Multi-Perspective Stereoscopy from Light Fields
Final Project Report for CSE 528

Sushant Ojal
Stony Brook University
s.ojal@cs.stonybrook.edu

Abstract

The main objective of this report is to describe the details of implementation of the paper "Multi-Perspective Stereoscopy from Light Fields" by Kim et. al. This is a self contained document giving an analysis of the different sections of the paper and their respective implementations.

1. Introduction

This paper focuses on generating stereo pair of images from a given light field. The proposed methodology can be used to generate stereo content with a per pixel control over disparity based on multi perspective imaging from light fields. This technique is mainly helpful in 3D content modification and editing post production. The stereo images are produced as a result of piecewise continous cuts through the light field, while minimizing certain energy parameters, which ensure there are no sudden disparity changes.
Section 2 of this report deals with light field representations and the dataset which has been used to create a light field for the current implementation. Section 3 describes the generation of images once the light field has been contructed. The next section deals with extracting cuts form this light field which correspond to a stereo pair. The generated cuts are in accordance to a pre-determined goal disparity range. Section 5 describes the formulation of this task as an energy minimization problem and the algorithms used to implement this. I will conclude by highlighting the results and stating out the challenges which I faced during the course of this project.

2. Representation of light field

Light slab representations are the essentially the norm of representing a 4D light field. For our discussion, we are interested in generating stereo pair of images from light field, or generating 3D information given a light field. For both of these purposes, we are interested only in the horizontal parallax. Therefore, we can get rid of the redundant fourth

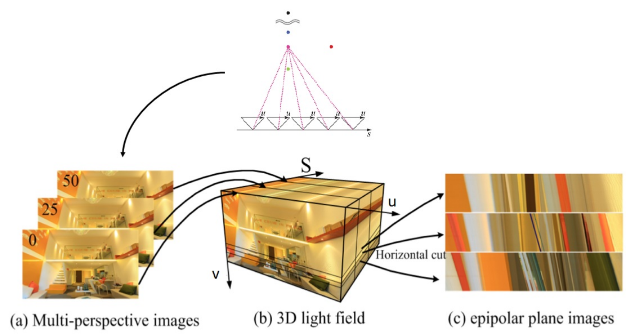

Figure 1. From left to right: (a) multi perspective images taken from a linearly translating camera (b) stacked up volume of these images or the 3D light field (c)epiploar plane images generated by cuts in the 3D light field

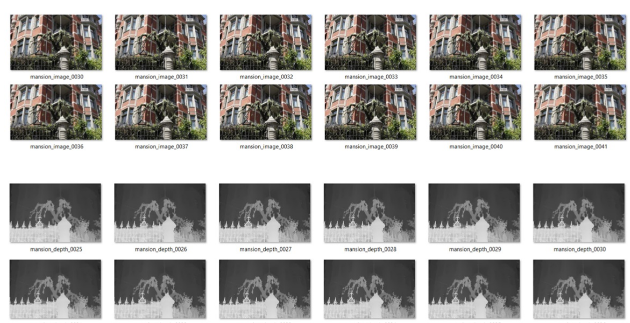

Figure 2. A snapshot of the mansion database which was used for implementing this paper

dimension and work with a three dimensional light field. A 3D light field can be represented by a 3D volume of perspective images taken from a linearly translating camera.A vertical cut through this volume gives us an image through a new view point or perspective. A horizontal cut represents an epipolar plane image.

For the implementation of this paper, I am using the "mansion" dataset which was created by Disney Research. (https://www.disneyresearch.com/project/lightfields).

