Advanced Depth Estimation and Refocusing with Light Fields

Tzu-Sheng Kuo
Stanford University
tskuo@stanford.edu

Xianzhe Zhang
Stanford University
xianzhez@stanford.edu

Abstract

Light field photography opens abundant opportunities that are not feasible with traditional cameras. In this project, we explore two applications of light field photography. First, we implement Adelson and Wang's method for depth estimation with light fields. Secondly, we implement Ng's Fourier Slice Photography Theorem for refocusing. We evaluate our results using the 4D Light Field Benchmark and the Stanford Light Field Archive datasets. We report our findings both qualitatively and quantitatively.

1. Introduction

The concept of light field photography has existed since the early 20th century. In 1903, Ives proposed using parallax barriers to capture images from different directions [6]. Similarly, Lippmann invented integral imaging to capture light fields using lenslets in 1908 [12]. However, due to the limited technology available at that time, light field photography did not have many real-world applications.

Recently, thanks to the advancement in technology, light field cameras have finally come into the market. For example, companies such as Lytro[1] and Raytrix[2] were founded in 2006 and 2008, respectively, aiming to develop commercial light field cameras. These movements enable researchers to experiment with the theorems and algorithms of light fields using real-world data. For example, we implemented depth estimation and refocusing algorithms in homework 5, and evaluate the result using a real-world image captured by a Lytro camera. This assignment motivates us to explore more advanced algorithms for depth estimation and refocusing with light fields.

In this project, we reproduce two algorithms for depth estimation and refocusing, respectively. For depth estimation, we implement Adelson and Wang's [1] method and evaluate it with the 4D Light Field Benchmark [5], which includes synthetic images with ground truth disparity. For refocusing, we implement Ng's Fourier Slice Photography Theorem [16] and evaluate it using The New Stanford Light Field Archive[3], which includes real-world light field images. Figure 1 shows a high-level illustration of the goal of our project. We also describe our analysis both qualitatively and quantitatively in this report.

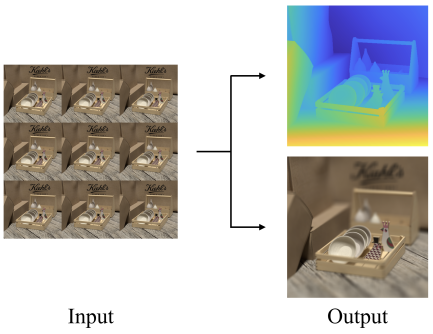

Figure 1. An illustration of our script, which takes a light field as input and outputs both a disparity map and refocused images.

2. Related Work

Although our work mainly focuses on reproducing two algorithms, we share related work of these areas and refer the interested readers to these references.

2.1. Depth Estimation

The goal of depth estimation is to take a set of light field images as input and generates a disparity or a depth map. The goal of a depth estimation algorithm is to take a set of light field images as input and generates a disparity or a depth map. Adelson and Wang are the pioneers in this area, where they estimate the disparity based on spatial and viewpoint derivatives of a light field [1]. However, their approach can only estimate disparity at the edges.

More recently, researchers have been working on algorithms that make dense depth estimation beyond edges. For

[1] https://en.wikipedia.org/wiki/Lytro
[2] https://en.wikipedia.org/wiki/Raytrix
[3] http://lightfield.stanford.edu/

example, Tao et al. [18] leverage both the defocus and correspondence cues for estimation by analyzing the spatial and angular variance in the 4D epipolar image. Kim et al. [10] reconstruct depth maps from light fields with a two-stage pipeline, where they first computes more reliable estimates on boundaries and then process the interior regions with a fine-to-coarse procedure. Jeon et al. further develop sub-pixel-wise depth estimation algorithm that refines initial estimation throughout iterations [9].

Following the above approaches, we are interested in extending Adelson and Wang's method for dense estimation beyond edges. We explain our method later in Section 3.

2.2. Refocusing

Refocusing is a critical applications of light field photography. Most refocusing methods follow one of the two pathways: integral photography methods in spatial domain and Fourier Slice Theorem methods in Fourier domain.

Most of the integral photography methods build upon [7] and [13], such as Levoy and Hanrahan 's work on light field rendering [11] and Ng et al.'s work on light field photography with a hand-held plenoptic camera [17]. According to Ng's work [16], the time complexity of light field photography using integral method in spatial domain is $O(n^4)$, which limits its applications to some extent.

