VIEW SYNTHESIS-BASED DISTRIBUTED LIGHT FIELD COMPRESSION

M. Umair Mukati[†], Milan Stepanov[*], Giuseppe Valenzise[*], Frédéric Dufaux[*], Søren Forchhammer[†]

[†]DTU Fotonik, Technical University of Denmark,
Ørsteds Plads, Kgs. Lyngby - 2800, Denmark

[*]Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes,
91190, Gif-sur-Yvette, France

ABSTRACT

Light field imaging is becoming a key technology, which provides users with a realistic visual experience through the capability of dynamic viewpoint shifting. This ability comes at the cost of capturing huge amounts of information, leaving the problem of compression and transmission a challenge. The encoder complexity is the key to achieve efficient coding in conventional light field coding schemes, where a complicated prediction process is essentially used at the encoder side to exploit the redundancy present in the light field image. We employ Distributed Source Coding (DSC) for light field images, which can extensively lift the computational requirement from the encoding side at the expense of increased computational complexity at the decoder side. The efficiency of DSC is heavily dependent on the quality of side information at the decoder. Therefore, we propose to leverage a learning-based view synthesis method, which takes into account the light field structure to generate high-quality side information. We compare our approach to Distributed Video Coding and Distributed Multi-view Video Coding schemes adapted to the light field framework and relevant standard-based approach, and demonstrate that the proposed view synthesis-based approach can achieve similar performance, while substantially reducing the number of key views to be transmitted.

Index Terms— Light field, distributed source coding, view synthesis

1. INTRODUCTION

Light field (LF) is a relatively new paradigm in image acquisition technology. It offers some off-the-shelf capabilities such as refocusing, aperture adjustment and view-point shifting after capturing the scene. Contrary to traditional camera technology which captures the light intensity focused at the sensor plane, LF technology captures the intensity of light rays passing through it, thereby recording not only the spatial coordinate of the incident light ray but also its angular orientation. A major challenge comes while storing or transmitting this information due to the amount of captured data. A typical LF image captured by LYTRO Illum camera offers only a 0.25 megapixel resolution albeit occupying about 218 megabyte of hard disk space[1]. This also limits the rate of its transmission due to high bandwidth requirement. Therefore, LF coding is considered as an important research topic.

In the literature, an encoder usually exploits the redundancy present in the input data to compress it. Generally, the method of exploiting redundancy is highly complex, resulting in a high computational demand at the encoder. Contrary to these schemes, in Distributed Source Coding (DSC), the correlation is exploited at the decoder side, which effectively lifts the complex computations from the encoder. From the LF acquisition perspective, DSC can thus release the burden of the camera processor while still guaranteeing efficient data transmission. DSC is based on the theoretical results of Slepian-Wolf and Wyner-Ziv (WZ) theorems [1]. According to them, two correlated sources can be coded with a total rate lower bounded by their joint entropy (after quantization), even if only one of the two sources is available at the decoder.

In practical Distributed Video Coding (DVC) [2] schemes, video frames are divided into two groups: key frames and WZ frames. Key frames are encoded using traditional, hybrid coding schemes. Conversely, WZ frames are initially estimated based on the decoded key frames; this *side information*, available at the decoder, is then corrected through channel codes requested from the encoder. Since generating parity bits (e.g., syndromes [3]) is computationally much lighter than temporal prediction, the complexity cost at the encoder is reduced by decreasing the number of key frames. This framework has been later extended to Distributed Multi-view Video Coding (DMVC) [4], and has been applied to LF as well in the preliminary works [5][6]. However, distributed coding of LF has remained little explored till now.

In this work, we build on top of the latest state-of-the-art method in DMVC [7], and we propose improving the estimation at the decoder side. More precisely, we replace the typically employed optical flow [8] or overlapped block mo-

This project has received funding from EU's H2020 ITN programme, under the MSCA grant agreement No 765911 (RealVision).

