

# COMPRESSION OF UNFOCUSED PLENOPTIC IMAGES USING A DISPLACEMENT INTRA PREDICTION

*Yun Li, Roger Olsson, and Mårten Sjöström.*

Dept. of Information and Communication Systems  
Mid Sweden University  
SE-851 70 Sundsvall Sweden  
Roger.Olsson@miun.se

## ABSTRACT

Plenoptic images are one type of light field contents produced by using a combination of a conventional camera and an additional optical component in the form of microlens arrays, which are positioned in front of the image sensor surface. This camera setup can capture a sub-sampling of the light field with high spatial fidelity over a small range, and with a more coarsely sampled angle range. The earliest applications that leverage on the plenoptic image content is image refocusing, non-linear distribution of out-of-focus areas, SNR vs. resolution trade-offs, and 3D-image creation. All functionalities are provided by using post-processing methods. In this work, we evaluate a compression method that we previously proposed for a different type of plenoptic image (focused or plenoptic camera 2.0 contents) than the unfocused or plenoptic camera 1.0 that is used in this Grand Challenge. The method is an extension of the state-of-the-art video compression standard HEVC where we have brought the capability of bi-directional inter-frame prediction into the spatial prediction. The method is evaluated according to the scheme set out by the Grand Challenge, and the results show a high compression efficiency compared with JPEG, i.e., up to 6 dB improvements for the tested images.

*Index Terms*— Light field, plenoptic, HEVC, B-coder

## 1. INTRODUCTION

The intensity and the direction of light can be represented by a light field, which can be sub-sampled and captured by using plenoptic cameras. Based on the captured images, refocused images along different depth planes and multi-view images are possible to be rendered through post-processing. This is done by algorithmically combining the Elementary Images (EI) from the captured sensor image. Nevertheless, such a sampling of the light field will create redundant information in the image and manifest as cross-correlation between

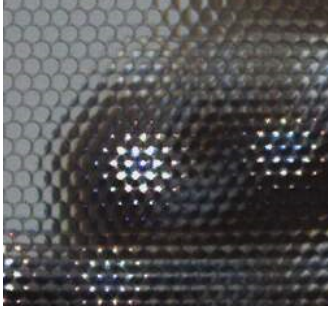
EIs. The video coding standard High Efficiency Video Coding (HEVC) is very efficient at reducing redundancy in conventional two-dimensional images using its intra-prediction tool-set, but was not intended to exploit the type of correlation present in plenoptic images [1]. For that the compression tools used to reduce temporal redundancy may be more suitable as the set of EIs (and any other transformation forms of the plenoptic image), to a large degree, presents a certain degree of correlation. The question addressed in this paper is to what extent competing compression efficiency can be obtained by using spatial displacement intra prediction with more than one hypothesis for the coding of plenoptic 1.0 images.

Gabriel Lippmann introduced the first plenoptic camera in 1908 [2]. However, the commercially available plenoptic camera was firstly produced by Lytro, Inc. founded by Ng et al. in 2006 [3] [4]. Lytro camera captures the distribution of light rays as described by the light field. This capability is achieved by putting a microlens array in front of the photo sensor. Because the focal plane of the microlens is on the photo sensor, only angular information is captured in each EI. This camera set-up is named plenoptic 1.0 system. In addition to this system, focused plenoptic cameras [5] have also been devised as an alternative to capture the light field. In images captured by focused plenoptic cameras, each EI is essentially a small cropped multi-view image from a specific viewing angle. As a contrast, in the content captured by plenoptic 1.0 systems, each EI captures the angular information of a spatial point, and there exists gaps between EIs that break the spatial continuity of a natural scene. Therefore, a coding scheme that explores the inter-EIs correlation can be beneficial for the coding of plenoptic 1.0 contents. A small part of this type of images is shown in Fig. 1.

In previous works, vector quantization approach was proposed [7] to utilize a subset of vectors to represent an entire vector space. The vectors are derived from the 4D representation of light field images. Moreover, in order to provide progressive coding scalability, Discrete Wavelet Transform (DWT) is commonly used [8] [9] for the coding. Another

---

This work has been supported by grant 20140200 of the Knowledge Foundation, Sweden.



**Fig. 1.** Part of a Plenoptic 1.0 image after demosaicing and devigneting [6].

alternative for an efficient compression of light field data is the predictive hybrid coding schemes, e.g., in [10], the EIs of images are transformed into Sub Images (SI) and the SIs are encoded by using the multi-view extension of H.264. In addition, Self Similarity (SS) modes have been proposed in [11] and [12] for H.264 and HEVC to reduce the redundancy inherited from the repetitive patterns of focused plenoptic images. The SS modes allow the encoder to encode the plenoptic images efficiently without knowing any camera geometrical information, which can be used to process the EIs.