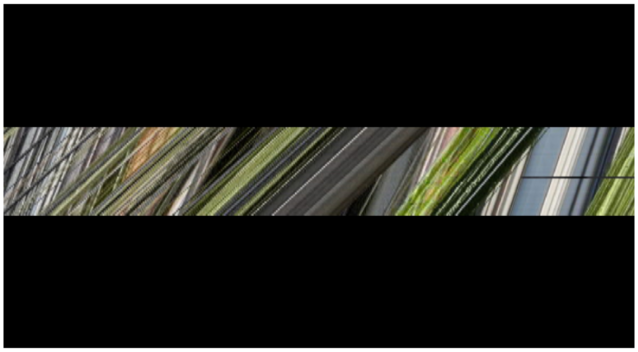

Figure 3. An epipolar plane generated from the mansion light field along the v-axis at v = 300

These images were acquired by a camera of focal length 50mm with baseline differences of 10mm. Each image is of the resolution of 5490x3450. Since working with a light field containing such high resolution images is quite memory intensive, as a first step, each of these images was downscaled to 600x400. The light field was constructed by stacking up 50 of such images along the 's' axis. (or the 'z' axis).

Cutting this volume at different values of 'v' gives us different epipolar planes.

3. Image generation from light field

Any cut through the generated EPI volume gives us a new image. The depth along the s-axis at which the cut was made determines the disparity of the individual pixels in the resultant image. To gain a better control over this disparity, it would be logical to generate more finely grained images. In other words, we would like to synthesis more images from the given light field by *interpolating* the values of the images in the current epipolar volume.

Since the depth of the individual pixels is known, we can estimate the slope of every pixel in the epipolar volume. Generating similar pixels along this slope will give us new images along the s-axis of the volume. The more granularity of this line, the more will be the number of generated images.

Figures 4 and 5 highlight the difference observed in the epipolar planes generated from light fields generated from 50 and 500 images respectively. As expected, there is a lesser granularity in the second image.An obvious downside of this is that there are ten times more pixels which have to stored and computed on in case of further operations.

I will now move on to explain how stereoscopic images can be generated from this light field which obey certain constraints.

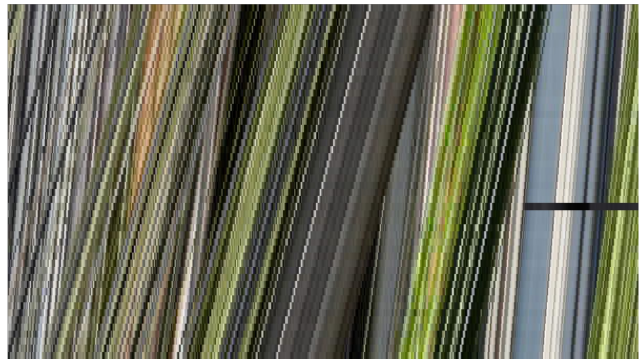

Figure 4. A magnified view of an epipolar plane of a light field of 50 images

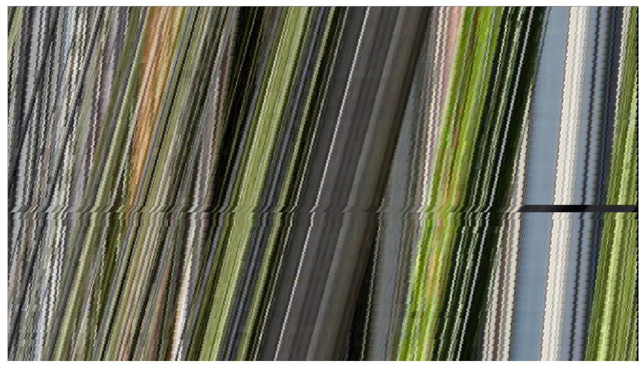

Figure 5. A magnified view of an epipolar plane of a light field of 500 images generated by interpolating from a light field with 50 images

4. Stereoscopic Light Field Cuts

In order to create a stereoscopic image pair from a given 3D light field, a corresponding 3D disparity volume is first created. The value of every pixel in this volume is a scaled reciprocal of the depth of that pixel. This volume can therefore be interpreted as a normalised disparity volume, such that the image disparity of a pixel p in an image $I_s^{'}$ with respect to a reference image I_s is defined as :

$$\mathcal{T}_s(\mathbf{p}, s') = \Delta(s', s)\mathcal{D}(\mathbf{p}, s')$$

The difference $\Delta(s^{'}, s) = s^{'} - s$ is proportional to the camera baseline between two images $I_s^{'}$ and I_s. The volume obtained as a result of this product is a true disparity volume, which contains the actual disparities of all the images in the light field with respect to a reference image. Figures 6 and 7 show the epipolar planes for both of these volumes.

