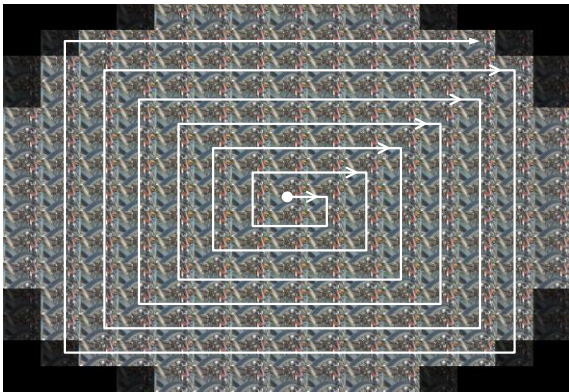

Fig. 3. Spiral scan for 4D-LF pseudo-sequences.

To analyze the motion characteristics of the 4D-LF pseudo-sequence generated above, the sum absolute difference (SAD) between two consecutive sub-aperture images is computed as the following equation:

$$SAD_{4D-LF} = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \left| SAI_{left}(x,y) - SAI_{right}(x,y) \right|$$

$$(3)$$

Here, SAI_{left} and SAI_{right} are two consecutive sub-aperture images, (x,y) is the pixel location in the SAIs with the size of N×M.

Fig. 4 shows SAD comparison between the natural videos, i.e., Foreman, Soccer [19] and 4D LF pseudo-sequences
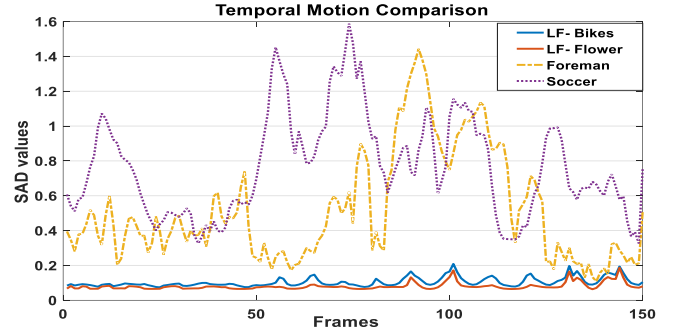

Fig. 4. Motion comparison between 4D-LF pseudo-sequences and natural sequences

As shown in Fig. 4, the SAD values computed for natural videos are significantly higher than that of the 4D-LF pseudo-sequences. This means, the temporal correlation along sub-aperture images of 4D-LF pseudo-sequences is higher than that of the natural sequences. In this case, the WZ coding solution which exploits the temporal correlation at the decoder may be a suitable coding solution for LF compression which requires the low encoding but still achieving high compression performance

B. Proposed LF Image Compression Architecture

To achieve a practical WZ coding solution for 4D-LF sub-aperture pseudo-sequence, we follows the Stanford like DVC coding approach [20] in which the 4D-LF pseudo-sequence can be divided into two sub-sequences.

While the sub-aperture images of the even positions, called key frames, are encoded with the conventional HEVC standard [7], the sub-aperture images of the odd positions, called WZ frames, are encoded with the WZ coding structure [20]. In this case, the source information, X, is the WZ frames while the SI, Y, is created at the decoder side using the common motion compensated temporal interpolation (MCTI) algorithm [21].

Since the 4D-LF images are highly correlated (see Section 3.A), a skipping mode decision is applied in the proposed framework. The skipping mechanism is detailed in Section 3. C.

Fig. 5 illustrates the proposed LF image compression architecture which can be performed as the following steps:

- **At the encoder:**

First of all, the LF data is unpacked and decoded into 4D-LF images composed of sub-aperture images. The sub-aperture images are then grouped into a pseudo-sequence using a spiral scanning order as stated in Section 3.A. The LF image compression is now cast as a common video coding problem.

