유사 비디오 시퀀스 기반의 라이트필드 영상 부호화를 위한 움직임 탐색 영역 제한

*임종훈, Vinh Van Duong, Thuc Nguyen Huu, 전병우
성균관대학교 전자전기컴퓨터공학과
e-mail: {*yjh6522, duongvinh, thuckechsu, bjeon}@skku.edu

Limiting Motion Search Range for the Pseudo Video Sequence-based Light Field Image Coding

*Jonghoon Yim, Vinh Van Duong, Thuc Nguyen Huu, and Byeungwoo Jeon
Department of Electrical and Computer Engineering
Sungkyunkwan University

Abstract

The large data volume of light field (LF) image has motivated much research on how to compress the data volume more efficiently. One of the approaches is to compress LF images after representing them in the form of pseudo video sequence. In this way, the pseudo temporal redundancy between views can be exploited by motion estimation and compensation. Based on our observation that images obtained by LF cameras have small range of disparity values between adjacent views, we propose to limit the motion search range to reduce the time complexity of motion estimation. Our experimental results show that a smaller motion search range reduces the encoding time while not affecting the bitrate of H.266/VVC much.

1. Introduction

Light Field (LF) images can provide various functionalities that conventional 2D images cannot such as depth estimation, multi-view rendering, post-capture refocusing, rapid scan-less volumetric microscopy and saliency detection [4]. Under strong application prospective for immersive applications, Moving Picture Experts Group (MPEG) has been paying attention on 3D graphics and 3D display based on light fields [11] already. However, LF images have an issue of high data volume. For example, the dimension of the raw image captured by Lytro Illum LF camera [12] is 7728 x 5368 pixels, corresponding to 8K resolution (76804320) and consumes 318 megabytes without compression assuming 8 bits per pixel. There has been various research carried out to efficiently compress the LF images [5]. M. B. Carvalho *et al.* applied 4D-discrete cosine transform (DCT) to exploit the high dimensional redundancy within and across the multiple light field views [13]. M. Rizkallah *et al.* applied a view synthesis scheme which allowed the LF to be reconstructed with high quality even with a small number of views [6]. V. V. Duong *et al.* tested the intra coding tools in H.265/HEVC and H.266/VVC to encode a LF image in the lenslet format [7].

Recently, there are several research works on LF compression focused on converting the LF image into a pseudo video sequence (PVS) by stacking the multiple views of a LF as illustrated in Figure. 1 [4]. In view of this, in this paper, we mainly address compressing the LF by feeding the PVS to existing video codecs. Section 2 explains our observations and motivations. We look at the experimental result in Section 3. Finally, concluding remark is given in Section 4.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2020R1A2C2007673).

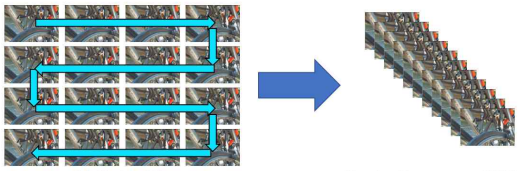

Figure 1. Formation of PVS by stacking multiple SAIs in an LF image. This figure illustrates the process of creating PVS by connecting the PVS in a zig-zag scan order.

2. Observations and Motivations

By rearranging the pixels of a lenslet image captured by the LF camera, we can obtain multiple sub-aperture images (SAI) which carry scene information obtained from different viewpoints [1]. When we capture a scene from different viewpoint, the objects appear at different coordinate on the images, and the pixel coordinate difference of the object across different views is referred to as the disparity [8]. We can create a PVS by connecting multiple SAIs. Here, the SAIs corresponding to slightly different viewpoints are aligned along the pseudo temporal axis. When we compress the PVS by feeding it to a conventional video encoder, the correlation between each frame is to be exploited by inter coding techniques and the disparity is captured by motion vectors which records the displacement of the scene object between different frames. As stated in [9], disparity between the adjacent views does not exceed [-10, 10] pixels for the scene captured by LF cameras. We also observe that the motion vectors of a PVS of a light field "Bikes" [1] tend to be distributed in the region of certain radius as shown in Figure 2. In this paper, we propose to use a smaller motion search range (SR) when encoding the LF PVS.

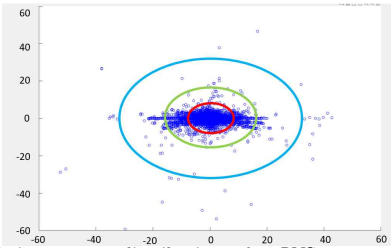

Figure 2. Motion vector distribution of a PVS converted from the light field "Bikes" [1] encoded in Low-Delay P configuration at QP 22. A PVS is created by stacking SAIs in a zig-zag scan order as illustrated in Figure 1. Red, green and blue circles indicate the region with the radius of 8, 16, and 32, respectively.

