Light Field Reconstruction from Focal Stack based on Depth Estimation

Stock Sawasdee
Stanford University
paphop@stanford.edu

Abstract

Focal stack of an image can be used to reconstruct light field data. In this project, we reconstruct light field from focal stack using depth estimation. We proposed a new method of creating a depth map from focal stack by finding the label with highest sharpness score in a kernel.

I. Introduction

Unlike traditional photography where light is compressed to 2 dimensions on a film or a sensor of a camera, light field photography is an imaging technology that captures incoming light from different directions, resulting in 4-dimensional images. The 4 dimensional images from light field photography have many applications including digital refocusing and perspective shift; however, taking light field photographs can be costly. One popular approach to capture light field is to use plenoptic cameras where micro-lenses are placed in front of the sensor to create micro images from lights in different directions.

Instead of using specially designed cameras to generate light field images, we are interested in using focal stack images, which are the projection of 4-D light field onto 3-D light field, to reconstruct the 4-D light field. The approach in this project is to estimate depth of the image and apply plenoptic function to the pixels at each depth accordingly.

II. Related work

Levin et al. [1] proposed a method to reconstruct light field by considering focal stacks as a convolution of light field data and to generate light field, they find the deconvolution of the focal stacks. Mousnier et al. [2] suggested a different approach to solving the problem. They use focal stacks to estimate depth of the image and then back-project the focused regions of the scene to create perspective shifts. The algorithm for depth estimation used in [2] was graph cut algorithm from [3] by minimizing the sum of the sharpness score of each pixel and the smoothness of the variation of labels.

III. Method & Algorithm

3.1 Pipeline

The pipeline is adopted from what is discussed in [2], but instead of using images taken from a camera we first generate the focal stack images from light field dataset so that we can use them as a ground truth for comparison.

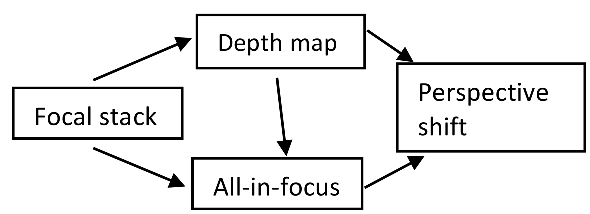

Figure 1: The pipeline of light field reconstruction based on depth estimation

Figure 1 shows an overview of the pipeline used in this project. Given focal stack images, we create a depth map by label the region based on gradient using algorithm discussed in the following section. Once the depth map is created, we can obtain an all-in-focus image by selecting pixels corresponding to the depth map from the focal stack.

To create a perspective shift or the light field, we select the focused regions at each depth and shift them according to the plenoptic equation (1).

$$L(x, y, u, v) = L_{u,v}(x - su, y - sv) \quad (1)$$

(x, y) represents coordinates on sensor plane and (u, v) represents coordinates on lens plane. s is the slope which is determined by the distance from the center of the epipolar images.

The shifted images are then filled in from the background first to the foreground.

3.2 Depth estimation

The depth estimation algorithm we proposed consists of two steps, generating blurred gradient field and sharpening the depth map by using highest gradient score.

First, we calculate gradient in each of the focal stack image. Then we apply a Gaussian blur to the gradients so than the high gradient is dispersed in to the surrounding region. This step is to solve the problem when we have objects with little details and does not produce so much gradient.

After the blurred gradients are created, we label each pixel with a number corresponding to the image in focal stack that has the highest gradient. The number can be easily remapped into depth if we know the parameters on camera when it took the images. In our case, since we generate focal stack from light field data, we will use the same parameter used when generating the focal stack.

The next step is to sharpen the label map or the depth map. We run a 3x3 kernel through the image. For each 3x3 pixels we encounter we score each label based on the true gradient at that pixel. If the label with the highest gradient has a score with a proportion out of total score in the kernel higher than some threshold, we change all label in the current kernel to that label. We do nothing otherwise.

The idea behind this algorithm is that we have confidence that pixels with low gradient that are close to pixel with high gradient will be at the same depth. If pixels in the kernel have roughly equal gradient score, it could mean either the scores are low and we are not sure at which depth the pixels belong to or the kernel encounters an edge where there are at least two depth with high gradient and we have no winner.