[1] 15x15 views, 10 bit, 3 color channels

978-1-7281-1485-9/20/$31.00 ©2020 IEEE

tion compensation [9] to generate side information (SI) with a learning-based view synthesis approach, which estimates the scene geometry and inpaints occlusion, to obtain higher-quality estimates. We compare to distributed LF coding approaches based on optical flow to generate SI in two scenarios: Pseudo Video Sequence (PVS) and sub-aperture images representation, motivated by DVC and DMVC, respectively. Furthermore, we show that a view synthesis approach, that efficiently leverages the LF structure to synthesize intermediate views, can provide competitive coding performance even if only a small number of key views are transmitted. This enables to significantly reduce the computation requirements at the encoder side.

2. RELATED WORK

We divide the related work in three parts: DSC of LFs, DVC and DMVC approaches, and view synthesis.

In [5], Zhu et al. used DSC to encode camera views in the pixel-domain. At the decoder, they synthesize SI using neighboring views through geometry-based image rendering. Aaron et al. [6] encoded LF views in the transform domain and utilized scene geometry calculated at the encoder using original images to estimate SI. More recently, Cong et al. [10] proposed to generate PVS from a LF image to encode it in a distributed manner. We have implemented similar approach and compared with our proposed method.

Conversely, DVC and DMVC have received more attention. More precisely, novel approaches proposed improving SI, as this plays a major role in the overall RD performance. The quality of generated SI can be improved by utilizing more adjacent frames [11] or multiple SI generation techniques [12, 13], which usually results in more than one SIs. Maugey et al. [14] proposed three schemes to fuse the SI. Among the schemes, the fusion scheme utilizing the reciprocal of the residual and the reciprocal of vector magnitude as weights have superior performance compared to the former two fusion schemes. Salmistraro et al. [7] propose a DMVC approach which exploits temporal and inter-view redundancies at the decoder side by generating multiple SIs. Moreover, a robust fusion method is employed by fusing likelihoods estimated from each SI.

View synthesis generates a view at a novel perspective from views given at different perspectives. The application of machine learning methods allowed further improvement in the view synthesis domain by allowing the generation of higher-quality views from sparser input sets. In their seminal work, Kalantari et al. [15] proposed a machine learning approach for view synthesis which outperformed previous conventional approaches. They processed four corner views of a LF image through a series of convolutional layers which estimated the disparity at the novel view, warped the input views and merged them to generate the final novel view. Srinivasan et al. [16] proposed to generate the whole LF from a single

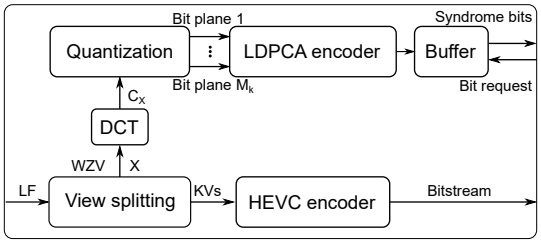


Fig. 1. Block diagram of the TDWZ encoder.

image by predicting the scene geometry, estimating Lambertian surfaces using the estimated geometry and finally modeling the occluded areas and non-Lambertian parts. Recently, Navarro et al. [17] proposed a three-part network which estimates the disparity map of each of the four corner views in order to better treat occluded regions and fuse warped corner views using learned weight to synthesize a novel view. We select the method in Navarro et al. [17] which achieves superior performance compared to other methods and incorporate it in our SI generation block.

3. PROPOSED METHOD

Here, we describe the whole coding scheme with the encoder in Sec. 3.1 and the decoder in Sec. 3.2. In the Sec. 3.2.1, we present our proposed method to generate SI employing the view synthesis based approach.

3.1. Encoding of light field views

At the encoder, as illustrated in Fig. 1, key views (KVs) are encoded using the High-Efficiency Video Coding [18] encoder in Intra mode while the WZ views are encoded following a Transform Domain WZ (TDWZ) architecture, where each view is initially transformed using a 4×4 Discrete Cosine Transform (DCT) operator [2]. Subsequently, the DCT coefficients are uniformly quantized using a quantization matrix from a proposed set [19] to achieve rate adaptivity and rearranged into bands of the same frequency. Starting from the lowest frequency, each frequency band is then passed to the LDPCA encoder [3] for encoding.

