Light Field Image Compression with Sub-apertures Reordering and Adaptive Reconstruction

Chuanmin Jia[1], Yekang Yang[1], Xinfeng Zhang[2], Shiqi Wang[3], Shanshe Wang[1], and Siwei Ma[1(✉)]

[1] Institute of Digital Media, Peking University, Beijing, China
swma@pku.edu.cn
[2] Rapid-Rich Object Search (ROSE) Lab, Nanyang Technological University, Singapore, Singapore
[3] Department of Computer Science, City University of Hong Kong, Kowloon Tong, Hong Kong

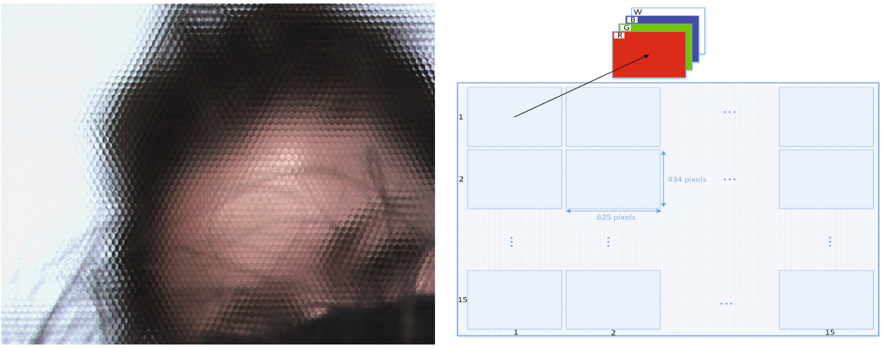
Abstract. Light field (LF) attracts tremendous attention due to its capability of recording the intensity of scene objects as well as the direction of the light ray, which also dramatically increases the amount of redundant data. In this paper, we explore the structure of the light field images, and propose a pseudo-sequence based light field image compression with sub-aperture reordering and adaptive reconstruction to efficiently improve the coding performances. In the proposed method, we firstly decompose the lenslet image into sub-aperture images, and then design an optimized sub-aperture scan order to rearrange them sequentially as a pseudo-sequence. Third, we take advantage of the state-of-the-art video codec to compress the pseudo-sequence by leveraging both intra- and inter-view correlations. Considering the interpolation and transform induced by the reconstruction procedure from sub-aperture images to lenslet image, we propose an enhanced reconstruction method by applying region-based non-local adaptive filters which extracts the non-local similarities for collaborative filtering to promote the quality of reconstructed lenslet images. Extensive experimental results show that the proposed method achieves up to 15.7% coding gain in terms of BD-rate.

Keywords: Light Field · Image compression
Sub-apertures arrangement · Adaptive reconstruction

1 Introduction

Light field (LF) image compression and the standardization progress of JPEG pleno [1] have attracted tremendous attentions recently. Light field is able to capture the intensity of objects and record the information of light rays. However, due to the dense sampling during LF imaging, large amount of redundant data

B. Zeng et al. (Eds.): PCM 2017, LNCS 10735, pp. 47–55, 2018.
https://doi.org/10.1007/978-3-319-77380-3_5



(a) (b)

Fig. 1. (a) Crop of an uncompressed lenslet image in RGB. (b) LF structure produced by the LF toolbox [10], dimension: $15 \times 15 \times 434 \times 625 \times 4$.

are introduced which show less hospitality for transmission, storage and further LF applications. To address this issue, high efficiency compression methods for LF image are urgently required.

Conventional light field image compression schemes [2–4] treat LF as a natural image with large amount of intra-redundancy, and optimizes intra prediction for redundancy removal, which do not dive into the structure of LF. In other words, better compression performance can be achieved by taking LF characteristics into consideration.