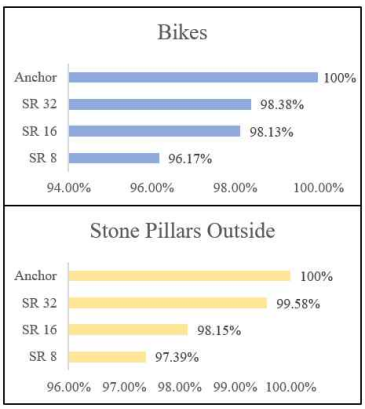

Figure 3. Relative encoding time complexity ratio of different motion SRs.

3. Experiments

This section provides the experimental conditions and results. There are many ways to create PVS out of LF such as scanning the SAIs in the zig-zag order, z-order, or spiral order [10]. However, it is stated in [4] that smoother SAI rearrangement tends to have better coding performance, so we chose the zig-zag scan order as shown in Figure 1. We used two light fields "Bikes" and "Stone Pillars Outside" from EPFL dataset [1] which contains scenes captured by Lytro Illum plenoptic camera. For both light fields, central 99 views are cropped from 1515 views to avoid extreme vignetting effects at border views as described in [14]. When making PVS, each SAI in LFs is converted to YUV 420 format before being encoded. We carried out an experiment with VTM-16.0 [3] on the desktop having intel-i5-8600K CPU @ 3.60 GHz and 16 GB RAM under Low Delay P (LDP) configuration. LDP is chosen to better exploit the disparity between adjacent SAIs in light fields. The quantization parameter (QP) is set up at 22, 27, 32, and 37. Against the VTM-16.0 anchor which estimates the motion in TZ-search pattern with motion SR of 64, we compared the encoding performance of PVSs in terms of the BD-Rate [2] and encoding time complexity when motion SR is set equal to 8, 16, and 32. A smaller value of BD-Rate represents better coding efficiency.

As one can see in Figure 3, we can see that the encoding time complexity can be reduced up to 4% by using a smaller SR. Also, as shown in Table 1, the average BD-Rate tends to slightly increase as SR gets smaller. This result demonstrates that a smaller SR is sufficient in coding the LF PVS.

Table 1. BD-Rate (Y-PSNR) performance (anchor: VVC)

Search Range	8	16	32
Bikes	0.05%	0.00%	0.03%
Stone Pillars Outside	0.00%	0.11%	0.08%
Average	0.02%	0.06%	0.06%

4. Conclusions

In this paper, we limited the motion search range for coding the PVS of LF images. Our experimental results showed that a smaller motion SR reduced the encoding time complexity and slightly increased the BD-Rate. In the future, we plan to study more efficient LF PVS coding scheme such as developing a novel motion search method and a new model for the disparity.

References

[1] S. Wanner, et al., "Datasets and benchmarks for densely sampled 4d light fields," in *Proc. Int. Symposium on Vision, Modeling and Visualization (VMV)*, 2013, pp. 225-226.

[2] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," ITU-T Q.6/SG16 VCEG 13th Meeting VCEG-M33, 2001.

[3] (online) VVC reference software, available at https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM

[4] 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, 2017.

[5] C. Conti, "Dense light field coding: a survey," in *IEEE access*, vol. 8, pp. 49244-49284, 2020.

[6] M. Rizkallah, et al., "Graph-based transforms for predictive light field compression based on super-pixels," in Proc. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 1718-1722.

[7] V. V. Duong, et al., "Light field image compression using versatile video coding intra prediction," in *Proc. Korea Institute of Broadcasting and Media Engineering Summer Conf.*, 2019.

[8] N. Meng, et al., "Light field view synthesis via aperture disparity and warping confidence map," in *IEEE Trans. on Image Processing*, vol. 30, pp. 3908-3921, 2021.

[9] H. Jeon, et al., "Accurate depth estimation from a lenslet light field camera," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 547-1555.

[10] T. N. Canh, et al., "Boundary handling for video-based light field coding with a new hybrid scan order," in *Proc. Int. Workshop on Advanced Image Technology (IWAIT)*, 2019.

[11] MPEG-I, "MPEG-I Visual activities on 6DoF and light fields," *ISO/IEC JTC1/SC29/WG11*, Macao, China, N17285, 2017.

[12] R. Ng, :Light field photography with a hand-held plenoptic camera," *Technical report*, Stanford University, 2005.

[13] M. B. Carvalho, et al., "A 4d dct-based lenslet light field codec," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, 2018.

[14] L. Mignard-Debise, et al., "Light -field microscopy with a consumer light-field camera," in *Proc. IEEE Inc. Conf. on 3D Vision*, 2015, pp. 335-343.