Ng et al. propose Fourier Slice Photography Theorem [16] based on Fourier Slice Theorem [2] [3], which is widely used in many medical imaging techniques [14]. Based on these existing foundations, Ng et al. apply Fourier Slice Theorem on image refocusing with light field cameras and work on the fidelity with full-aperture photographs that have finite depth of field. They reduce the time complexity to $O(n^2logn)$ by pre-computing the Fourier transformation of the light field in $O(n^4logn)$.

3. Method

3.1. Depth Estimation

For depth estimation, we first follow Adelson and Wang's method [1] to calculate the disparity of each pixel based on this equation:

$$disparity = \frac{\sum_P (I_x I_{v_x} + I_y I_{v_y})}{\sum_P (I_x^2 + I_y^2)},$$

where I is the image intensity, I_x and I_y are spatial derivatives, and I_{v_x} and I_{v_y} are viewpoint derivatives. Here, P denotes an integration patch that blurs over sharp variations. According to the original paper, the authors suggest a patch size from 5×5 to 9×9.

Adelson and Wang's method also calculate the confidence of each pixel based on the following equation:

$$confidence = \sum_P (I_x^2 + I_y^2).$$

The value of confidence means how confident the algorithm is for the estimated disparity at each pixel. For example, high confidence means that the estimated disparity at a pixel is more likely to be true. Since the confidence is basically the sum of the square of spatial derivatives, we may expect that edges and boundaries are more likely to have high confidences.

To propagate the disparity from pixels with high confidence to those with low confidence, we first assign the value of pixels to NaN if it is below a threshold. Then, we use the inpaint_nans Matlab File Exchange toolbox developed by D'Errico [4] to interpolate these pixels based on the disparity of remaining pixels with high confidence. More specifically, we use its method two to estimate these values by solving a direct linear system of equations.

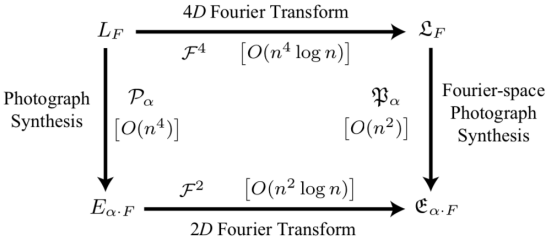

Figure 2. Fourier Slice Photography Theorem [16].

3.2. Refocusing

Generally, we are implementing the algorithm following Fourier Slice Photography Theorem straightforward. Ng et al. has presented detailed mathematical derivations in their work [16]. Here we just generalize the whole pipeline and implementations.

The figure 2 shows the Fourier Slice Photography Theorem. As we can see on the left part of the figure, integral method for photograph synthesis in spatial domain will take $O(n^4)$ time. We could also follow the path by transforming the light field imaging (L_F) from 4D spatial domain to 4D Fourier domain (i.e. $\mathcal{F}^4$, time complexity: $O(n^4logn)$), reducing the dimension from 4D to 2D by applying the Fourier Photography Operator (i.e. $\mathfrak{P}_\alpha$, time complexity: $O(n^2)$), and then transforming the reduced 2D Fourier representation from Fourier domain to spatial domain (i.e. $\mathcal{F}^{-2}$, time complexity: $O(n^2logn)$), which would produce the image ($E_{F'}$) we want. Here n is the number of samples in each dimension, F is the separation between the lens and the film, F' is the separation between the lens and the refocus plane, and $\alpha = F'/F$.

Therefore, the Fourier Slice Theorem could also be represented by this equation [16]:

$$\mathcal{P}_\alpha \equiv \mathcal{F}^{-2} \circ \mathfrak{P}_\alpha \circ \mathcal{F}^4$$

.

Specifically, we could divide the algorithm into 2 phases: preprocess and refocusing. We compute the 4D Fourier transformation ($F^4[L_F]$) in the preprocess phase, which would take $O(n^4 logn)$ time using Fast Fourier Transformation algorithm.

In the refocusing phase, we first apply the Fourier Photography Operator $\mathfrak{P}_\alpha$ on the 4D Fourier transformation of the light field. Here,

$$\mathfrak{P}_\alpha \equiv \frac{1}{F^2} \mathcal{S}_2^4 \circ \mathcal{B}_\alpha^{-T}$$

$\mathcal{S}_2^4$ is the slicing operator that reduces a 4 dimensional function down to 2 dimensional one by zero-ing out the last 2 dimensions. $\mathcal{B}_\alpha$ is a basis changing operator which could be represented by $\mathcal{B}[f](x) = f(\mathcal{B}^{-1}x)$. And $\mathcal{B}_\alpha$ could also be explicitly represented by a matrix related to α:

$$\mathcal{B}_\alpha = \begin{bmatrix} \alpha & 0 & 1-\alpha & 0 \\ 0 & \alpha & 0 & 1-\alpha \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Applying $\mathcal{B}_\alpha^{-T}$ will take $O(n^2)$ time.