The LDPCA encoder first converts a frequency band into bit planes. These bit planes are successively encoded from the most significant bit to the least significant bit. Before transmission, an accumulated syndrome is calculated for each bit plane using a predefined low-density parity-check (LDPC) matrix. The syndrome is then fed into a bit accumulator as described in [3] to achieve accumulated syndrome. Along with the cyclic information of the bit plane, the accumulated syndrome is transmitted in small parts each time the WZ decoder requests for more information until the bit plane is successfully decoded.

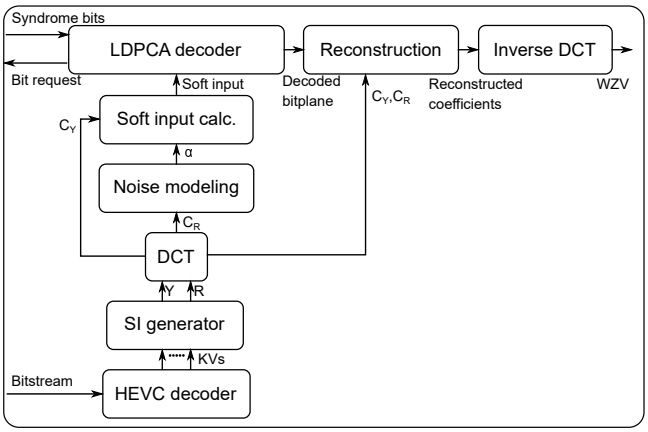


Fig. 2. Block diagram of the TDWZ decoder.

3.2. Joint decoding of light field views

At the decoder, as illustrated in Fig. 2, decoded key views are used in the SI generation block to estimate a WZ frame and the corresponding residual signal. For this purpose, we propose employing a deep learning method that considers LF structure and present it in the following sub-section. We conclude the section by briefly describing the WZ view decoding and reconstruction.

3.2.1. Side information generation

SI represents a combination of an estimated WZ view and noise residual signal. The noise residual signal is the view estimation error, which is calculated by taking the difference between the estimated and the original views. Due to the unavailability of the original view at the decoder, the noise residual signal is estimated at the decoder.

This process is crucial as its accuracy highly determines the quality of the reconstructed view. The SI is utilized in the decoding of the bit planes as well as in the view reconstruction phase. To generate SI for a view (s,t), we assume that all the required KVs are already decoded. Recent advances in deep learning-based view synthesis methods showed improvement over traditional view synthesis schemes and we propose to leverage these advances to improve SI generation. We implement the scheme from Navarro et al. [17] which achieves higher quality of synthesized views by implicitly treating occlusions, by estimating four disparity maps. The approach consists of three networks: a feature extraction network f_f, a disparity estimation network f_d and a selection network f_s, trained in an end-to-end fashion. Given four decoded corner KVs $\{\bar{I}_i\}$ and angular coordinates of a WZ view (s,t), f_f extracts independently features

$$F_i = f_f(\bar{I}_i, s, t) \tag{1}$$

of each corner view $i \in \mathbf{I}$, with $\mathbf{I} = \{(0,0),(0,N),(M,0),(M,N)\}$. The extracted features

$$\mathbf{F} = (F_{0,0}, F_{0,N}, F_{M,0}, F_{M,N}) \tag{2}$$

are provided to f_d which estimates the disparity $d_{m,n}$ of each corner view with respect to the view being synthesized.

$$(d_{0,0}, d_{0,N}, d_{M,0}, d_{M,N}) = f_d(\mathbf{F}, s, t) \tag{3}$$

Each disparity map is used to warp each pixel x of corner views to the view position (s,t) following

$$W_i(x) = \bar{I}_i\left(x + d_i(x)\right). \tag{4}$$

The selection network f_s learns contributions of each warped view

$$(w_{0,0}, w_{0,N}, w_{M,0}, w_{M,N}) = f_s(\mathbf{F}, \mathbf{W}, s, t) \tag{5}$$

to the final predicted view (s,t)

$$\bar{Y}^{s,t}(x) = \sum_{i\in\mathbf{I}} w_i(x) W_i(x) \tag{6}$$

where $\mathbf{W} = (W_{0,0}, W_{0,N}, W_{M,0}, W_{M,N})$.