The obtained sub-aperture image pseudo-sequence is then split into key and WZ frames in which the key frames are encoded with the HEVC Intra coding [7]. For the remaining WZ frames, a skipping mechanism is activated to decide which frame should be encoded with the WZ structure and which frames are skipped.
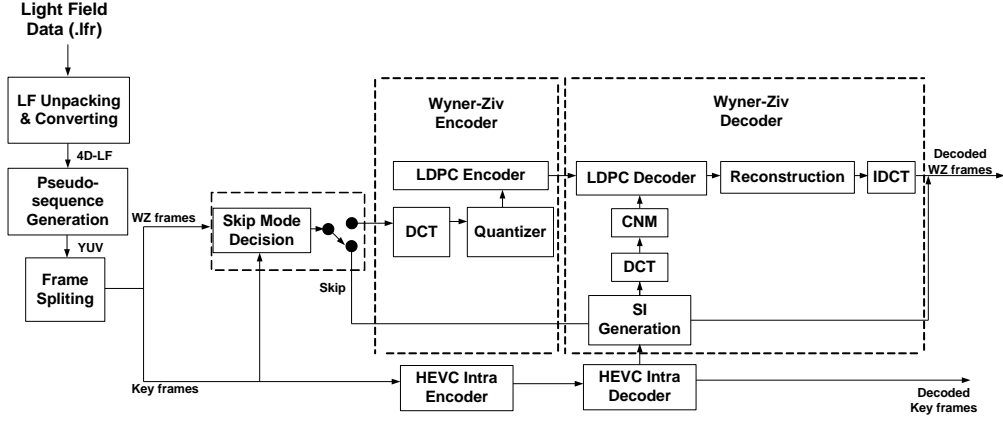
Fig. 5. Proposed LF image compression architecture

For WZ coding mode, the discrete cosine transform (DCT) follow with a uniform quantizer and Low Density Parity heck (LDPC) code are applied to compress the original WZ frames [22]. To signal the skipping mode, a flag is embedded into bit-stream for each frame.

- **At the decoder:**

If the Skip mode is selected from the encoder, the SI is naturally used as the final WZ reconstruction. Otherwise, the common WZ decoder process is applied, i.e. SI generation, LDPC decoder, Correlation Noise Modelling (CNM) and reconstruction.

SI generation: The obtained key frame bitrate is firstly decoded using the HEVC Intra decoder. After that, the SI is created using the decoded key frames [21].

LDPC decoder: This module decodes of a bit plane given the input value of SI from CNM and parity bits transmitted from the encoder. This decoding procedure is repeated for every increasing of number of parity bit requests from the decoder.

CNM: This module characterises the statistical relationship between the SI frame and the original frame through a distribution model. It is a complex task since the original information is not available at the decoder and SI quality varies throughout the sequence. If the model accurately describes the WZ and SI relationship, the coding performance is high and vice-versa. A Laplacian distribution model is applied in our architecture for its good trade-off between model accuracy and complexity.

Reconstruction: the parity bits, obtained from the LDPC decoder, together with the SI and the correlation noise information which are estimated from previous steps are used to reconstruct the WZ frame. Finally, the decoded key and WZ frames are merged to form the final 4D-LF images.

C. Frame Skipping Mechanism

The frame skipping mechanism is based on the technique wherein the motion activity between two consecutive 4D-LF images is measured through a sum absolute difference (SAD) metric as Eq (3).

This SAD metric is then compared to an experimentally derived threshold to decide whether or not the SKIP mode is used as Fig. 6.
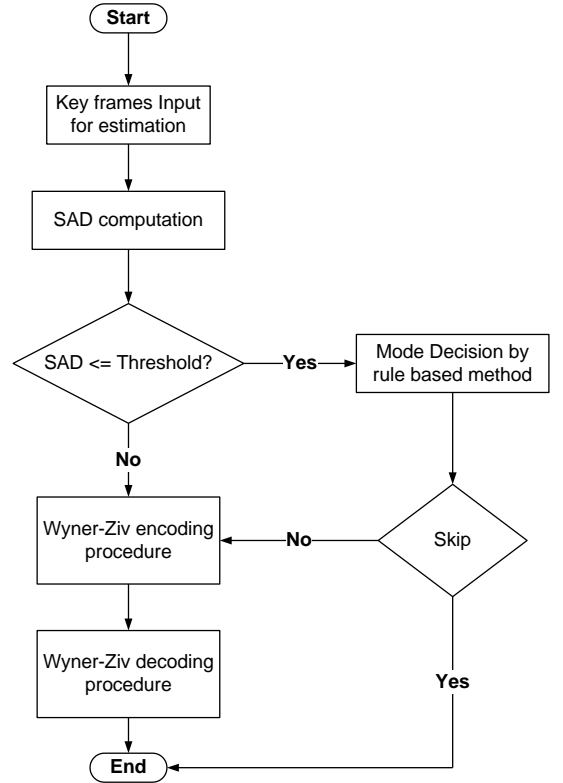

Fig. 6. Flowchart of frame skipping decision

The threshold value is adaptively calculated by averaging SAD of 4D-LF sub-aperture pseudo-sequences as:

$$Threshold = \frac{\left(\sum_{x=0}^{N-1}\sum_{y=0}^{M-1}|SAI_{left}(x,y) - SAI_{right}(x,y)|\right)}{(H * W)}$$

(4)

Where H and W are size of height and width of sequences, respectively.