For better understanding of the LF characteristics, two concepts widely used in LF are introduced in this paragraph: elemental images and LF structures. As for elemental images, a micro lens array is placed in front of the image sensor, which optically subdivides the image sensor into a matrix of sub-images. We refer to such sub-images as elemental images (EI). Figure 1(a) shows a crop of lenslet image in RGB for the illustration of EI. The LF structures, however, is a 5-*Dimension* representation of image data: the first two dims denote the number of sub-apertures in row and column of lenslet image respectively while the following two dims indicate the resolution of each sub-aperture. The last dimension is the number of color channels. Given EI and LF structures, the sub-aperture images can be generated by resampling pixels who have the same horizontal and vertical location within each EI. The only difference among each sub-aperture is that different sub-aperture records different angles of light ray during imaging. Figure 1(b) illustrates the sub-apertures obtained by decomposition with LF MATLAB Toolbox [10].

Once sub-apertures are obtained, intra prediction in LF compression can be naturally converted to inter prediction since we have already transferred intra redundancy in the form of inter correlations. Hence, inter prediction in existing video codecs can be adopted to achieve better coding performance. Generally, all sub-apertures can be arranged as a pseudo-sequence for both intra- and inter-redundancy removal [9,11,13]. However, the compression performances of LF sub-apertures is sensitive to the sub-apertures arrangement mechanism, as different arrangement methods reveal different reference relationships and inter prediction capabilities. For the lenslet decomposition procedure from lenslet image to sub-apertures, the affine transform and interpolation make this process irreversible, which limit the quality of reconstructed lenslet image.

In this paper, we explore the structure of LF, and propose a pseudo-sequence based LF compression with sub-aperture reordering and adaptive reconstruction. To reduce the intra- and inter-view redundancies as much as possible, we first decompose the lenslet image into sub-apertures, and rearrange them into a pseudo-sequence according to the proposed optimal scan order for sub-apertures. Subsequently, the elaborately designed video coding methods, e.g., HEVC, can be directly utilized to reduce these redundancies via intra and inter predictions. To alleviate the distortions generated in lenslet image decomposition and reconstruction, we propose a region-based adaptive non-local filtering technique for the reconstructed lenslet image by searching non-local similar patches to arrange them as groups and designing low-pass filters for the groups in transform domain. Based on our experiments, the proposed algorithm achieves obvious improvement in terms of bit-rate reduction.

The rest of the paper is organized as follows. Section 2 briefly reviews related work in LF image compression. Section 3 introduces the proposed sub-aperture reordering scheme and non-local adaptive reconstruction for lenslet image. Extensive experimental results are reported in Sects. 4 and 5 concludes this paper.

2 Related Work

Many specific compression methods for LF images have been proposed in literatures, which can be roughly classified into two categories: (i) non-local intra prediction based compression [2–4] and (ii) sub-apertures based compression [11–13].

Intra Prediction Based Compression. Methods in this category focus on directly compressing the lenslet images. Li *et al.* [2] proposed an intra prediction algorithm by combining a sparse image set and its associated disparities for the prediction of highly correlated EIs. To further reduce the redundancy, Conti *et al.* [4] formulated a self-similarity model into HEVC by searching similar blocks in a nonlocal reconstructed region for prediction. Zhong *et al.* [3] proposed a super-pixel based L1 norm linear weighted intra prediction scheme for lenslet compression. However, the intrinsic LF structure of the above methods has been largely ignored.

Sub-apertures Based Compression. As for the sub-aperture based compression, LF image is firstly decomposed into sub-apertures by its calibration information, and then those sub-apertures are rearranged sequentially as a pseudo-sequence for further compression. Liu *et al.* [11] proposed a reference management and rate allocation scheme for sub-apertures. However, the complexity of reference scheme may limit the usage in practical LF coding scenario. Chen *et al.* [12] integrated the sparse coding into LF compression and selected key sub-apertures to train sparse dictionary for coding. This algorithm is efficient under low bit rate coding circumstance, which may lack its generalization ability. Zhao *et al.* [13] provided a hybrid arrangement order for sub-apertures by combing z scan order and U-shape order. Though promising performance has been achieved, in-depth performance analyses are still required to further leverage the prediction structure for coding efficiency improvement.

