Analysis of Light Field Sampling Using Robots and CNNs

Carson Vogt

February 5, 2018

Abstract

It is the goal of this project to explore the use of CNNs to construct a light field partially collected by a robot. While other groups have used deep learning to create new images in a light field, none have applied it to the robotic collection use case which comes with its own set of challenges. It is my hope that by the end of this, there will be a demonstrable network that fills in a sparsely sampled, unstructured light field collected by a robot.

The Problem

There are currently no papers addressing the reconstruction of unstructured light fields. While there is one paper on the subject [1], construction of novel views is limited to barycentric interpolation.

While most light fields are captured using a light field camera, angular capture is limited to the camera itself. For situations where an adaptable capture system is required, a robot moving a camera makes sense as the available movements and locations for data capture are diverse, and each location is also given a position and orientation. Be it analysis of a structure that is difficult for a human to reach, or analysis of a subject in a hazardous environment, a robot collected light field makes a lot of sense. Therefore, exploring reconstruction of a partially collected, unstructured light field from a robot is a worthwhile pursuit.

Related Work

Several papers are especially applicable to this project: Probably the most similar, *DeepStereo: Learning to Predict New Views from the World's Imagery* uses two CNNs to generate a novel image. The input to this system is two posed images and a novel image pose, with the output being a novel view. However, the true input to the network is a plane sweep volume in an effort to increase training speed.

Of interest is the larger baseline between images than is seen in other methods presented. As well, rather than a stereo rig, the images are taken sequentially from a moving vehicle making them unstructured.

One of the downsides of using this method directly is the fact that it would not take advantage of any of the rest of the collected light field.

Learning-Based View Synthesis for Light Field Cameras This paper uses a similar CNN structure to the previous, however it takes four images as input and a novel image pose.

The results from the method in this paper have been very good, and the simple input is alluring. Much of the light field is utilized in training, however, in reconstruction only the four corner images are used as input.

The data for this method is very structured and the baselines are quite small as the data set is captured by a light field camera.

Light Field Reconstruction Using Deep Convolutional Network on EPI Uses a CNN to recover the high frequency details in a light field image after performing

an anti-aliasing blur operation.

This method requires only a single CNN, while the previous papers use two. It also does not seem to require as specific a sampling pattern, but it appears to only have been tested on highly structured data taken from a Lytro.

The authors suggest the maximum disparity for reasonable results is five pixels.

Unstructured Light Fields This paper is the original inspiration for this idea. In it, the authors describe a created system with which a user can move a camera around to collect images on a surface defined on a monitor. A densely sampled light field is collected and novel views are created via barycentric interpolation.

The advantage here is the unstructured nature of the data and the quick rendering time. No training is required.

Dense data is required for a reasonable reconstruction, and even then the reconstructions show many artifacts.

A Convolutional Network for Real-Time 6-DOF Camera Relocalization An interesting paper, though not necessarily as relevant and fairly simple. The network takes a posed image as the input, with a pose as the output.

This is a simpler application of the CNN, but the similarities don't seem dismiss-able. It may be interesting to explore the concept further in the future as the system put forward in the paper relies on structure from motion, and the others report a localization error on the order of meters.

Light Field Reconstruction Using Sparsity in the Continuous Fourier Domain This is not a paper that has come up in deep learning applications of light fields. However, aspects of it seem like they can be improved upon or utilized to limit the amount of required data.

Currently, other research papers focus on the 1D slicing technique used for the initialization of the reconstruction algorithm developed in the paper. This would in fact limit it to requiring a very structured dataset. However, further research into the subject may prove it to be useful in a deep learning application.

Progress to Date

Ready to Go

Have already acquired at least two large datasets of posed images taken in a traditional light field pattern (in this case a sphere). However, collecting for a cylindrical or planar case would not be difficult as a pipeline is in place to collect the images easily with the PR2.

A method is in place for extracting and comparing reconstructed images using the opencv $L2$ normal function.

The current or baseline reconstruction in place is the method used in [1], which uses a Delaunay triangulation followed by a barycentric interpolation Currently there is no pre or post filtering involved.

I have a working understanding of TensorFlow and could confidently build an adequate CNN for the problem, though determining the structure of the CNN is still to be determined.

Needs Work

Depending on the input method, it could take some time to implement the appropriate method, all are well documented.

Where to Submit

Immersive Mixed and Virtual Environment Systems - March 1
International Symposium on Mixed and Augmented Reality - March 15
Visual Analytics Science and Technology - March 21 (abstract)
International Journal of Robotics Research - April 1
British Machine Vision Conference - April 30
SIGGRAPH Asia - June 6
International Conference on Computational Photography - June 13
Winter Conference on Applications of Computer Vision - June 29
Virtual Reality Software and Technology - August 15

Stages

Basic Reconstruction of Unstructured Light Field Using CNN

This largely encompasses what was described above.

More Advanced Reconstruction

This could take a number of forms.

- Varying the input method. As the related work section shows, the input to the CNN is the least similar part of the individual methods using CNNs.

- Varying the CNN architecture.

- Take a closer look at applying aspects of [3] to the model. At least one of the papers, [4] ignores the method applied to [3] due to the constraints from the sampling pattern.

Minimization of Light Field Input Images

A further step of interest would be to determine what/which method yields the best results as a function of camera spacing. After, experiment to determine what about that method makes it superior, and how might it be improved upon.

Bibliography

[1] Davis, A., Levoy, M., and Durand, F., *Unstructured Light Fields* (2012). Eurographics 2012, no. 31.

[2] Flynn, J., Neulander, I., Philbin, J., and Snavely, N. *DeepStereo: Learning to Predict New Views from the World's Imagery* (2015).

[3] Kalantari, N. K., Wang, T. and Ramamoorthi, R., *Learning-Based View Synthesis for Light Field Cameras* (2016). SIGGRAPH Asia '16.

[4] Shi, L., Hassanieh, H., Davis, A., Katabi, D. and Durand, F., *Light Field Reconstruction Using Sparsity in the Continuous Fourier Domain* (2014). ACM Transactions on Graphics. no. 34.

[5] Wu, G., Zhao, M., Wang, L., Dai, Q., Chai, T., and Liu, Y., *Light Field Reconstruction Using Deep Convolutional Network on EPI* (2017). CVPR.