Finally, we just need to apply inverse Fast Fourier Transformation algorithm on the result we get in the last step.

Ng et al.'s work [16] also mentioned some techniques to reduce artifacts such as roll off error, aliasing artifacts. The source of artifacts is discussed in [16] and other many works [8] [15]. In a nut shell, it mainly comes from the imperfect band-limited filter of the camera which will is different from an ideal band-limited filter.

In our implementation, besides applying different α values, we explore these techniques to reduce artifacts, such as boarder padding. In Ng et al's work, they only discussed the conditions for full-aperture light field. We also conduct some experiments and try to explore the relationship between the view points in the light field and the aperture size, and the impacts of view points on the generated images.

4. Evaluation

In this section, we describe the datasets and parameters that we use to evaluate the algorithms. We will show the evaluation results in the next section.

4.1. Dataset

4.1.1 4D Light Field Benchmark

This dataset contains synthetic light fields with 9×9 viewpoints. The size of each image is 512×512 with 3 channels. Moreover, each set of light field also has its depth and disparity map as ground truth for evaluation. Their website [4] also offers tools that import and export light field easily. We use this dataset for evaluating the depth estimation algorithm that we implemented. Note that this dataset is sometimes called the HCI synthetic dataset in other literature.

4.1.2 The New Stanford Light Field Archive

This dataset [5] contains real-world light fields with 16×16 viewpoints. The size of each image is 1024×1024 with 3 channels. We use this dataset for evaluating the refocusing algorithm that we implemented.

4.2. Parameters

4.2.1 Depth Estimation

After several preliminary studies, we chose 0.3 as the confidence threshold. For the size of the integration patch, we use the suggested value 5×5 according to the paper.

4.2.2 Refocusing

The dataset provides images with resolution of 1024×1024 which is too big for our experiments. We down-sampled these images in real time with a factor of 0.25 to save memory and accelerator our experiments.

5. Results

5.1. Depth Estimation

Figure 3 shows example qualitative results of the depth estimation algorithms. The first row is the center view of the input light fields. The second row is the raw disparity map generated with Adelson and Wang's method. The white and black pixels in the third row are those with confidence larger and less than the threshold, respectively. The fourth row is the final disparity map after interpolation. The last row is the ground truth disparity. The colors yellow and blue denote closer and farther depth, respectively. Note that the depth and disparity are interchangeable.

Confidence is critical to the quality of the final disparity map. As we can see in figure 3, images with clear and bounded confidence edges generally have better results, such as town, platonic, and greek. In contrast, images with complex confidence edge have worse results, such as tower, dishes, and antinous.

The object in the middle and farther depth may be confused sometimes. For example, the center regions in dino and dishes are incorrectly estimated as the farthest object in the images. This result may originate from the estimation