In order to estimate noise residual signal of the corresponding predicted view (s,t), warped corner views (Eq. (4)) are subtracted from the estimated view:

$$R_i^{s,t}(x) = \bar{Y}^{s,t}(x) - W_i(x), \tag{7}$$

where x denotes a spatial pixel position, and W_i is a warped view. The estimated individual residual noise signals are merged following:

$$R^{s,t}(x) = \sum_{i\in\mathbf{I}} w_i^{\text{residual}}(x) R_i^{s,t}(x), \tag{8}$$

where

$$\begin{aligned} w_i^{\text{residual}}(x) &= \frac{\log \prod_{j\in\mathbf{I}\setminus i} |R_j(x)|}{\sum_{k\in\mathbf{I}} \log \prod_{j\in\mathbf{I}\setminus k} |R_j(x)|} \\ &= \frac{\sum_{j\in\mathbf{I}\setminus i} \log |R_j(x)|}{\sum_{k\in\mathbf{I}} \sum_{j\in\mathbf{I}\setminus k} \log |R_j(x)|}. \end{aligned} \tag{9}$$

The level of uncertainty in the estimation process is well represented with the degree of agreement of the warped KVs. Therefore, the SI with higher uncertainty should contribute less to the final residual. Thus, the reciprocal of noise value is better suited to model the contribution of the noise value. But, the sum of reciprocal value introduces a multiplication operation which becomes highly sensitive to changes in residual value. Therefore, we apply the natural logarithm function to achieve a more stable solution as proposed in Eq. (9).

3.2.2. Wyner-Ziv view decoding and reconstruction

After SI is generated, it is transformed using a 4×4 DCT. Then, the resulting SI coefficients, the transformed estimated view C_Y and the transformed estimated residual C_R, are passed to the noise modeling block. We employ the Laplacian distribution to model the noise residual. The Laplacian parameter α indicates the reliability of the estimated view. An accurate noise model directly impacts the number of syndrome bits requested to the encoder in order to correct the estimation errors. We use noise modeling at coefficient level following the approach in [20].

The estimated parameter α and the transformed estimated view C_Y are used to compute the soft input by using C_Y as the mean and α as the parameter of the Laplacian distribution. Then, by calculating the area under the distribution where the bit is one, the likelihood of bit being equal to one is computed.

The LDPCA decoder is a probabilistic decoder, which requires soft input and syndrome bits to decode a bit plane. A soft input for the whole bit plane gives the likelihood of a particular bit being one or zero. It is calculated for each bit plane of the quantized DCT coefficients by transforming the estimated view from the SI using the same set of operations applied to the original view at the WZ encoder (i.e. the 4×4 DCT transformation, the coefficient quantization and the division into bit planes). After receiving syndrome bits from the encoder, the bit plane is decoded in an iterative fashion using the message passing algorithm" as described in [21].

Finally decoded bit planes are combined together and along with SI they are used to reconstruct the final DCT coefficients following [22].

4. EXPERIMENTAL RESULTS

In this section, we define training and testing conditions, then, we describe the anchors to evaluate our proposed approach and finally, we present the results.

4.1. Training conditions

We have implemented the viewsynthesis model, described in Sec. 3.2.1, in PyTorch. The LF dataset provided by Srinivasan et al. [16] which consists of 3343 LF images captured by LYTRO Illum camera has been used for the training. We choose 3243 images for the training set while the rest of the images are used for the validation set. In each iteration, a spatial position of a 192×192 patch from each corner view and a target view at the angular position (s, t), and the angular position of the target view are randomly selected. They are then used to learn the weights of the model in the supervised manner. We minimize the sum of the L1-norm of the difference between the target view and the network output, and the L1-norm of the difference between their gradients using ADAM optimizer with a batch size 10.

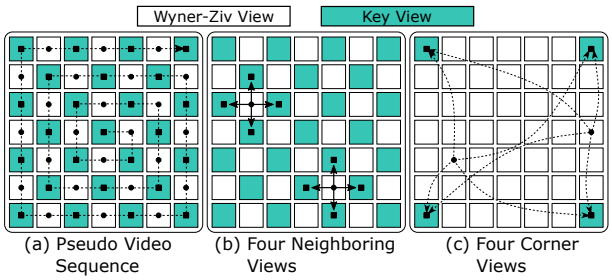


Fig. 3. View splitting modes.