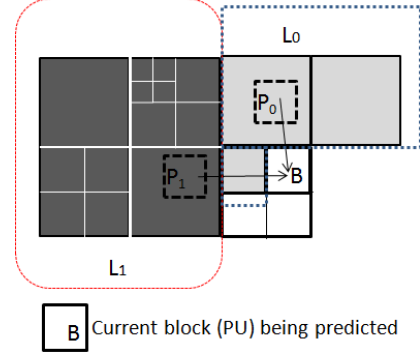
We have previously proposed an image B-coder for the coding of focused plenoptic contents [13][14]. The B-coder introduces the entire inter-prediction scheme in HEVC into the context of intra-prediction. This achieved a significant improvement over HEVC intra for the focused plenoptic images. In this work, the compression efficiency of the B-coder for the plenoptic 1.0 images is evaluated by investigating specific rate-distortion ratios for a set of plenoptic images.

The remainder of the paper is organized as follows. The B-coder is illustrated in Section 2. Test arrangements and evaluation criteria are presented in Section 3. Section 4 shows the results and analysis. We conclude the paper in Section 5.

## 2. PROPOSED METHOD

The image B-coder has been previously proposed in [13][14]. In addition to the conventional HEVC intra prediction, which only utilizes the adjacent reconstructed pixels for the prediction of a current block, i.e., current Prediction Unit (PU), the B-coder predicts the current PU by using blocks from its reconstructed neighbors. The details of the prediction are illustrated in Fig. 2.

The reconstructed blocks within a search range from the current PU are split into two parts (shown as dark and light gray in Fig. 2), which are loaded into the reference picture list  $L_0$  and  $L_1$ , respectively. The prediction candidates for the current PU are 1) the best matching block from  $L_0$ , 2) the best matching block from  $L_1$ , and 3)  $\frac{P_0+P_1}{2}$ .  $P_0$  and  $P_1$  are two blocks from  $L_0$  and  $L_1$ , respectively. More specifically,  $P_0$  is obtained by a refinement search around the best matching



**Fig. 2.** Bi-prediction within an image. Two parts in color light gray and dark gray are assumed as two reference pictures and available in the reference list  $L_0$  and  $L_1$ .

block from  $L_0$ , and  $P_1$  in the same way from  $L_1$ . The best among the three candidates is selected as reference blocks for the prediction of the current PU. The best is measured in terms of least rate-distortion.

The Advanced Motion Vector Prediction (AMVP) is also applied to the B-coder. In the case of images, the displacement vector for the current PU, i.e. the vector pointing to the reference block, is encoded predictively from its predictor, which is obtained from neighboring reconstructed PUs.

The proposed prediction is evaluated along with the conventional intra prediction in HEVC using a Rate Distortion Optimization (RDO) criterion. The best prediction mode is chosen for the current PU, and the residues from the prediction are encoded by following the original HEVC coding scheme.

## 3. TEST ARRANGEMENT AND EVALUATION CRITERIA

The test was conducted according to the requirement of the ICME 2016 Grand Challenge: Light-Field Image Compression [6]. The proposed B-coder was applied to the 12 images and encoded using a set of four specified compression ratios: 10:1, 20:1, 40:1, and 100:1. We evaluated the resulted quality of the compressed images by using a four-point PSNR. The proposed B-coder was configured using the "Low delay-Main" setting in JCTVC-L1100 [15], and the rate control algorithm of HEVC for the B-coder was switched on with equal bit allocation to meet the bit rate target.

The images were padded to the size of an integer number times of 64 for the encoding. In the evaluation process, the bit stream was then decoded, and the image cut to its original size. This was followed by the rendering of light field structure by using the LFTtoolbox 0.4 [16][17]. The  $PSNR_{YUV_{mean}}$ ,  $PSNR_{Y_{mean}}$ , and  $PSNR_{YUV}$  for each view as well as  $PSNR$  for other channels were com-

puted according to the specification [6]. However, only  $PSNR_{Y_{UV}mean}$  and  $PSNR_{Y_{mean}}$  are shown in this paper, and the computation takes respective channels into account on the rendered views from the original images and the decoded images. A subset of the results from the proposed scheme were compared with JPEG and HEVC intra by using BD-PSNR [18] on the  $PSNR_{Y_{mean}}$  component. The HEVC reference encoder version 11 was used for the testing with Quantization Parameters (QPs) 22, 27, 31, 37.