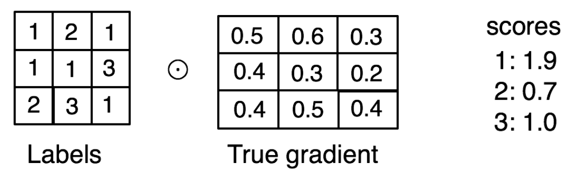

Figure 2: An example of how score for each label in a kernel is calculated

IV. Results

4.1 Depth estimation

The result of depth map or label map is shown in Figure 3. In the Gaussian blur for gradient, we use a standard deviation of 15. When we find labels with highest blurred gradient, the label map is rounded. After applying the high gradient score filter with a threshold of 70%, the depth map become sharper and capture the edge of the table, the first and the last book. The algorithm even captures different depth of the first book and the regions on the table close to the camera. However, it cannot distinguish between the second and the last book.

Figure 3: The original image, all in focus (top); label map generated from blurred gradient (middle); label map after high gradient score filter (bottom).

4.2 Light field reconstruction

Although we can generate the light field or perspective shifts in the entire (u, v) domain, for simplicity of display, we generate only light field in $(u, 0)$ and $(0, v)$ domains.

You can find a perspective shift video in the data and code folder under Video/depthmap.avi. One of the frame is shown in Figure 4. We can see that the method in this project produces discontinuity artifact at the edge of the first book.

We also calculate peak signal-to-noise ratio (PSNR) and structural similarity analysis index (SSIM) of images from the two algorithms and compare them against the ground truth. The results in Table 1 show that our method has higher PSNR and SSIM values.

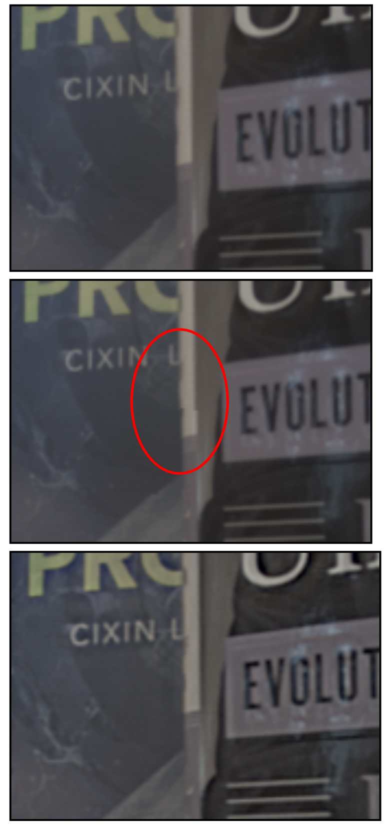

Figure 4: The original light field data (top); light field reconstructed using the method in this project (middle) the shift artifact is shown in the red circle; light field generated from the method mentioned in [1] (bottom).

	Our method	Levin et al. [1]
PSNR	9.2876	8.0648
SSIM	4.8898e-04	9.6455e-04

Table 1: The comparison of PSNR and SSIM of the results at (u, v) = (7, 0) from our method and [1] against the ground truth.

V. Discussion

The results show that reconstructing light field from focal stack images as suggested in [2] is possible through the use of depth estimation. However, there are still limitations to this method.

First, although the method can produce a high quality all-in-focus image, it does not resolve occluded objects. Blurred edges of images in focal stack containing light information of objects behind can be resolved using deconvolution method but not from depth estimation. Second, when focal stack does not cover enough resolution in depth, that is the sampling rate in depth is not enough, the difference of gradient cannot distinguish some objects at different depths, resulting in the objects being mistakenly labeled as the same depth. Lastly, the method requires a high accuracy of depth estimation. As we can see that incorrect depth estimation can easily cause artifacts when the image is shifted.

Our method for depth estimation although cannot capture depth difference between the middle book and the last book, it can estimate different depths of the table and the book close to the camera. When this depth map is input in the perspective shift in the video, natural movement can be observed.

One challenge of studying the reconstruction of light field is choosing quantitative metrics to compare the results. In this study, PSNR and SSIM is used for comparison. Our method has shown to perform better than the method in [1]. One reason is that we the all-in-focus image which is very similar to the original light field. On the other hand, deconvolution method in [1] could introduce some artifacts in the regularization of Wiener filter that increase the mean square error. However, qualitatively it performs better at distinguishing between two objects at different depth.

VI. Reference

[1] Levin, Anat & Durand, Fredo. (2010). Linear View Synthesis Using a Dimensionality Gap Light Field Prior.
[2] Mousnier, A & Vural, E & Guillemot, Christine. (2015). Partial light field tomographic reconstruction from a fixed-camera focal stack.
[3] Y. Boykov, O. Veksler, and R. Zabih. (2001). Efficient approximate energy minimization via graph cuts.