[4] https://lightfield-analysis.uni-konstanz.de/tools

[5] http://lightfield.stanford.edu/

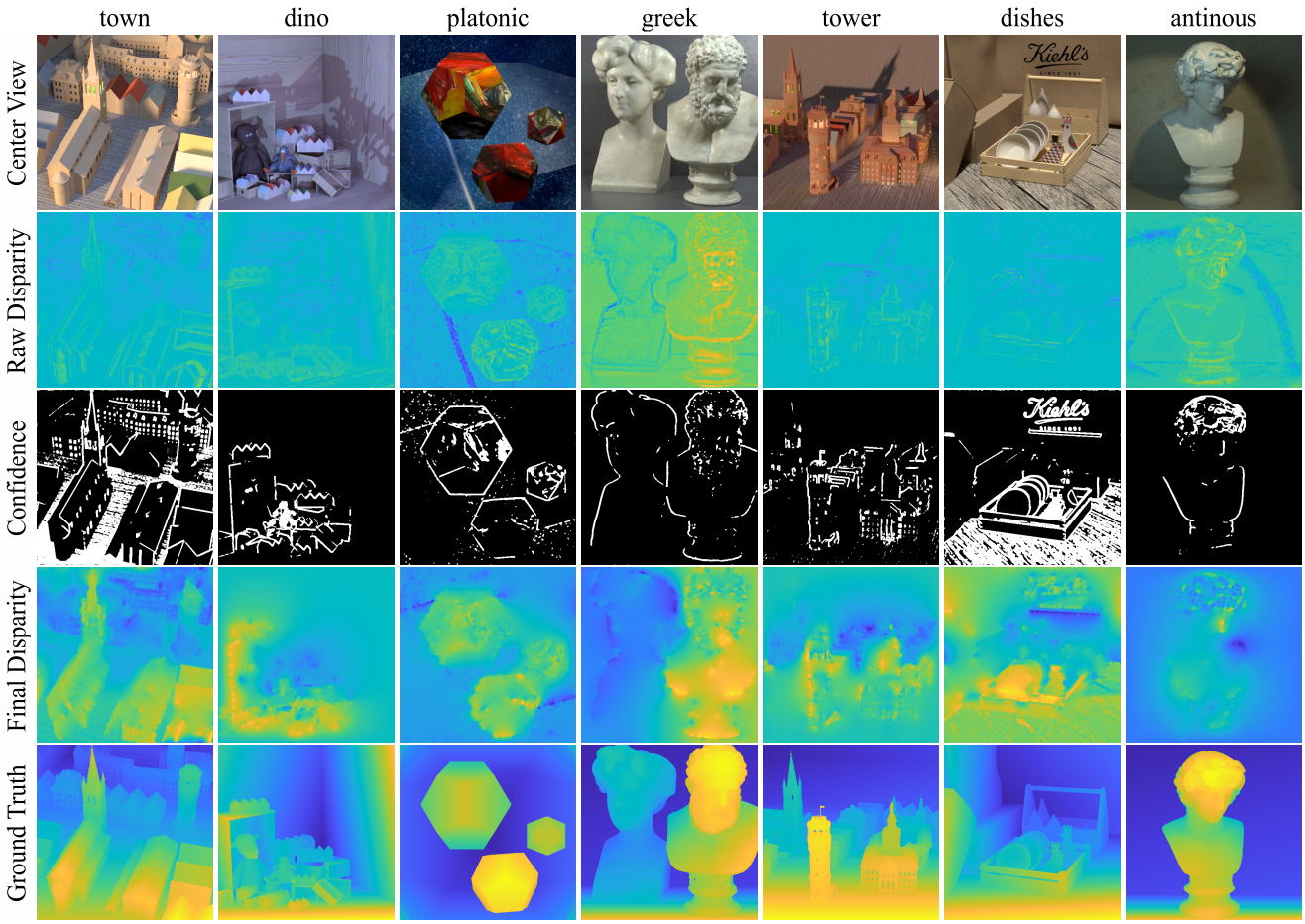

Figure 3. Examples of qualitative results.

scene	town	dino	platonic	greek	tower	dishes	antinous
MSE	1412	2421	3776	4084	5477	5506	5939
PSNR	16.63	14.29	12.36	12.02	10.75	10.72	10.39

Table 1. Evaluation of the Heidelberg dataset with MSE and PSNR.

of raw disparity, where most of the pixels are assigned the green color (middle depth) because of the lack of disparity cue in these regions.

Table 1 shows the quantitative results of the images in Figure 3. For comparison, we scale both the ranges of the final disparity and ground truth to [0,255], then calculate the mean squared error (MSE) and peak signal-to-noise ratio (PSNR). Images on the left have smaller MSE and larger PSNR, meaning that their disparity map may have better quality, which is aligned with our observation of the qualitative results.

5.2. Refocusing

In this section, we will show the results of our experiments and discuss the findings we have. We first test the correctness and the ability of this algorithm to refocus on the planes in different depths, we use different resolutions of light field view points to explore the impacts on the generated image, and we also compare the results of different padding sizes.

5.2.1 Refocus plane

As we mentioned in section 3.2, $\alpha = F'/F$, it is the key to control the depth we want to refocus. We use different α on several light fields to refocus on different planes. Figure 4 shows the center view, images refocusing on different depths, and corresponding visualizations of their Fourier transformation. We can clear see the refocusing effects and the results of applying the Fourier Photography Operator. The changes in Fourier domain give us the final expected