#### 4. RESULTS AND ANALYSIS

The actual bit rates for the four tested bit rate points according to the compressed ratio (R1, R2, R3, and R4) are obtained from the encoding as shown in Tab. 1. It can be observed that the rate control for the B-coder does not produce the exact required bit rate, and the accuracy is declining with decreasing bit rates. The compression ratio is presented in Tab. 2. The lack of exactly matching the stipulated compression ratios is a result of the rate control process in the HEVC encoder that our B-coder is built upon.

**Table 1.** Encoded bit rate in bytes

Image ID	Coded bit rate			
	R1	R2	R3	R4
I01	5317653	2734292	1409699	592316
I02	3872536	2504618	1235703	517482
I03	5105787	2521350	1349116	673750
I04	5095566	2520762	1368424	646055
I05	5021668	2390953	1050402	423982
I06	5202174	2604127	1121786	150525
I07	4735782	2390410	1202195	503395
I08	5099375	2502790	850755	173046
I09	5794953	3065775	1628954	757972
I10	4053311	2539736	1191879	468550
I11	5228771	2627661	1311775	351541
I12	5180096	2594780	1131055	451689

The  $PSNR_{Y_{mean}}$  and the  $PSNR_{Y_{UV}mean}$  are illustrated for the four bit rate points in Tab. 3 and Tab. 4, respectively.

Tab. 5 and 6 shows the BD-BSNR/rate for the proposed B-coder compared with JPEG and HEVC intra, respectively. It is illustrated that the B-coder has achieved a bit rate reduction of over 70 percent for the tested images when compared with the JPEG anchor. In addition, the proposed method also outperforms HEVC with up to 20 percent bit rate reduction. The improvement is smaller for plenoptic 1.0 images than for focused plenoptic images [13][14] with respect to the B-coder. This is because each EI of an image from plenoptic 1.0 camera captures the angular information of a spatial point in a scene. Therefore the cross-EIs correlation is more dependent on the scene surface, especially after the post-processing with demosaicing and devigneting, which is the case for the

**Table 2.** Compression ratio

Image ID	Compression ratio			
	R1	R2	R3	R4
I01	9.8	19.0	36.8	87.6
I02	13.4	20.7	42.0	100.2
I03	10.2	20.6	38.4	77.0
I04	10.2	20.6	37.9	80.3
I05	10.3	21.7	49.4	122.3
I06	10.0	19.9	46.2	344.5
I07	11.0	21.7	43.1	103.0
I08	10.2	20.7	61.0	299.7
I09	9.0	16.9	31.8	68.4
I10	12.8	20.4	43.5	110.7
I11	9.9	19.7	39.5	147.5
I12	10.0	20.0	45.9	114.8

**Table 3.**  $PSNR_{Y_{mean}}$  in dB for corresponding bit rates

Image ID	$PSNR_{Y_{mean}}$			
	R1	R2	R3	R4
I01	43.7	41.3	39.0	36.1
I02	41.0	39.6	37.2	34.2
I03	41.8	39.3	36.9	34.3
I04	45.1	42.8	40.8	38.2
I05	39.3	37.8	36.5	34.8
I06	46.5	44.7	43.2	41.1
I07	42.3	40.5	39.1	37.4
I08	44.6	42.8	41.6	39.7
I09	41.3	39.4	37.4	35.1
I10	41.5	40.9	40.0	38.4
I11	39.2	37.7	36.7	35.3
I12	44.1	42.4	41.0	38.4

**Table 4.**  $PSNR_{Y_{UV}mean}$  in dB for corresponding bit rates

Image ID	$PSNR_{Y_{UV}mean}$			
	R1	R2	R3	R4
I01	42.3	40.3	38.3	35.7
I02	39.7	38.4	36.3	33.7
I03	40.1	37.8	35.7	33.5
I04	43.4	41.3	39.7	37.5
I05	38.6	37.2	36.1	34.6
I06	45.1	43.5	42.3	40.5
I07	41.2	39.5	38.1	36.6
I08	43.6	42.0	40.9	39.2
I09	40.1	38.4	36.7	34.7
I10	40.8	40.2	39.3	37.9
I11	38.1	36.7	35.7	34.4
I12	42.7	41.2	39.9	37.6

**Table 5.**  $BD - PSNR/rate$  for the B-coder compared with JPEG

Image ID	BD-PSNR	BD-rate
I01	5.45	-79.26
I02	6.11	-82.89
I03	4.40	-77.59

**Table 6.**  $BD - PSNR/rate$  for the B-coder compared with HEVC

Image ID	BD-PSNR	BD-rate
I01	0.94	-22.73
I02	0.72	-18.83
I03	0.25	-6.65

images using in the test. However, we observed that better compression efficiency can be obtained by using the B-coder for the raw images, i.e., images captured from the plenoptic 1.0 camera without post-processing.

