라이트필드 영상 슈퍼픽셀 분할의 시점간 일관성 개선

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

Improving View-consistency on 4D Light Field Superpixel Segmentation

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

Abstract

Light field (LF) superpixel segmentation aims to group the similar pixels not only in the single image but also in the other views to improve the computational efficiency of further applications like object detection and pattern recognition. Among the state-of-the-art methods, there is an approach to segment the LF images while enforcing the view consistency. However, it leaves too much noise and inaccuracy in the shape of superpixels. In this paper, we modify the process of the clustering step. Experimental results demonstrate that our proposed method outperforms the existing method in terms of view-consistency.

1. Introduction

Superpixel segmentation is a grouping of similar pixels in images into a larger region. In computer vision and image processing fields, superpixel is one of the keys to reduce the computational complexity since it can lessen the number of processing units [6]. Recently, with the advent of light field (LF) cameras such as Lytro [7], users can have hundreds of multi-view images in a single capture. Due to its versatility, the demand for LF image processing is rising [8]. Considering the large data volume of the LF data and the small view change among light fields, *i.e.,* its huge data redundancy, LF superpixel segmentation is promising for making many LF processing algorithms computationally feasible and easier. There have been lots of works carried out to efficiently segment the LF images. X. Lv et al. [3] proposed a graph model to achieve optimal computational efficiency and segmentation accuracy of the LF superpixels. H. Zhu et al. [2] segmented the central view of the LF images and propagated the labels onto the other views using pre-calculated depth information. N. Khan et al. [1] utilized the edge slopes of the epipolar image (EPI) to enforce the view consistency of the LF superpixels.

(a) (b)

Figure 1. (a) shows the portion of central view (5,5) in the light field papillon in HCI dataset [5]. (b, Top) and (b, Bottom) shows the segmentation result of view (1,2) and (5,5) of papillon with the method in [1], respectively. Each color represents different superpixel label. As marked in circles, there is much noise in the result and superpixel shape does not match with the objects in the scene.

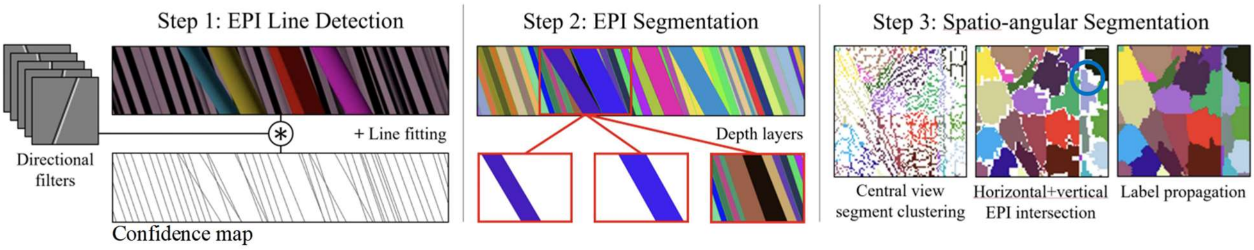
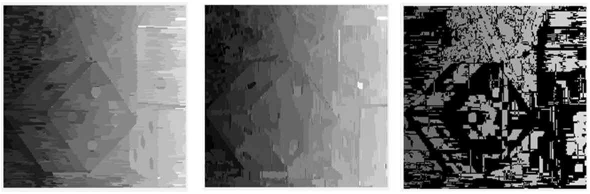


Figure 2. Illustration of the algorithm in [1]. In Step 1, N. Khan et al. find lines in vertical and horizontal EPIs by applying Prewitt edge filters. In Step 2, they match the pair of lines to form segmented regions in EPIs with explicit depth ordering. In Step 3, these EPI segments are clustered, and labels are propagated to fill the empty pixels without label, i.e., pixels in white area as marked with a blue circle. (Figures are taken from [1].)



(a) (b) (c)

Figure 3. In the figures, each gray-colored region denotes the different superpixel label in the central view of an LF image buddha. (a) shows the result of horizontal EPI clustering. (b) shows the result of vertical EPI clustering. As shown in (c), we discard the labels of pixels where horizontal and vertical clustering result do not match, and those pixels are colored in black.

In this paper, we are interested in the LF superpixel segmentation method in [1]. Its horizontal and vertical EPI segmentation groups the pixels in a central view with those in their corresponding locations in the other views. The method makes the superpixels view consistent, i.e., the shape of the superpixels does not abruptly change across the different views. Moreover, this technique can operate without relying on prior depth information and it can handle the occluded regions in non-central views. However, it leaves noise and outshoots on the superpixels on non-central views and the shape of the superpixels does not match with the objects in the scene as can be seen in Fig. 1. We observe that its clustering lets many pixels in non-central views be assigned with wrong labels. This is because horizontal and vertical EPI segmentation does not completely align as Fig. 3 demonstrates. Therefore, we propose to solve the problem by selecting either horizontal or vertical EPI segmentation in the clustering step.