In this work, we propose a sub-apertures arrangement algorithm to deal with two main problems in LF image coding. First, a novel sub-apertures reordering scheme is proposed to deal with the lack of inter-prediction precision in conventional z-scan. Second, non-local adaptive reconstruction for lenslet image is also proposed to compensate the irreversible distortion induced by affine transform and interpolation during lenslet decomposition. Experimental results show that the proposed algorithm achieves obvious bit-rate reduction.

3 Sub-aperture Reordering and Adaptive Reconstruction

In this section, we first introduce the generation of sub-apertures and pseudo-sequence via proposed scan order. Then, we describe the details of proposed non-local adaptive reconstruction for lenslet image.

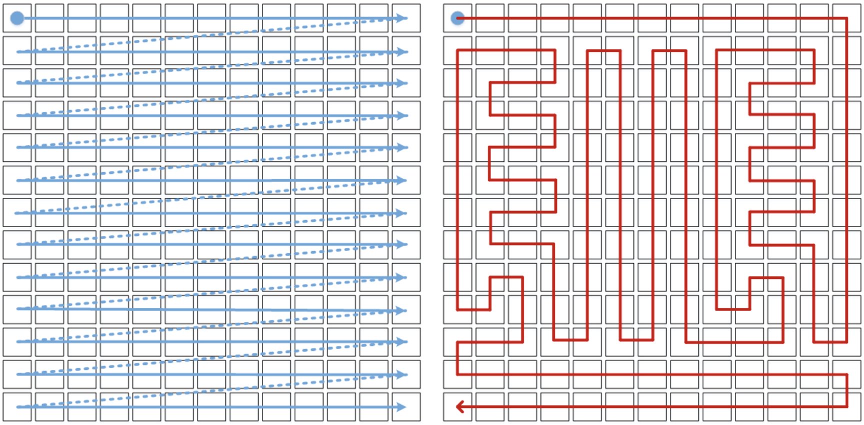
3.1 Sub-apertures Generation

To explore the correlations between sub-apertures, the lenslet images are converted into 5-D LF using MATLAB LF Toolbox [10]. The conversion process is listed as follows. (i) affine transform for lenslet images based on the camera calibration information, (ii) interpolation and (iii) resamping pixels in each EI to generate sub-apertures. The test LF images are chosen from EPFL LF dataset [14] with resolution 7728×5368 in YUV420 color space. After the decomposition process, the shape of 5-D LF structure is in $15 \times 15 \times 434 \times 625 \times 4$, which has been illustrated in Fig. 1(b). The resolution of each sub-aperture is 434×625.

3.2 Pseudo-sequence Generation

Since the scan order of subapertures will directly affect the inter-prediction efficiency. In this subsection, we design an optimized scan order for sub-apertures to generate the pseudo-sequence for compression. According to the Call for Proposal (CfP) for LF coding [1], we mainly consider the inner 13×13 views, which means the total frame numbers of generated pseudo-sequence is 169.

It is obvious that different sub-apertures reordering methods lead to different inter prediction capability such that coding efficiency varies for different reordering methods. To promote the coding efficiency, similar sub-apertures should be temporally adjacent for better inter prediction. Our proposed method gathers similar sub-apertures during pseudo-sequence generation. In view of this, a new sub-apertures arrangement mechanism is proposed, as shown in Fig. 2(b).



(a) (b)

Fig. 2. (a) The anchor sub-apertures scan order in [1]. (b) The proposed scan order.

3.3 Non-local Adaptive Reconstruction for Lenslet

This subsection describes the lenslet reconstruction from sub-aperture images. For lenslet reconstruction, our method conducts inverse decomposition to reconstruct lenslet from the coded pseudo sequence. We first convert the pseudo sequence back to 5-D LF structures by arranging each frame of pseudo sequence with proposed reordering method. Since the pseudo sequence only contains $13 \times 13 = 169$ frames. The outer most sub-apertures of 5-D LF structures are filled with their nearest sub-aperture. However, the transform during decomposition makes the lenslet reconstruction irreversible therefore the upper bound of reconstructed lenslet quality is limited. To address this issue, region based adaptive reconstruction is proposed into the method by drawing lessons from the high efficiency of in-loop filters in video coding [5,6].