## 5. CONCLUSION

In this paper, the compression efficiency of a previously proposed image B-coder has been evaluated with respect to plenoptic 1.0 images. The B-coder has introduced the entire inter-frame prediction scheme of HEVC into the spatial domain. The results in terms of PSNR for the plenoptic rendered images were plotted for four bit rate points of 12 images. The results show a high compression efficiency compared to JPEG and HEVC intra for the tested contents. The bit rate reduction for the proposed method can reach over 70 percent compared with JPEG and over 20 percent compared with HEVC intra.

## 6. REFERENCES

- [1] B. Bross, W.-J. Han, J.-R. Ohm, G.J. Sullivan, and T. Wiegand, "High efficiency video coding (HEVC) text specification working draft 10," *JCT-VC Document, JCTVC-L1003*, 2013.
- [2] Lippmann, G., "Épreuves réversibles donnant la sensation du relief," *J. Phys. Theor. Appl.*, vol. 7, no. 1, pp. 821–825, 1908.
- [3] Ren Ng, "Digital light field photography," *Doctoral thesis, Stanford University*, 2006.
- [4] Ren Ng, Marc Levoy, Gene Duval, Mark Horowitz, and Pat Hanrahan, "Light Field Photography with a Handheld Plenoptic Camera," *Stanford Tech Report CTSR*, 2005.
- [5] Todor Georgiev and Andrew Lumsdaine, "Focused plenoptic camera and rendering," *Journal of Electronic Imaging*, vol. 19, no. 2, pp. 021106, Apr. 2010.
- [6] M. Rerabek, T. Bruylants, T. Ebrahimi, F. Pereira, and P. Schelkens, "Icme 2016 grand challenge: Light-field image compression," *Call for proposals and evaluation procedure*, 2016.
- [7] M Levoy and Pat Hanrahan, "Light field rendering," *Proceedings of the 23rd annual conference on computer graphics and interactive techniques*, pp. 31–42, 1996.
- [8] Chuo-Ling Chang, Xiaoqing Zhu, Prashant Ramanathan, and Bernd Girod, "Light field compression using disparity-compensated lifting and shape adaptation.," *IEEE transactions on image processing*, vol. 15, no. 4, pp. 793–806, Apr. 2006.
- [9] X Dong, D Qionghan, and X Wenli, "Data compression of light field using wavelet packet," *ICME '04. 2004 IEEE International Conference*, pp. 1071–1074, 2004.
- [10] J. Dick, H. Almeida, L. D. Soares, and P. Nunes, "3D Holoscopic video coding using MVC," *2011 IEEE EUROCON - International Conference on Computer as a Tool*, pp. 1–4, Apr. 2011.
- [11] Caroline Conti, João Lino, and Paulo Nunes, "Improved Spatial Prediction for 3D Holoscopic Image and Video Coding," *Proc. of the European Signal Processing Conference (EUSIPCO)*, pp. 378–382, 2011.
- [12] Caroline Conti, Luis Ducla Soares, and Paulo Nunes, "Influence of self-similarity on 3D holoscopic video coding performance," *Proceedings of the 18th Brazilian symposium on Multimedia and the web - WebMedia '12*, p. 131, 2012.
- [13] Yun Li, Mårten Sjöström, Roger Olsson, and Ulf Jennehag, "Efficient Intra Prediction Scheme For Light Field Image Compression," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florence, Italy*, May 2014.
- [14] Y. Li, M. Sjoström, R. Olsson, and U. Jennehag, "Coding of focused plenoptic contents by displacement intra prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, pp. 1–1, 2015.
- [15] F Bossen, "Common test conditions and software reference configurations," *Joint Collaborative Team on Video Coding (JCT-VC), JCTVC-L1100*, 2013.
- [16] "LFtoolbox (EPFL)," [FTP://tremplin.epfl.ch](http://tremplin.epfl.ch), retrieved: 03, 2016.
- [17] "LFtoolbox0.4," <http://www.mathworks.com/matlabcentral/fileexchange/49683-light-fieldtoolbox-v0-4>, retrieved: 03, 2016.
- [18] Gisle Bjontegaard, "Calculation of average PSNR differences between RD-curves," *ITU-T VCEG-M33*, 2001.