4.2. Testing conditions

The performance was evaluated on four LF images from EPFL LF dataset [23] suggested by JPEG Pleno initiative [24]. Raw input lenslet images were decoded using LFToolbox version 0.4 [25] by demosaicing, devignetting and resampling to a rectangular grid, followed by color and gamma correction. The resulting LF image provides a set of 15×15 views of 434×625 pixels with 10-bit precision. In our experiments, we crop LF to 7×7 views due to noticeable artifacts at peripheral views which would degrade the SI generation block. To demonstrate results, only luminance channel of LF image is considered in the coding process. Firstly, the effective resolution of each view is set to 436×628 (governed by 4×4 DCT operation which demands that the resolution of a view be a multiple of four). The DCT generates 109×157 coefficients for each frequency band, resulting in a binary source code of 17113 length for each bitplane. We designed LDPCA codes for this length following the procedure described in [3].

Key views are decoded using HEVC Intra decoder (HM reference software, v.16.0, with Range Extension and Main10 profile). For each RD point, different quantization parameters are selected during HEVC coding such that the decoded KVs and reconstructed WZVs have a similar quality, as specified in Table 1.

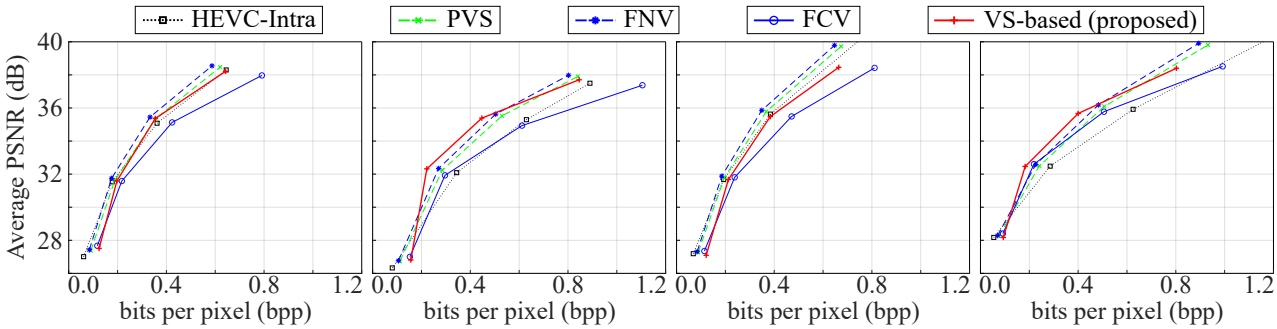
Table 1. Key frames quantization parameters for the four RD points.

Sequence	Q_1	Q_4	Q_7	Q_8
Bikes	46	38	32	27
Danger de Mort	45	35	30	27
Fountain Vincent2	45	38	32	25
Stone Pillars Outside	43	34	29	24

4.3. Anchors

In this section, we present variations of DSC approaches for LF compression which will be compared with the proposed method.

First, we consider a scenario similar to the GOP2 structure, i.e. one-half of the views are encoded using a conventional non-DSC approach while the rest of the views are encoded using the WZ approach. Two possible approaches to encode a LF are evaluated: PVS, where the SI is generated using two adjacent pseudo frames; and sub-aperture image,



(a) Bikes (b) Danger de Mort (c) Fountain Vincent 2 (d) Stone Pillar Outside

Fig. 4. RD performance for test LF images.

where four neighboring views (FNV) are used to estimate the SI. In the PVS case, as illustrated in Fig. 3 (a), we examine the LF coding set in the DVC scenario by rearranging LF views in a stream following the spiral order. Before decoding a WZV we assume that the respective neighboring frames are already decoded. Then, in the SI generation part, optical flow [8] is used to estimate the disparity vectors between KVs which are warped to procure an estimate of the WZV. Then their average is taken as the final view estimate while their difference is taken as the estimate for the residual signal. In FNV case, as illustrated in Fig. 3 (b), the views are split in a checkerboard pattern allowing to utilize horizontal and vertical adjacent neighbors to estimate SI for WZV in a similar manner as in DMVC [7]. Contrary to DMVC, an additional angular dimension and LF image signal are considered. Effectively, the setup allows to generate two SIs, one for horizontal neighbors and one for vertical neighbors, which are independently utilized to calculate two soft inputs. A single soft input is passed to the LDPCA decoder after fusing the two soft inputs utilizing the correspondingly computed Laplacian parameter α as their weights.