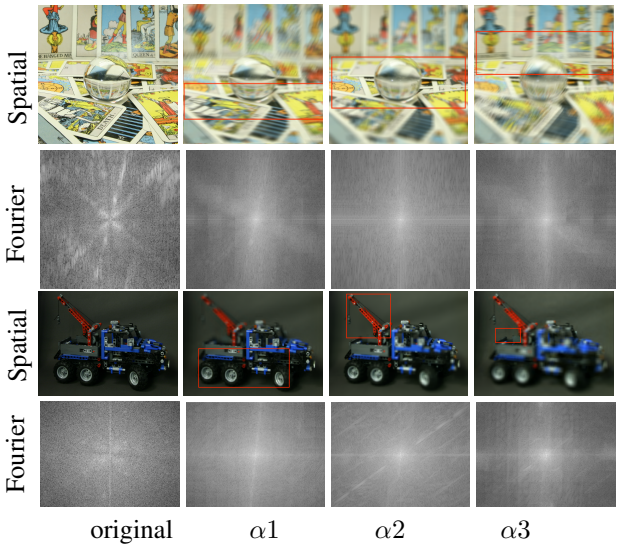

Figure 4. Refocusing on different planes with different α. For every two rows, the first image is the center view of the light field and its Fourier transformation visualization, following by images refocusing on different planes. For the tarot image, $\alpha 1 = 0.95$, $\alpha 1 = 1.0$, $\alpha 1 = 1.05$. For the LEGO truck image, $\alpha 1 = 0.96$, $\alpha 1 = 1.0$, $\alpha 1 = 1.04$.

refocus images.

5.2.2 View points resolution

We also notice that the view points resolution will shed great impacts on the generated images. As shown in figure 5, we can clearly observe that the top row with view points resolution 16×16 looks blurrier than the bottom row with resolution 8×8. On one hand, a larger light field size mimics a larger aperture size which could increase the size of the circle of confusion and give us blurrier effects. On the other hand, a larger light field size could also bring about a refocused image with higher fidelity, since more view points could provide more information to reconstruct the refocused image. We can see that the images in the bottom row has more artifacts than the images in the top row, even the images in the bottom row seem clearer than the top ones.

5.2.3 Eliminating Artifacts

Ng et al.'s work[16] suggests that adding paddings to the boarders of the light field in the dimensions of view points could correct rolloff error. We experiment with different padding size for the light field. In figure 6, we are comparing two padding size. For example, $1\times$ means we increase the light field size from 16x16 to 32x32 by adding more zero paddings. As shown in figure 6, larger padding size did increase the quality of the refocused image especially when

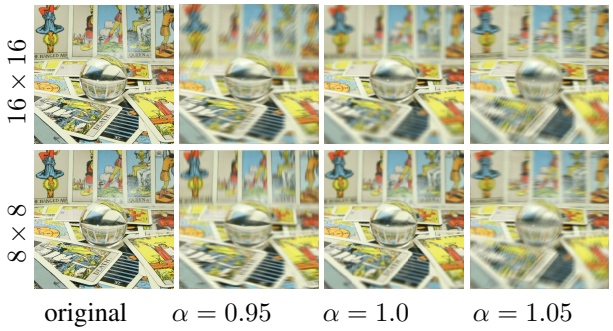

Figure 5. Refocus results with different light field size.

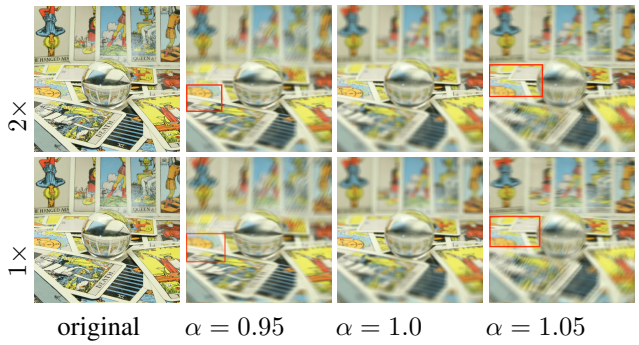

Figure 6. Refocus results with different padding size.

$\alpha \neq 1.0$. We can see that more the focused area is clearer with $2\times$ padding than $1\times$. Intuitively, We believe this is because the Fourier Photography Operator maps the original information to a bigger space. since we are just giving it an zero value when the mapped index is out of the boundary, we could capture more mapped information with more paddings and contribute to the generated images finally.

6. Discussion

6.1. Depth Estimation

The final disparity is a significant enhancement from the raw estimation of Adelson and Wang's method; however, there are rooms for improvements to be made. Since our method follows a two-step pipeline, we could improve either the initial estimation or the inpainting algorithms, or both.