According to observation, the range of threshold for 4D-LF sub-aperture pseudo-sequences is about 0 to 0.2 in order to activate the skip mode decision, otherwise it processes the WZ coding procedure.
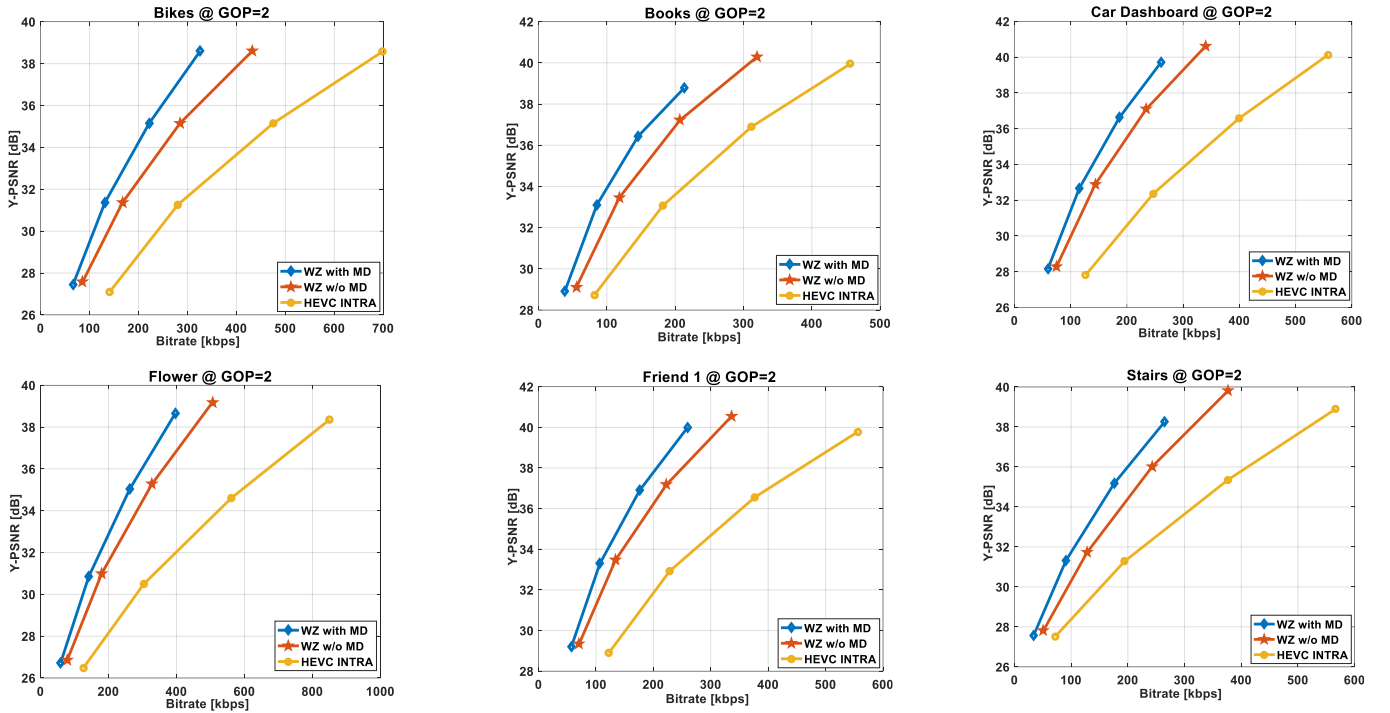
Fig. 7. RD performance evaluation

IV. EXPERIMETAL RESULTS

A. Test Methodology

In this paper, to assess the performance of the proposed LF compression solution, six common LF images [23] are examined with group of picture (GOP) size 2. The test methodology focused on the rate-distortion (RD) performance comparison between the proposed LF coding solution with the most relevant LF image coding benchmark, notably the HEVC intra; and also encoding complexity (measured in processing time) comparison.