Figure 6. Normalised disparity

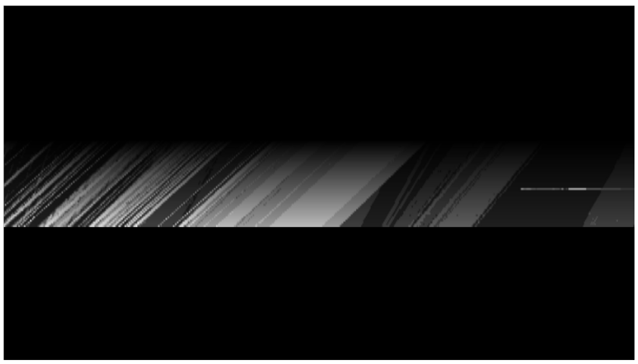

Figure 7. True disparity with respect to the first image of the light field.

4.1. Goal based multiperspective cuts

The main objective of the paper can be now be defined more precisely, ie, to calculate a non-planar cut through this true disparity volume which satisfies a given set off disparity constraints. To this end, I created a goal disparity map in which the value of every pixel specifies the disparity that is desired in the output stereo pair. Since, the true disparity volume represents the actual disparity of a point with respect to a reference image, the volume which contains the difference of the true disparity volume and the goal disparity image, represents the deviation of a pixels disparity from the desired goal disparity. Figure 8 shows this *difference* volume, if the goal disparity is chosen as the disparity image halfway through the true disparity volume.

5. Formulation as an Energy Minimization Problem

The energy measuring the deviation of a 2D cut s_C can be expressed as:

$$E_{\mathrm{d}}(s_C) = \sum_{\mathbf{p}} |\mathcal{T}_s(\mathbf{p}, s_C(\mathbf{p})) - \mathcal{G}(\mathbf{p})|$$

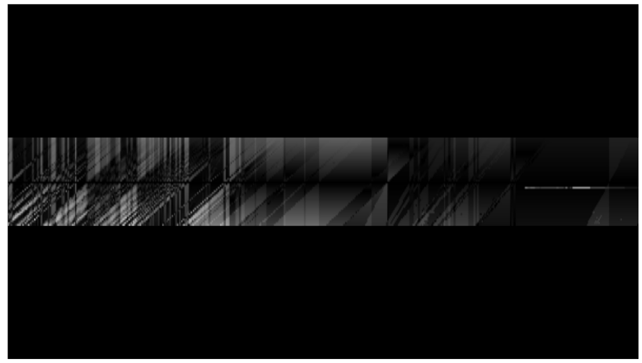

Figure 8. The difference volume, in which every point represents the deviation in disparity from the goal disparity.

If the cut is computer exactly in accordance with values which minimize this formulation, we might end up getting pixels which are adjacent and not of continuous depth. This might become very prominent if these pixels comprise the visually salient parts of the image. Therefore certain constraints need to be applied on the cut so that it is continuous and sudden changes in depth values of adjacent pixels are penalized. This can be done by using the following formulation:

The energy measuring the deviation can be measured as:

$$\begin{aligned} E_{\mathrm{s}}(s_C) = & \sum_{(\mathbf{p},\mathbf{q})\in N_u} |s_C(\mathbf{p}) - s_C(\mathbf{q})|\, p_u(*) + \\ & \sum_{(\mathbf{p},\mathbf{q})\in N_v} |s_C(\mathbf{p}) - s_C(\mathbf{q})|\, p_v(*), \text{ with} \\ p_u(*) = & \min\left(p_{\max}, |\partial_s \mathcal{D}(*)| + \lambda \mathcal{D}(*) + \kappa\, |\partial_s \mathcal{L}(*)|\right) \\ p_v(*) = & \min\left(p_{\max}, |\partial_s \mathcal{D}(*)| + \lambda \mathcal{D}(*) + \kappa\, |\partial_u \mathcal{L}(*)|\right) \end{aligned}$$