Recently, non-local structure filters [7,8] provide significant coding performance promotion in video coding. We illustrate the proposed non-local adaptive reconstruction algorithm in Fig. 3 which can be categorized into four stages, patch extraction & matching, group construction, low-pass filtering and weighted reconstruction. First, the whole constructed lenslet image can be overlaply divided into small patches whose size is $\sqrt{B_s} \times \sqrt{B_s}$ (small red square). Then for each patch, we find K nearest neighbors according to the Euclidean distance between different image patches within a search window $W_s \times W_s$ (dashed blue square).

$$d(x_i, x_j) = \| x_i - x_j \|_2^2, \tag{1}$$

The patch group X_{G_i} can be generated by vectorizing K patches and forming a two-dimension matrix whose size is $B_s \times K$.

$$X_{G_i} = [x_{G_i,1}, x_{G_i,2}, ..., x_{G_i,K}], \tag{2}$$

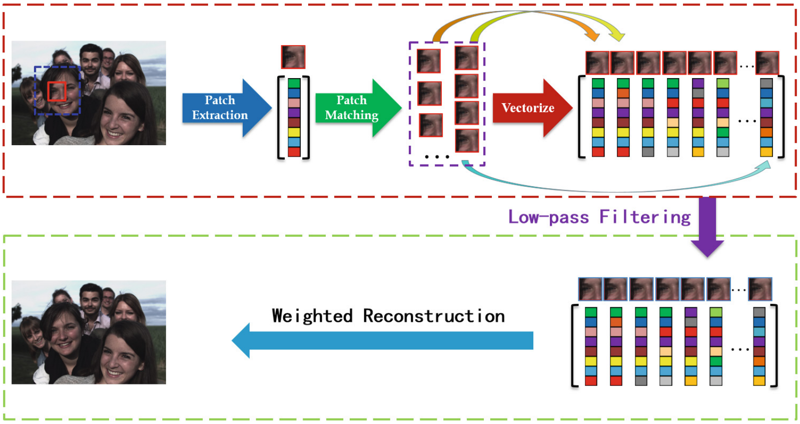


Fig. 3. The flowchart of proposed non-local adaptive reconstruction. (Color figure online)

After a singular value decomposition (SVD) procedure to patch group, the low-pass filtering can be performed with the singular values.

$$X_{G_i} = U_{G_i} \Sigma_{G_i} V_{G_i}^T, \tag{3}$$

where Σ_{G_i} is a diagonal matrix namely as singular value matrix. The low-pass filtering process is done by hard thresholding each element of the singular value matrix as follows,

$$\alpha_{G_i} = hard(\Sigma_{G_i,k}, \tau), \tag{4}$$

where k in Eq. 4 denotes the index for singular value in singular value matrix and τ is the hard threshold. For simplicity, we use the filtering threshold proposed in [8]. Since the filtering process is applied to all patch groups X_{G_i}, we finally average all group patches and reconstruct the filtered image. The non-local adaptive reconstruction is applied for both luma and chroma channels.

4 Experimental Results

In this section, we first describe the dataset and configurations we use in our experiment. Second, the performances of the proposed sub-aperture reordering and non-local adaptive reconstruction are provided. The bit-rate reduction for luma channel is reported in this paper.

4.1 Dataset

As mentioned in previous section, all lenslet images are selected from EPFL LF dataset [14] which contains 12 single frame lenslet images with resolution 7728×5368. The 5-D LF structure can be acquired by decomposing lenslet image with MATLAB toolbox [10]. According to JPEG call for proposal [1], inner $13 \times 13 \times 434 \times 625 \times 4$ sub-apertures out of the $15 \times 15 \times 434 \times 625 \times 4$ LF structure are re-arranged to generate the pseudo-sequence for compression. When reconstructing lenslet from pseudo-sequence, the outer sub-apertures of the $15 \times 15 \times 434 \times 625 \times 4$ LF structure are filled with its nearest sub-aperture.