For content visualization, thumbnails of selected LF images are shown in Fig. 8.
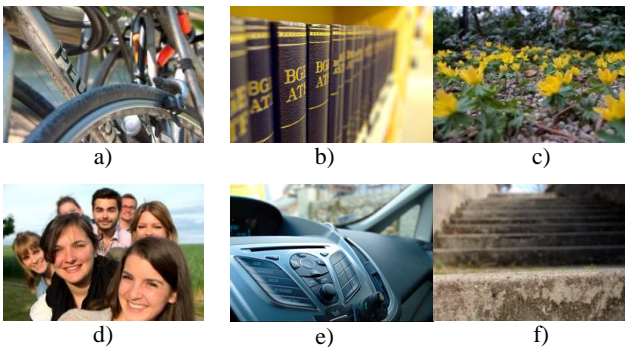

Fig. 8. Thumbnails of Light field images: (a) Bikes, (b) Books, (c) Flowers, (d) Friends 1, (e) Car_Dashboard, (f) Stairs

B. Compression evaluation

For video content, HEVC compression standard is currently state-of-the-art and provides the best compression performance compared to other standards. In this evaluation, the RD performance is compared and presented in Fig. 7., and Bjøntegaard Delta (BD)-Rate [24] saving compared to the HEVC Intra is computed in Table I.

TABLE I. BD RATE [%] SAVING COMPARED TO HEVC INTRA

LF sequences	WZ with MD	WZ w/o MD
Bikes	-53.85	-40.84
Books	-52.14	-37.41
Car_Dashboard	-54.30	-44.77
Flower	-55.79	-45.53
Friend-1	-55.40	-45.61
Stairs	-52.76	-39.54
Average	***-54.04***	***-42.28***

From the BD rate assessment of WZ coding solution with and without mode decision (MD) compared with HEVC Intra, as shown in Table I and Fig. 7, some conclusions can be drawn:

- Our proposed LF image compression architecture can provide much stable in bitrate saving compared to conventional distributed coding architectures.
- The proposed WZ with Mode Decision (labeled as WZ with MD) solution outperforms the HEVC Intra coding solution saving 54% bitrate in GOP2 while providing similar perceptual quality.
- The proposed WZ with MD solution gains significantly by 5dB compared to the most relevant HEVC Intra benchmark and by 2dB with WZ without Mode Decision (label as WZ w/o MD).

C. Complexity evaluation

Examining compression complexity is an essential part of performance evaluation. For this evaluation, the coding solutions are tested on the same PC with an Intel core i7-7700HQ (2.8 GHz) processor, 16GB RAM, and Windows 10-Home OS. The test is run on all quantization parameters (QPs) of 40, 34, 29 and 25, with GOP2. The results shown in Table II and III are for QP 40 and QP 25. To avoid the effect of

multi-thread processing during the test, the results of 5 repetitions of the same compression setting are averaged.

From these complexity results, some points can be observed:

- Our proposed WZ with MD is almost 2 times faster than HEVC Intra and also slightly faster than the WZ without MD.
- Based on the GOP size process, our proposed solution may achieve faster encoder with higher GOP sizes

TABLE II. TIME COMPLEXITY (S) COMPARED BETWEEN WYNER-ZIV AND HEVC INTRA ENCODER AT QP 40

LF sequences	QP 40@GOP2		
	WZ with MD	WZ w/o MD	HEVC INTRA
Bikes	116.41	118.26	207.48
Books	95.40	97.55	182.66
Car_Dashboard	156.99	158.87	194.96
Flower	144.00	147.27	195.50
Friend-1	150.31	152.44	195.44
Stairs	139.57	141.73	177.89
Average	*133.78*	*136.02*	*192.32*

TABLE III. TIME COMPLEXITY (S) COMPARED OF WYNER-ZIV AND HEVC INTRA ENCODER AT QP 25

LF sequences	QP 25@GOP2		
	WZ with MD	WZ w/o MD	HEVC INTRA
Bikes	170.66	178.09	289.14
Books	144.46	154.44	252.26
Car_Dashboard	204.56	209.48	259.36
Flower	232.18	241.03	289.34
Friend-1	192.71	199.31	262.12
Stairs	199.43	208.27	248.37
Average	*190.67*	*198.44*	*266.77*

V. CONCLUSION

This paper proposes a novel WZ coding based LF image compression solution and compares the solution state-of-the-art HEVC Intra codec. The proposed LF coding solution significantly outperforms the relevant HEVC Intra, for both coding structures with and without Skip Mode Decision. In particular, the proposed LF coding solution also provides a lower computational complexity than the HEVC Intra approach. This is very important for the future exploitation of LF images.

