The application of deep learning on depth from multi-array camera

The Application of Deep Learning on Depth from Multi-Array Camera

Hossein Javidnia, Shabab Bazrafkan, Peter Corcoran

Department of Electronic and Electrical Engineering, College of Engineering, National University of Ireland, Galway
{h.javidnia1, s.bazrafkan1, peter.corcoran}@nuigalway.ie

***Abstract*-- Consumer-level multi-array cameras are a key enabling technology for next generation smartphones imaging systems. The present paper aims to analyze the accuracy of the depth estimation while using different camera combinations in a multi-array camera. This is done by providing a framework of deep neural networks to determine depth map from a sequence of images captured by a multi-array camera. Capturing depth information enables users to perform a range of post-capture edits such as refocusing, and creating a 3D model of any scene. Thus it is essential to calculate an accurate depth map while using the minimum computational resources.**

I. INTRODUCTION

Multi-array cameras such as Lytro [1] and PiCam [2] are becoming widely available in CE devices; and capturing accurate 3D information allows users to have a better and different experience with their personal digital photography.

In the past few years, personal photography has started to generate considerable interest due to inexpensive and still, high quality cameras. But capturing 3D information for most consumers has remained extravagant.

So far, plenoptic cameras such as Lytro [1] and Raytrix [3] are the most common and available consumer cameras enabling users to capture 3D information; however these cameras are not adjustable to be implemented in smartphones, which are the most popular imaging device in consumer electronics.

Based on InfoTrends' recent worldwide image capture forecast in "2016 U.S. Mobile Imaging End User Study" [4], consumers will approximately take 1.2 trillion photos in 2017 which shows the growth rate of 9% from the previous year. It is expected for 85% of these photos to be taken by mobile phones, 4.7% by tablets and 10.3% by other types of digital cameras.

Recently PiCam [2] and LinX [5] prototyped sets of cameras in a grid structure which enable users to have the same experience while capturing image using their smartphones as plenoptic cameras. A key problem with much of the literature in relation to this context is the expensive computation [6-8]. This issue rises due to the use of all the cameras in the array and it makes them not suitable for hand held consumer electronic devices.

This paper tackles these issues by taking advantage of deep neural networks. Five Siamese architectures [9] are designed to be trained on a synthetic dataset simulating images captured using a plenoptic camera. These experiments reveal a number of important aspects of depth calculation using multi-array cameras which will help us to answer the following questions:

1- Is it required to use all the cameras in the array to estimate depth map?
2- How efficient it is to combine horizontal and vertical displacement of the cameras in depth estimation?
3- How the variation of baselines can affect the output estimated depth?
4- Is deep neural network a good replacement for parallax computation?

II. ALGORITHMS & IMPLEMENTATION

A. Camera Array Structure

In this paper we present a matching CNN that outputs depth map from a set of images captures by multi-array camera.

The concentration in the state of the art is on processing the parallax to determine depth values from an array of images. Considering the parallax, the image features are shifted in different frames of the image array as a function of the 3D structure of the scene. What is known about depth estimation from multi-array cameras is largely based on the use of all of all the images in the array. That considerably increases the computational resources. The proposed solution examines the depth map obtained from different camera combinations rather than considering the whole array.

There are five experiments done in this paper by taking advantage of deep neural networks. The first network is designed to accept two images of an array of 9×9 camera. Fig. 1 illustrates the selected camera indices for the first experiment.

C1	C2	C3	C4	C5	C6	C7	C8	C9
C10	C11	C12	C13	C14	C15	C16	C17	C18
C19	C20	C21	C22	C23	C24	C25	C26	C27
C28	C29	C30	C31	C32	C33	C34	C35	C36
C37	C38	C39	C40	C41	C42	C43	C44	C45
C46	C47	C48	C49	C50	C51	C52	C53	C54
C55	C56	C57	C58	C59	C60	C61	C62	C63
C64	C65	C66	C67	C68	C69	C70	C71	C72
C73	C74	C75	C76	C77	C78	C79	C80	C81

Figure 1. Selected camera indices for the first experiment

C1	C2	C3	C4	C5	C6	C7	C8	C9
C10	C11	C12	C13	C14	C15	C16	C17	C18
C19	C20	C21	C22	C23	C24	C25	C26	C27
C28	C29	C30	C31	C32	C33	C34	C35	C36
C37	C38	C39	C40	C41	C42	C43	C44	C45
C46	C47	C48	C49	C50	C51	C52	C53	C54
C55	C56	C57	C58	C59	C60	C61	C62	C63
C64	C65	C66	C67	C68	C69	C70	C71	C72
C73	C74	C75	C76	C77	C78	C79	C80	C81

Figure 2. Selected camera indices for the second experiment

The second experiment is done by accepting two inputs from the first and last cameras. Fig. 2 presents the camera indices chosen for the second experiment. This experiment

reveals the importance of the baseline selection in depth estimation using deep neural networks.

The third network is designed to accept four images of the array as the inputs. This experiment uses the cameras located at the corners of the grid as illustrated in Fig. 3.

C1	C2	C3	C4	C5	C6	C7	C8	C9
C10	C11	C12	C13	C14	C15	C16	C17	C18
C19	C20	C21	C22	C23	C24	C25	C26	C27
C28	C29	C30	C31	C32	C33	C34	C35	C36
C37	C38	C39	C40	C41	C42	C43	C44	C45
C46	C47	C48	C49	C50	C51	C52	C53	C54
C55	C56	C57	C58	C59	C60	C61	C62	C63
C64	C65	C66	C67	C68	C69	C70	C71	C72
C73	C74	C75	C76	C77	C78	C79	C80	C81

Figure 3. Selected camera indices for the third experiment

By adding the central camera to the list of the inputs, the fourth observation is performed as shown in Fig. 4.