In the second experiment, we note that in the formerly presented approaches the complexity reduction is limited, as half of the key views still needs to be processed by the HEVC encoder. We consider a similar scheme as our proposed method, which requires only 4 KVs, while replacing the SI generation method with optical flow. We denote this scheme as FCV (Four Corner Views). An illustration of this scheme is given in Fig. 3 (c). Each corner view is used as a reference to estimate disparity vectors to all other reference views resulting in a total of 12 disparity maps. For each reference view, the disparity vectors are normalized and averaged to procure the final disparity map of the view. In the final step, the disparity vectors are scaled based on the distance between a WZV at angular position (s, t) and reference views, and then used to warp each reference view. Finally, the proximity of index (s, t) to each reference view is used as a weight to fuse all the warped images.

In addition, we also provide comparison with HEVC Intra to encode all the 49 KVs as a relevant standard-based solution.

4.4. Performance analysis

We compare our proposed method with the approaches described in the previous section. The comparison is performed in terms of PSNR which is computed for a LF images as the average quality across all the views.

Regarding distributed LF coding methods, it can be observed from Fig. 4 that FNV achieves superior performance compared to PVS. The results reflect that with additional information in the former case it is possible to improve the prediction process, e.g., reproducing the occluded regions. It is noticeable that rate-distortion performance of our proposed VS-based approach is comparable to FNV, even if we use only four key views. For the highest bitrate, we can observe underperformance of our approach which we believe comes from the inability of the VS network to generate fine details due to the larger baseline along the corner views, in contrast to the high-quality SI generated by FNV due to the higher correlation in the adjacent neighbors. At the lowest bitrate, we can observe gains in favor of FNV which we believe appear because the VS network was trained on undistorted LF dataset. For the remaining two bitrates, our approach outperforms FNV across the contents *Danger de Mort* and *Stone Pillars Outside* while for the contents *Bikes* and *Fountain Vincent 2* we can observe reduction in performance. The reason is the poor disparity estimation in the former cases for optical flow due to the repetitive patterns and high-frequency contents. The comparison with HEVC Intra suggests that our approach achieves similar performance at the contents *Bikes* and *Fountain Vincent 2* while it outperforms HEVC for *Danger de Mort* and *Stone Pillars Outside*.

The overall performance across contents demonstrates that similar performance can be achieved with very low encoding complexity with a high-end SI generation scheme. For example, our measurements show that, on average, we can reduce the encoding time by $\sim$25 times for the highest bitrate, and up to $\sim$55 times for the lowest bitrate. If we keep the encoding complexity and the number of side information the same, we can notice that our approach significantly outperforms the equivalent optical flow-based method FVC, thanks to the improved SI generation scheme.

5. CONCLUSION

In this work we have presented a novel approach for distributed LF compression based on view synthesis. View synthesis complements the distributed coding paradigm as it enables generating high-quality novel views from a sparse set of key views, effectively allowing to reduce the number of key views to be coded and transmitted. Plugging a deep learning-based view synthesis method into a distributed coding scheme leads to better coding performance compared to the HEVC Intra benchmark. Furthermore, we achieve similar performance as previously proposed DSC schemes, but using a much smaller fraction of required key views.

Apart from the view estimation process, we have observed that residual estimation is critical in achieving better RD performance. Therefore, in future work we will explore means of improving the estimation of the residual noise signal.

6. REFERENCES

[1] T.M. Cover and J.A. Thomas, "Elements of information theory," *John Wiley & Sons, Inc.*, 1991.

[2] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. IEEE-93, pp. 71–83, 2005.

[3] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive codes for distributed source coding," *EURASIP Trans. SP*, vol. SP-86, pp. 3123–3130, 2006.