This energy expression penalizes the cuts if there is a change in the slope of the cut for the adjoining pixels along the s axis. The maximum penalty $p_{\max}$ ensures that the cut can be discontinious at some places. The overall energy function is therefore:

$$E(s_C) = E_{\mathrm{d}}(s_C) + kE_{\mathrm{s}}(s_C)$$

The minimization of this function can be achieved by using the graph cut optimization.[Boykov and Kolmogorov 2004]. (http://www.csd.uwo.ca/faculty/yuri/Abstracts/pami04-abs.shtml). I am using the maxflow-mincut algorithm implemented by the authors to compute the mincut of the graph. To formulate the light field as a graph, I took the following steps:

- For n input images of dimension w $*$ h, construct a regular graph of size $w * h * (n + 1)$.

- A ray at position (u, v, s) is associated with a directional graph edge between the corresponding two nodes along the s-axis, and the edge weight is chosen as difference of true disparity and goal disparity.

- Bi-directional edges between neighboring nodes along the u-axis and v-axis are weighted with the corresponding smoothness values.

The graph cut algorithm was successfully able to compute the mincut, if there was one edge (outdegree) per node of the graph. On adding the remaining four edges per node, which is necessary to enforce the smoothness of disparities between adjoining pixels, there was a shortage of memory. The results are therefore based on only on minimizing the energy deviation from the goal disparity.

6. Results

I have used the programmable pipeline of OpenGL to display the stereo pair of images. In the fragment shader, the red color of every pixel of the reference image is combined with the green and blue colors (cyan) of the computed frame to give the output stereopair.

I used two goal disparity maps, to evaluate the linear and the non-linear remapping. Both the results have the first image of the light field as the reference image. Figures 9 and 10 show these both outputs.

Figure 9. Linear remapping. Note the large disparities in pixels of the foreground as well as background.

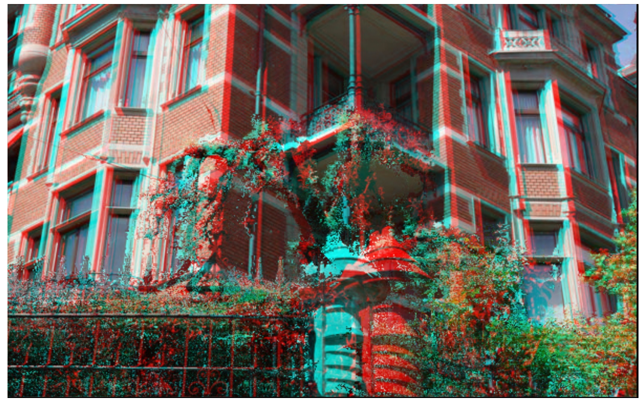

Figure 10. Non linear remapping. Note the larger disparity in the foreground pixels as opposed to the smaller disparity in the pixels of the background.

A very obvious limitation in the second result is the presence of ”holes” or the absence in coherence in pixels along the edges of the foreground pixels. This can be attributed to the absence of edge weights penalizing the absence of smoothness, while computing the min cut.

7. Conclusion

I have tried to realise the main objective of this paper by computing the piecewise cuts which allow per pixel disparity control. With easier access to plenoptic cameras, optimised versions of such solutions can find multiple applications. Additionally, I feel there is a scope for further research in terms of calculating min-cuts of the graph formulation of a light field with a huge number of images.

8. Libraries and resources used

- OpenGL
- OpenCV (for reading images and doing basic pixel operations
- ”mansion dataset” by Disney Research. (https://www.disneyresearch.com/project/lightfields/)
- implementation of the maxflow algorithm for energy minimization (http://www.csd.uwo.ca/faculty/yuri/Abstracts/pami04-abs.shtml)

9. References

- An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. (Yuri Boykov and Vladimir Kolmogorov)
- Multi-Perspective Stereoscopy from Light Fields (Changil Kim, Alexander Hornung, Simon Heinzle, Wojciech Matusik, Markus Gross)