For the initial estimation algorithm, which is exactly Adelson and Wang's method in this project, more advanced algorithms could be considered. For example, Kim et al. [10] perform the initial depth estimation using pixels in the epipolar-plane image (EPI) for better results, even their confidence is merely the difference in color variation.

For the inpainting algorithm, more advanced methods

might also improve the performance of the Matlab File Exchange toolbox that run out of the box. For example, [9] propagate the depth with higher confidence with using multi-label optimization, which minimizes an objective function after iteration.

More recently, deep learning based approaches have further pushed the boundaries of what depth estimation from light fields can achieve. For example, Wu et al. [19] develop a convolutional neural network that operate in the EPI domain for depth estimation. These methods are often end-to-end rather than the two-step pipeline in this project.

6.2. Refocusing

Fourier Slice Photography Theorem provides us a more efficient way to generate refocused image with light field. It reduces the time complexity from $O(n^4)$ to $O(n^2 logn)$ with an $O(n^4 logn)$ preprocessing phase. But we also found that the refocusing effect might decrease greatly with bigger deviation from 1.0 for α as show in figure 4. This might be result of the Fourier Photography Operator. Even though larger padding size would give us better refocused image, it will also greatly increase the computation time. Since the time complexity is highly related to the size of every dimension.

There are many interesting experiments we could do if we got more time. Such as we could explore more techniques to suppress artifacts such as over-sampling, some advanced filter, etc. And there are more different ways to change the aperture size, we could try to extend the algorithm to enable it to generate refocusing image with appropriate convolution in the Fourier domain as suggested in the work [16].

7. Conclusion

We implement two algorithms for depth estimation and refocusing in this project, which serves as an extension of homework 5. We also provide a detailed qualitative and quantitative analysis of these algorithms. Finally, we discuss our observations of the results and refer the interested readers to various references we found useful.

Acknowledgement

We would like to thank professor Gordon Wetzstein for his invaluable guidance and feedback on this project. We would also thank TA David Lindell for his helpful support and comments.

References

[1] E. H. Adelson and J. Y. A. Wang. Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):99–106, Feb 1992.

[2] R. N. Bracewell. Strip integration in radio astronomy. *Australian Journal of Physics*, 9(2):198–217, 1956.

[3] S. R. Deans. *The Radon transform and some of its applications*. Courier Corporation, 2007.

[4] J. D'Errico. inpaint_nans, 2020. MATLAB Central File Exchange.

[5] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldluecke. A dataset and evaluation methodology for depth estimation on 4d light fields. In *Asian Conference on Computer Vision*. Springer, 2016.

[6] F. E. Ives. Parallax stereogram and process of making same, 1903. U.S. Patent 725567.

[7] H. E. Ives. Parallax panoramagrams made with a large diameter lens. *JOSA*, 20(6):332–342, 1930.

[8] J. I. Jackson, C. H. Meyer, D. G. Nishimura, and A. Macovski. Selection of a convolution function for fourier inversion using gridding (computerised tomography application). *IEEE transactions on medical imaging*, 10(3):473–478, 1991.

[9] H. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. Tai, and I. S. Kweon. Accurate depth map estimation from a lenslet light field camera. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1547–1555, June 2015.

[10] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross. Scene reconstruction from high spatio-angular resolution light fields. *ACM Trans. Graph.*, 32(4), July 2013.

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

[12] G. Lippmann. Épreuves réversibles donnant la sensation du relief. *J. Phys. Theor. Appl.*, 7(1):821–825, 1908.

[13] G. LIPPMANN. La photographie inte grale. *Comptes-Rendus*, 146:446–551, 1908.

[14] A. MACOVSKI. Medical imaging systems. In *Prentice Hall.*

[15] D. P. Mitchell and A. N. Netravali. Reconstruction filters in computer-graphics. *ACM Siggraph Computer Graphics*, 22(4):221–228, 1988.

[16] R. Ng. Fourier slice photography. *ACM Trans. Graph.*, 24(3):735–744, July 2005.

[17] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, P. Hanrahan, et al. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2(11):1–11, 2005.

[18] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi. Depth from combining defocus and correspondence using light-field cameras. In *Proceedings of the 2013 IEEE International Conference on Computer Vision*, ICCV '13, page 673–680, USA, 2013. IEEE Computer Society.

[19] G. Wu, M. Zhao, L. Wang, Q. Dai, T. Chai, and Y. Liu. Light field reconstruction using deep convolutional network on epi. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1638–1646, July 2017.