[4] X. Guo, Y. Lu, F. Wu, W. Gao, and S. Li, "Free viewpoint switching in multi-view video streaming using Wyner-Ziv video coding," in *SPIE Trans. VCIP*. SPIE, 2006, vol. 6077, pp. 298 – 305.

[5] X. Zhu, A. Aaron, and B. Girod, "Distributed compression for large camera arrays," in *IEEE Proc. SSP*. IEEE, 2003, pp. 30–33.

[6] A. Aaron, P. Ramanathan, and B. Girod, "Wyner-Ziv coding of light fields for random access," in *IEEE Proc. MSP*. IEEE, 2004, pp. 323–326.

[7] M. Salmistraro, J. Ascenso, C. Brites, and S. Forchhammer, "A robust fusion method for multiview distributed video coding," *EURASIP Trans. ASP*, vol. ASP-2014, pp. 174, 2014.

[8] C. Liu, *"Beyond pixels: exploring new representations and applications for motion analysis"*, Ph.D. thesis, Massachusetts Institute of Technology, 2009.

[9] X. Huang and S. Forchhammer, "Improved side information generation for distributed video coding," in *IEEE Proc. MSP*. IEEE, 2008, pp. 223–228.

[10] H. P. Cong, S. Perry, and X. HoangVan, "A low complexity Wyner-Ziv coding solution for light field image transmission and storage," in *IEEE Proc. BMSB*, 2019, pp. 1–5.

[11] M. Ouaret, F. Dufaux, and T. Ebrahimi, "Fusion-based multiview distributed video coding," in *ACM Proc. VSSN*, 2006, pp. 139–144.

[12] X. Huang, C. Brites, J. Ascenso, F. Pereira, and S. Forchhammer, "Distributed video coding with multiple side information," in *IEEE Proc. PCS*. IEEE, 2009, pp. 1–4.

[13] X. Huang, L. L. Rakêt, H. Van Luong, M. Nielsen, F. Lauze, and S. Forchhammer, "Multi-hypothesis transform domain Wyner-Ziv video coding including optical flow," in *IEEE Proc. MSP*. IEEE, 2011, pp. 1–6.

[14] T. Maugey, W. Miled, M. Cagnazzo, and B. Pesquet-Popescu, "Fusion schemes for multiview distributed video coding," in *IEEE Proc. EUSIPCO*. IEEE, 2009, pp. 559–563.

[15] N. K. Kalantari, T. C. Wang, and R. Ramamoorthi, "Learning-based view synthesis for light field cameras," *ACM Trans. on Graphics*, vol. 35, no. 6, pp. 1–10, 2016.

[16] P. P. Srinivasan, T. Wang, A. Sreelal, R. Ramamoorthi, and R. Ng, "Learning to synthesize a 4D RGBD light field from a single image," in *IEEE Proc. ICCV*, 2017, pp. 2243–2251.

[17] J. Navarro and N. Sabater, "Learning occlusion-aware view synthesis for light fields," *arXiv preprint arXiv:1905.11271*, 2019.

[18] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. CSVT*, vol. 22, pp. 1649–1668, 2012.

[19] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The DISCOVER codec: architecture, techniques and evaluation," in *IEEE Proc. PCS*. IEEE, 2007.

[20] X. Huang and S. Forchhammer, "Cross-band noise model refinement for transform domain Wyner-Ziv video coding," *EURASIP Trans. SPIC*, vol. SPIC-27, pp. 16–30, 2012.

[21] W. Ryan, "An introduction to LDPC codes," *CRC Handbook for Coding and Signal Processing for Recording Systems*, 2004.

[22] D. Kubasov, J. Nayak, and C. Guillemot, "Optimal reconstruction in Wyner-Ziv video coding with multiple side information," in *IEEE Proc. MMSP*. IEEE, 2007, pp. 183–186.

[23] M. Rerabek and T. Ebrahimi, "New light field image dataset," in *IEEE Proc. QoMEX*. IEEE, 2016.

[24] JPEG PLENO, "Light field coding common test conditions," ISO/IEC JTC 1/SC29/WG1, JPEG, Vancouver, Canada, 2018.

[25] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *IEEE Proc. CVPR*. IEEE, 2013, pp. 1027–1034.