4.2 Test Configuration

The configuration follows the common test condition (CTC) of HEVC [16]. All tests are conducted with random access (RA) configuration with four QP values: 22, 27, 32, 37. To investigate the performance of the proposed two coding tools, we set 5 experiments to illustrate the coding efficiency of reordering and non-local adaptive reconstruction, the details of which will be discussed in the next subsection. The BD-rate performances are calculated using the method described in [17]. The peak-signal-noise-ratio (PSNR) is calculated between original lenslet and reconstructed lenslet. We report the BD-rate in luminance channel.

Table 1. Rate distortion performances of proposed method

Test image	Test 1	Test 2	Test 3	Test 4	Test 5
I01 Bikes	−4.0%	−3.8%	−9.2%	−9.0%	−12.7%
I02 Danger de Mort	−3.1%	−7.1%	−8.5%	−8.6%	−15.0%
I03 Flowers	−5.8%	−2.2%	−4.8%	−5.2%	−10.7%
I04 Stone Pillars Outside	−3.9%	−4.3%	−9.0%	−9.5%	−13.0%
I05 Vespa	−1.5%	−1.4%	−12.4%	−12.3%	−13.6%
I06 Ankylosaurus & Diplodocus 1	−4.0%	−4.0%	−14.3%	−14.3%	−14.2%
I07 Desktop	−1.2%	−2.6%	−4.1%	−4.4%	−6.7%
I08 Magnets 1	−1.1%	0.0%	−10.8%	−10.8%	−10.8%
I09 Fountain & Vincent 2	−1.8%	−3.4%	−3.8%	−3.8%	−7.1%
I10 Friends 1	−2.1%	−1.0%	−13.8%	−13.9%	−15.7%
I11 Color Chart 1	−1.7%	−1.0%	−7.7%	−7.7%	−7.7%
I12 ISO Chart 12	−2.0%	−1.9%	−9.8%	−9.9%	−11.6%
Average	**−2.7%**	**−2.7%**	**−9.0%**	**−9.1%**	**−11.6%**

4.3 Evaluation

This subsection provides the details of each experiment in Table 1. Test 1 shows the performance of proposed region-based non-local adaptive reconstruction with z scan for sub-apertures reordering, in which the anchor of test 1 is z scan for sub-apertures and no non-local adaptive reconstruction to the reconstructed lenslet. As for test 2, the anchor is proposed scan order with the non-local adaptive reconstruction turning OFF, while the proposed method is sub-apertures with proposed scan order (in Fig. 2(b)) with the non-local adaptive reconstruction turning ON. From the first 2 tests, we can learn that the proposed non-local adaptive reconstruction filters bring similar performances when different sub-apertures scan order is used. The average coding performances for test 1 and test 2 are 2.7% and 2.7% respectively which show the robustness of proposed non-local adaptive reconstruction algorithm.

For test 3 and test 4, these two tests provide the coding performances of proposed reordering scheme for sub-apertures. The anchor of test 3 is z scan while the proposed method of test 3 is with proposed scan order. It is worth noting that the adaptive reconstruction is turned OFF in test 3. Further, to show the performance of proposed scan order when non-local reconstruction turns ON. We conduct test 4. The anchor of test 4 is reordering all sub-apertures with z scan while the proposed method for test 4 is with proposed scan order. Both of the anchor and proposed methods in test 4 are with non-local adaptive reconstruction applied to the reconstructed lenslet. Generally speaking, the proposed scan manner achieves 9.1% bit-rate saving with adaptive reconstruction while 9.0% bit-rate can be reduced when there is no adaptive reconstruction. Similar to the first two tests, the test 3 and test 4 also provides the robustness for proposed scan order for sub-apertures.

For the last evaluation, test 5 shows the combination performances of proposed reordering scheme and non-local adaptive reconstruction. The anchor of test 5 is reordering all sub-apertures with z scan and without non-local adaptive reconstruction to reconstructed lenslet. Up to 15.7% bit-rate reduction can be achieved when applying both the proposed scan order and adaptive reconstruction. It is worth noting that the performance of reordering and non-local adaptive reconstruction is addable, which means that further coding performances can be achieved by using both of the two coding tools.

