LaTeX Author Guidelines for CVPR Proceedings

Thomas Lauer
UC San Diego
9500 Gilman Drive, San Diego CA
tlauer@ucsd.edu

Winston Durand
UC San Diego
9500 Gilman Drive, San Diego CA
wdurand@ucsd.edu

Abstract

Light field photography allows for the capture of images from multiple perspectives using a single camera array or array of microlenses. However these come at a tradeoff between spatial resolution and angular resolution. As the size of the microlens elements increases, the resulting lightfield will have higher angular resolution which is useful for depth estimation, but will have lower spatial resolution for novel view synthesis. We will be reimplementing a paper [1] which uses a learning based method to synthesize views from light field camera data.

1. Introduction

A Lytro is a common light field camera which uses microlenses to capture multiple perspectives of a scene in one light field.

Many geometric methods such as Levoy *et al*. [2] require relatively high sample counts to ensure full coverage of the 4D lightfield, as the simple linear interpolation used by Levoy do not work if there are occluded regions or missing data.

Kalantari summarizes several existing methods for interpolating lightfields and their drawbacks. Many rely on high quality input images and well defined orientations between views, which are difficult to achieve with consumer light field cameras, and impossible with light fields captured by hand with cellphone cameras.

Wanner *et al*. [3] utilizes an optimization based approach, which calculates disparities using traditional computer vision as a preprocessing step. However, because the disparity estimation is independent from the loss function, it can't be optimized as part of the loss function.

Kalantari *et al*. [1] propose a method to interpolate between sparsely sampled sub-apertures. They use an 8×8 subset of the full 14×14 subaperture, since the edge pixels are often black. The four corner sub-aperture images are fed into a series of two networks, the first is used to predict disparity which is then used to warp the sample images to the final perspective. The second network then takes these warped images, along with some additional metadata, and blends them together to produce the final RGB image of the novel view. Keeping this process differentiable allows both CNNs to be trained at the same time, which lets the disparity estimator become tuned to work with the final CNN.

2. Our Proposal

Our goal is to reimplement this paper and additionally try to apply this technique to non-Lytro camera array captures. Further, we would like to investigate rendering novel views outside of the planar quadrilateral formed by the source images.

3. Milestones

Below are our milestones.

3.1. Load Lightfields Into CNN

Our first goal will be to load datasets into PyTorch, and set up our preprocessing pipeline. This will involve creating dataloaders to convert the interlaced Lytro image into a stack of individual images, extract training batches, and set up our boilerplate training loop.

3.2. Naive CNN Approach

After initial preprocessing, we aim to implement a single layer CNN similar to the Naive approach outlined in the paper. This is a single convolutional network which accepts the 4 sampled images along with their position information as input, and attempts to produce the novel view requested. As Kalantari *et al*. show, this should be able to create the novel view, but the result should be blurry and low detail. This is a good progress check to ensure we can actually train the network correctly.

We estimate that up to this point should take approximately a week.

3.3. Two-Part CNN With Disparity Estimation

The next step will be implementing the two-stage CNN architecture outlined by Kalantari *et al*. Instead of feeding the raw images from the sampled views, we apply a disparity warp to all the images at a set number of disparity levels. These are accumulated into a feature vector and used as input to the disparity prediction network. More details can be found in section 3.1 of Kalantari *et al*. [1].

3.4. Investigate Using Different Sampled Orientations

An additional approach we would like to consider to get more spread out of the Lytor capture data is to choose our input images in a diamond pattern rather than the 8×8 grid from the original paper.

3.5. Investigate Using Camera Array Datasets

As a followup, we intend to investigate how model trained in this manner is able to generalize to other camera array datasets, both where the views don't form a regular square and camera orientation varies.

3.6. Exploration of novel view outside capture bounds

Finally, we would like to experiment with changing the virtual camera's depth, moving it gradually away from the capture plane. A further path to test is the ability of our approach to esimate novel views which are still on the same plane as our reference views, but outside the bounds of the quadrilateral which they form.

We generally are classifying this milestone as ambition

4. Technology

We intend to implement this using PyTorch, because it has high performance, automatically handles calculating derivatives, and we have previous experience with it. Additionally, we will likely be using OpenCV or something similar for feature extraction in the preprocessing stage.

5. Datasets

Our training datasets will be Lytro captures since it represents a standardized input format which is widely available, for example from Stanford at http://lightfields.stanford.edu/LF2016.html. We chose to use

6. Questions to be Answered

How the disparity from the first CNN compares to disparity produced by traditional computer vision methods, such as OpenCV.

References

[1] Nima Khademi Kalantari, Ting-Chun Wang, and Ravi Ramamoorthi. Learning-based view synthesis for light field cameras. *ACM Transactions on Graphics (TOG)*, 35(6):1–10, 2016.

[2] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42, 1996.

[3] Sven Wanner and Bastian Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):606–619, 2014.