C1	C2	C3	C4	C5	C6	C7	C8	C9
C10	C11	C12	C13	C14	C15	C16	C17	C18
C19	C20	C21	C22	C23	C24	C25	C26	C27
C28	C29	C30	C31	C32	C33	C34	C35	C36
C37	C38	C39	C40	C41	C42	C43	C44	C45
C46	C47	C48	C49	C50	C51	C52	C53	C54
C55	C56	C57	C58	C59	C60	C61	C62	C63
C64	C65	C66	C67	C68	C69	C70	C71	C72
C73	C74	C75	C76	C77	C78	C79	C80	C81

Figure 4. Selected camera indices for the fourth experiment

Fig. 5 represents the last experiment considering four more cameras added to the previous model.

C1	C2	C3	C4	C5	C6	C7	C8	C9
C10	C11	C12	C13	C14	C15	C16	C17	C18
C19	C20	C21	C22	C23	C24	C25	C26	C27
C28	C29	C30	C31	C32	C33	C34	C35	C36
C37	C38	C39	C40	C41	C42	C43	C44	C45
C46	C47	C48	C49	C50	C51	C52	C53	C54
C55	C56	C57	C58	C59	C60	C61	C62	C63
C64	C65	C66	C67	C68	C69	C70	C71	C72
C73	C74	C75	C76	C77	C78	C79	C80	C81

Figure 5. Selected camera indices for the fifth experiment

Fig. 6 illustrates the framework employed in each experiment to estimate depth from a camera array.

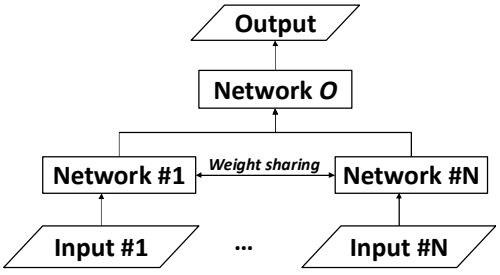

Figure 6. Framework employed by each experiment

B. Training Set

24 synthetic 4D light field image sets [10] are used for training and 4 sets for testing purposes. Each 4D set contains $9 \times 9 \times 512 \times 512 \times 3$ individual image.

III. Conclusions & Discussion

In this paper a framework is proposed to estimate depth from multi-array camera using deep neural network. Five different observations have been made based on the number of cameras and different baselines.

Table 1, shows the preliminary test error for the "Antonious" [10] set per experiment. Our preliminary testing suggests that there is only a very slight reduction in the error rate if more than 4 cameras are used. Detailed experiments and network architectures will be presented at ICCE 2018 and these should give a clear indication of the relative importance of horizontal/vertical separations and the trade-off in terms of image quality and error rate for increasing the number of cameras and the computational complexity.

Table 1. Test error for "Antonious" set with baseline = 100 mm and focal length = 100 mm

	Exp. #1	**Exp. #2**	**Exp. #3**	**Exp. #4**	**Exp. #5**
Test Error	0.00093	0.00091	0.00079	0.00083	0.00079

Acknowledgment

The research work presented here was funded under the Strategic Partnership Program of Science Foundation Ireland (SFI) and co-funded by SFI and FotoNation Ltd. Project ID: 13/SPP/I2868 on "*Next Generation Imaging for Smartphone and Embedded Platforms*".

References

[1] Lytro Redefines Photography with Light Field Cameras. Available: www.lytro.com/press/releases/lytro-redefines-photography-with-light-field-cameras

[2] K. Venkataraman, D. Lelescu, J. Duparre, A. McMahon, G. Molina, et al., "PiCam: an ultra-thin high performance monolithic camera array," ACM Trans. Graph., vol. 32, pp. 1-13, 2013.

[3] C. Perwass and L. Wietzke, "Single lens 3D-camera with extended depth-of-field," 2012, pp. 829108-829108-15.

[4] InfoTrends. (2016). 2016 U.S. Mobile Imaging End User Study. Available: store.infotrendsresearch.com/product_p/154700.htm [5]

Z. Attar, C. Aharon-Attar, and E. M. Wolterink, "System and Method for Imaging and Image Processing," ed: Google Patents, 2011.

[6] S. Zhang, H. Sheng, C. Li, J. Zhang, and Z. Xiong, "Robust depth estimation for light field via spinning parallelogram operator," Computer Vision and Image Understanding, vol. 145, pp. 148-159, 2016/04/01/ 2016.

[7] T. C. Wang, A. A. Efros, and R. Ramamoorthi, "Occlusion-Aware Depth Estimation Using Light-Field Cameras," in 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 3487-3495.

[8] H. G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. W. Tai, et al., "Accurate depth map estimation from a lenslet light field camera," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1547-1555.

[9] J. Zbontar and Y. LeCun, "Stereo matching by training a convolutional neural network to compare image patches," *J. Mach. Learn. Res.*, vol. 17, pp. 2287-2318, 2016.

[10] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldluecke, "A Dataset and Evaluation Methodology for Depth Estimation on 4D Light Fields," in Computer Vision – ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, pp. 19-34.