ACKNOWLEDGMENT

REFERENCES

[1] M. Levoy and P. Hanrahan, "Light field rendering," in Proc. SIGGRAPH, pp. 31–42, 1996.

[2] Lytro camera, https://www.lytro.com/

[3] Raytrix, https://www.raytrix.de/

[4] Ivo Ihrke, John Restrepo, and Loïs Mignard-Debise, "Principles of Light Field Imaging", IEEE Signal Processing Magazine, 2016.

[5] M. Levoy, K. Pulli, et al., "The Digital Michelangelo project: 3D scanning of large statues," in Computer Graphics (Proceedings SIGGRAPH 00), pp. 131–144, Aug. 2000.

[6] R. Conceição, M. Porto, B. Zatt and L. V. Agostini, "LF-CAE: Context-Adaptive Encoding For Lenslet Light Fields Using HEVC", 2018 IEEE International Conference on Image Processing (ICIP), Greece, Oct. 2018

[7] G. J. Sullivan, J. R. Ohm, W. J. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[8] N. Bakir, W. Hamidouche, O. Déforges, K. Samrouth and M. Khalil, "Light Field Image Compression based on Convolutional Neural Networks and Linear Approximation", 2018 IEEE International Conference on Image Processing (ICIP), Greece, Oct. 2018.

[9] "Algorithm Description of Joint Exploration Test Model 6", Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11), 6th Meeting, Hobart, Doc. JVET-F1001-v3, Apr. 2017.

[10] L. Li, L. Zhu, L Bin, L. Dong, and L. Houqiang, "Pseudo Sequence Based 2-D Hierarchical Coding Structure for Light-Field Image Compression," 2017 Data Compression Conference (DCC), Snowbird, UT, 2017

[11] C. Perra and P. Assuncao, "High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement", IEEE International Conference on Multimedia and Expo, Seattle, USA, Jul. 2016.

[12] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, 'Distributed video coding', in Proceedings of the IEEE, vol. 93, no. 1, pp. 71-83, 2005.

[13] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid and A. Abbas, "Modern Trends in Hyperspectral Image Analysis: A Review," in IEEE Access, vol. 6, pp. 14118-14129, 2018.

[14] X. Zhu, A. Aaron, and B. Girod, "Distributed compression for large camera arrays," in IEEE SSP '03 , Sept. 2003.

[15] G. Toffetti, M. Tagliasacchi, M. Marcon, A. Sarti, S. Tubaro, and K. Ramchandran, "Image compression in a multi-camera system based on a distributed source coding approach," in EUSIPCO '05, Sept. 2005.

[16] G. Wu *et al*., "Light Field Image Processing: An Overview," in IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 7, pp. 926-954, Oct. 2017

[17] E. H. Adelson and J. Y. A. Wang, "Single Lens Stereo with a Plenoptic Camera," IEEE Trans. Pattern Anal. Mach. Intell., vol. 14, no. 2, pp. 99–106, 1992.

[18] A. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder," IEEE Transactions on Information Theory, vol. 22, no. 1, pp. 1-10, Jan. 1976.

[19] Xiph.org Video Test Media [derf's collection], https://media.xiph.org/video/derf/

[20] X. Artigas *et al*, 'The discover codec: architecture, techniques and evaluation', in Proceedings of Picture Coding Symposium (PCS'07), Lisboa, Portugal, Nov. 2007

[21] J. Ascenso, C. Brites and F. Pereira, "Content Adaptive Wyner – Ziv Video Coding Driven by Motion Activity", IEEE International Conference on Image Processing, Atlanta, USA, Oct. 2006.

[22] D. Varodayan, A. Aaron and B. Girod, "Rate-Adaptive Codes for Distributed Source Coding", EURASIP Signal Processing Journal, Special Section on Distributed Source Coding, vol. 86, no. 11, Nov. 2006.

[23] M. Řeřábek and T. Ebrahimi, "New Light Field Image Dataset," 8th International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 2016.

[24] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves", Doc. ITU-T SG16 VCEG-M33, Austin, TX, USA, Apr. 2001.