The rest of the paper is organized as follows. Section 2 shows the superpixel segmentation framework in [1] and our proposed method. Based on that, we look at the experimental results in Section 3. Finally, concluding remark is given in Section 4.

2. Light Field Superpixel Segmentation Framework

We first look at each step of the algorithm in [1]. It consists of three steps as can be seen in Fig. 2.

1) Step 1 (EPI line detection): To effectively divide the EPIs into segments, N. Khan et al. [1] find the boundary of the segments by convolving the input EPI with Prewitt edge detecting filters. For each pixel in EPI, the maximum filter response value is taken as a confidence map which denotes the likelihood of a pixel belonging to an edge, and the index of the filter with the maximum response is taken as a disparity map. Finally, to robustly handle the occlusion problem, the detected lines are saved as the top and bottom intersections of the lines and the EPI boundaries.

2) Step 2 (EPI segmentation): The lines detected in the Step 1 are matched in a pair to form a segment in the EPI domain. To follow the scene depth order of the segments and enforce the occlusion awareness, bipartite graph matching is applied.

3) Step 3 (Spatio-angular segmentation via clustering): In the central view of an LF image, a clustering seed with its unique label number is spatially distributed with a uniform interval referred to as *superpixel size*. To cluster the EPI segments into a superpixel, pixels in each EPI segment are assigned a label of the seed which is closest to the centroid of the segment. As shown in Fig. 3(a) and (b), the result of clustering in horizontal and vertical EPI do not completely coincide. We intersect two directions of EPI segmentation, i.e., we discard the labels of the pixels

that do not match in both directions. Empty pixels in non-central views are assigned a label by propagating the label of the already assigned pixels. Finally, few remaining pixels are labeled by nearest neighbor assignment.

However, discarding the labels before the label propagation in Step 3 makes too many empty pixels, especially at the boundary of the superpixel (Fig. 3(c)). It causes the shape of the superpixel to disagree with the objects in the LF image and leaves noisy results as shown in Fig. 1(b). To handle the problem, instead of intersecting two directions of EPI segmentation, we choose either horizontal or vertical direction based on the view consistency metrics used in [1]: *self-similarity error* (SSE), and *numbers of labels per pixel* (LPP). To calculate the SSE, we project the center of superpixels from every view onto the central view using the ground truth disparity and calculate the average distance between the projections and a center of superpixels in the central view. A smaller error value indicates the better view consistency. To compute the LPP, we find how many different labels are assigned to a pixel in the central view on average by projecting the labels in non-central views to the central view via a ground truth disparity. A smaller number indicates the better view consistency. We choose the best direction as

$$(\text{Best direction}) = \underset{\text{mode} \in \{\text{Horr,Vert}\}}{\text{argmin}} \; w_S * SSE_{mode} + w_L * LPP_{mode} \quad (1)$$

where w_S and w_L are weight for the SSE, and LPP, respectively. Doing so drastically reduces the possibility of leaving empty pixels and assigning the wrong label. Further, to reduce the outshooting, we apply a median filter of size 1×3 or 3×1 before label propagation.

3. Experiments

We compare our method with the method in [1]. For experiment, we modify the source code publicly provided in [4]. We used seven scenes with a ground truth disparity map: buddha, buddha2, horses, medieval, mona, papillon, and stillLife from the HCI light field dataset [5]. They have 9×9 views with 768×768 spatial resolution except the horses having 1024×576 pixels in each view. Weights in Eq. (1) are set $w_S = 3.5$ and $w_L = 1$. Both algorithms are tested for the superpixel size of 15, 20, 25, 30, 35, 40, and 45.