5 Conclusion

In this paper, we propose a pseudo-sequence based light field image compression algorithm by reordering sub-apertures as well as applying the adaptive reconstruction filters for lenslet to promote the reconstruction quality. The proposed compression method first utilizes the HEVC codec to compress the pseudo-sequence, which is constructed by decomposing the lenslet image into sub-apertures and rearranging them according to a special designed order. Then, we propose a region-based non-local filters to further promote the quality of reconstructed lenslet image according to LF structures. Experimental results demonstrate that our proposed scheme can obviously improve the quality of the reconstructed lenslet LF images, and achieve significant coding gains.

Acknowledgement. This work was supported in part by the National High-tech R&D Program of China (863 Program, 2015AA015903) National Natural Science Foundation of China (61632001, 61571017, 61421062), and the Top-Notch Young Talents Program of China, which are gratefully acknowledged.

References

1. JPEG Pleno Final Call for Proposals on Light Field Coding. https://jpeg.org/items/20170208_cfp_pleno.html
2. Li, Y., Sjostrom, M., Olsson, R., Jennehag, U.: Efficient intra prediction scheme for light field image compression. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 539–543 (2014)
3. Zhong, R., Wang, S., Cornelis, B., Zheng, Y., Yuan, J., Munteanu, A.: L1-optimized linear prediction for light field image compression. In: 2016 IEEE Picture Coding Symposium, pp. 1–5 (2016)
4. Conti, C., Nunes, P., Soares, L.D.: HEVC-based light field image coding with bi-predicted self-similarity compensation. In: 2016 IEEE International Conference on Multimedia and Expo Workshops, pp. 1–4 (2016)
5. Zhang, X., Xiong, R., Ma, S., Gao, W.: Adaptive loop filter with temporal prediction. In: 2012 IEEE Picture Coding Symposium (PCS), pp. 437–440 (2012)
6. Zhang, X., Xiong, R., Lin, W., Zhang, J., Wang, S., Ma, S., Gao, W.: Low-rank based nonlocal adaptive loop filter for high efficiency video compression. IEEE Trans. Circ. Syst. Video Technol. (2016)
7. Ma, S., Zhang, X., Zhang, J., Jia, C., Wang, S., Gao, W.: Nonlocal in-loop filter: the way toward next-generation video coding? IEEE Multimed. **23**(2), 16–26 (2016)
8. Zhang, J., Jia, C., Ma, S., Gao, W.: Non-local structure-based filter for video coding. In: 2015 IEEE International Symposium on Multimedia, pp. 301–306 (2015)
9. Li, L., Li, Z., Li, B., Liu, D., Li, H.: Pseudo sequence based 2-D hierarchical reference structure for light-field image compression. arXiv preprint arXiv:1612.07309
10. Dansereau, D.G.: Light Field Toolbox for Matlab. Software Manual (2015)
11. Liu, D., Wang, L., Li, L., Xiong, Z., Wu, F., Zeng, W.: Pseudo-sequence-based light field image compression. In: 2016 IEEE International Conference on Multimedia and Expo Workshops, pp. 1–4 (2016)
12. Chen, J., Hou, J., Chau, L.P.: Light field compression with disparity guided sparse coding based on structural key views. arXiv preprint arXiv:1610.03684
13. Zhao, S., Chen, Z., Yang, K., Huang, H.: Light field image coding with hybrid scan order. In: 2016 IEEE Visual Communications and Image Processing, pp. 1–4 (2016)
14. Rerabek, M., Yuan, L., Authier, L.A., Ebrahimi, T.: EPFL light-field image dataset. ISO/IEC JTC 1/SC 29/WG1 (2015)
15. Tsai, C.-Y., et al.: Adaptive loop filtering for video coding. IEEE J. Selected Top. Sig. Process. **7**(6), 934–945 (2013)
16. Bossen, F., Flynn, D., Shring, K.: HEVC reference software manual. JCTVC-D404, Daegu, Korea (2011)
17. Bjontegaard, G.: Calculation of average PSNR differences between RD-curves. ITU SG16 Document VCEG-M33, April 2001
18. Girod, B.: The efficiency of motion-compensating prediction for hybrid coding of video sequences. IEEE J. Sel. Areas Commun. **5**(7), 1140–1154 (1987)