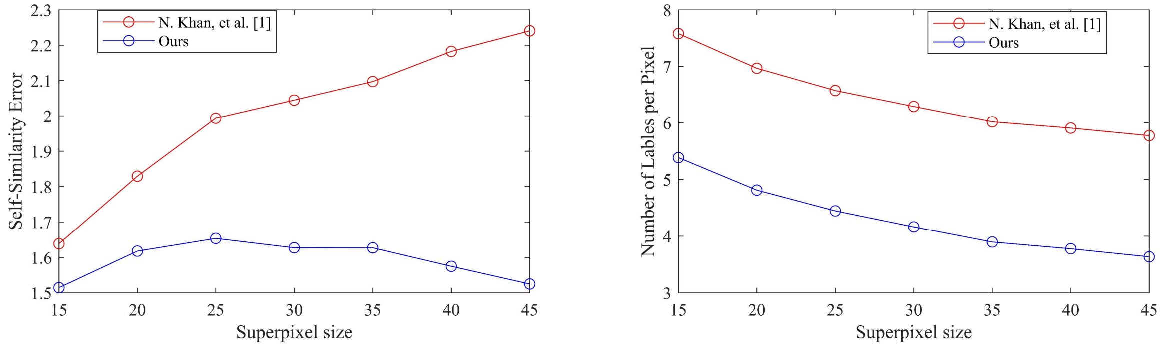
To quantitatively compare the view consistency of the LF superpixels of both methods, we use two evaluation metrics: SSE and LPP. Fig. 4 shows the metrics averaged for seven light fields. Our method outperforms the method in [1] in terms of both SSE and LPP for all superpixel sizes. We can also see that whereas SSE of [1] gets larger with increasing superpixel size, that of ours does not exceed a certain amount after the superpixel size gets bigger than 30. We also compare both methods visually in Fig. 5. Comparing the second and third columns, we see that our method effectively eliminated the noise. We also see that the perimeter of the superpixel became more consistent with the object on the scene, and the shape of the superpixels does not abruptly change as moving across the different views.

4. Conclusion

In this paper, we modified the clustering step of the light field superpixel segmentation method in [1] by selecting the best direction for EPI segmentation instead of intersecting two directions. The results showed that our proposed method outperform that of [1] in terms of view consistency metrics. In the future, we plan to further modify our method to make it optimized for other various post processing tasks such as view synthesis, object detection, and compression.

Acknowledgment

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)



(a) (b)

Figure 4. Comparison of numerical evaluation metrics for view consistency of light field superpixel segmentation between our proposed method and the method in [1]. (a) The average self similarity error (the lower the better) and (b) the average number of labels per pixel (the lower the better).

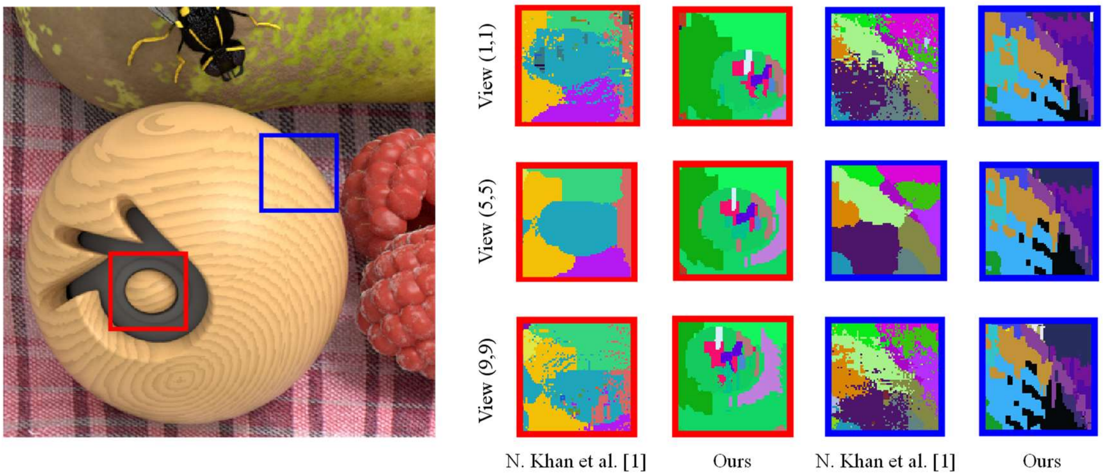


Figure 5. Light field superpixel segmentation results of [1] and our method. It shows some portions of the light field stillLife with the superpixel size of 30. Each color represents different superpixel label.

References

[1] N. Khan, et al., "View-consistent 4d light field superpixel segmentation," in *Proc. IEEE Int. Conf. on Computer Vision (ICCV)*, 2019, pp. 7810-7818.

[2] H. Zhu, et al., "4d light field superpixel and segmentation," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6709-6717.

[3] X. Lv, et al., "4d light field segmentation from light field super-pixel hypergraph representation," in *IEEE Trans on Visualization and Computer Graphics,* 2020.

[4] (Online) https://github.com/brownvc/lightfieldsuperpixels

[5] S. Wanner, et al., "Datasets and benchmarks for densely sampled 4d light fields," in *Proc. Int. Symposium VMV,* 2013, pp. 225-226.

[6] D. Stutz, et al., "Superpixels: an evaluation of the state-of-the-art," in *Computer Vision and Image Understanding*, vol. 166, pp. 1-27, 2018.

[7] N. Ren, et al., "Light field photography with a hand-held plenoptic camera," *Dissertation*, Stanford University, 2005.

[8] 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.