

HIGH ANGULAR RESOLUTION LIGHT FIELD RECONSTRUCTION WITH CODED-APERTURE MASK

Wanxin Qu¹, Guoqing Zhou¹, Hao Zhu¹, Zhaolin Xiao², Qing Wang¹, Rene Vidal³

¹School of Computer Science
Northwestern Polytechnical
University, Xi'an 710072, China

²School of Computer Science
Xi'an University of Technology
Xi'an 710048, China

³Center for Imaging Science
Johns Hopkins University
Baltimore MD 21218, USA

ABSTRACT

In the past decade, light field imaging has greatly extended the imaging capabilities of traditional photography. However, the applications of light field imaging are limited by the aliasing artifacts due to the plenoptic sampling trade-off between angular and spatial domains. We propose to use a coded aperture light field camera instead of the traditional one, which can get more angular information without losing spatial resolution. To that end, we exploit a theoretical model to explain the relationship between light field and the raw data captured by the sensor. Then, we design a mask to code the rays using compressive sensing. Last, the sparse characteristic of light field in gradient domain and the corresponding optimization methods are utilized to reconstruct the high angular resolution light field. Experimental results on synthetic data and real data demonstrate that our system can obtain high angular resolution light field by producing a low-aliasing refocused image and high PSNR multi-view images.

Index Terms— Light Field, Coded-Aperture, Angular Resolution, Sparse representation

1. INTRODUCTION

Light field is the parametric representation of the four-dimensional optical radiation field that contains both spatial and angular information. In geometric optics, the light rays can be uniquely determined by spatial and angular information. Levoy and Hanrahan [1] adopted two-parallel-planes model to parameterize the light field, *i.e.*, $L(s, t, u, v)$, where (s, t) and (u, v) describe the spatial and angular information respectively. In the past decade, light field imaging greatly extends the imaging abilities of traditional photography [2].

However, the applications of light field imaging are limited by the low imaging quality and aliasing artifacts, which can be traced back to plenoptic sampling trade-off between angular and spatial resolutions [3]. In order to improve the

spatial resolution, it is necessary to reduce the angular sampling. The low angular resolution is often deemed insufficient to produce high quality refocused images. As a result, the refocused images will exhibit strong aliasing artifacts due to angular undersampling [4]. The trade-off has become the bottleneck that limits the use of light field cameras and computational photography [5].

In this paper, we propose a light field imaging model with coded aperture, which is consisted of a micro lens array before the sensor and a mask before the aperture. It can be used to reconstruct higher angular resolution light field than classical plenoptic camera with single shot and don't sacrifice the spatial resolution. We improve angular resolution while ensuring the time efficiency.

2. RELATED WORK

In order to improve the angular resolution, many computational methods have been proposed without changing the light field acquisition devices [6, 7, 8]. Wanner *et al.* [6] improved the resolution in a variational Bayesian framework which depends on the accurate priori information of scene geometry, and it is difficult to be implemented in real applications. Levin *et al.* [7] explored the dimension gap between the 4D light field and 3D focal stack, and reconstructed the 4D light field by using a 3D measurement sets, it reduces the complexity of light field acquisition. Zhan *et al.* [8] proposed a phase-based light field synthesis architecture to get dense views. However, the accuracy of new views relies on the depth estimation and vice versa.

Apart from these computational methods, an alternative way is placing additional hardware into optical path. Representative works include placing the mask board close to the sensor [9, 10] and aperture [11, 12, 13] and mounting external encoding device on the main lens [14, 15]. Marwah *et al.* [9] proposed a compressive light field camera architecture that allows higher-resolution light field to be recovered in a single-shot. It needs much time to learn the dictionary. Xu *et al.* [10] captured the high resolution light field by using two attenuation masks and frequency-multiplexing. Inevitably, two masks decrease the light efficiency and signal noise ratio. Tambe *et al.* [11] designed multiplexing ma-

The work in the paper was supported by NSFC grants (61401359, 61531014, 61501370), Grant of Natural Science Basic Research Plan in Shaanxi Province of China (2015JQ6209), NPU Foundation for Fundamental Research (3102015JSJ0007), Aeronautical Science Foundation of China (No.2015ZC53044) and the Seed Foundation of Innovation and Creation for Graduate Students in NPU (Z2016154).

trices to resolve the resolution trade-off and used dictionary learning and sparse representations for robust reconstruction. Wang *et al.* [16] proposed a light field camera which includes a coded aperture and a random convolution CMOS sensor. The model has high reconstruction quality and light efficiency. Liang *et al.* [13] presented an adjusted aperture to capture the full sensor resolution light field. The multiple exposures will take much time for high angular resolution. In summary, these methods either need to increase the number of shots or change the mask patterns to increase more angles. Our model is based on the plenoptic camera to achieve the angular ascension. As the plenoptic camera has the inherence advantages of collecting multiple angle information, our model can reconstruct higher number of angles and only needs a single shot.

3. LIGHT-FIELD IMAGING MODEL WITH CODED APERTURE

3.1. Coded Aperture Light Field Camera

We adopt the plenoptic camera model [17] to capture the light rays and put a mask on the aperture to code them, as shown in Fig.1. According to the geometric optics, all rays emitted from one 3D space point are converged by the main lens and imaging on micro-lens array plane. Then, they are separated by the micro-lens and imaging on the sensor, and these light rays constitute a sub-image. One sub-image records different angular rays from the same point.

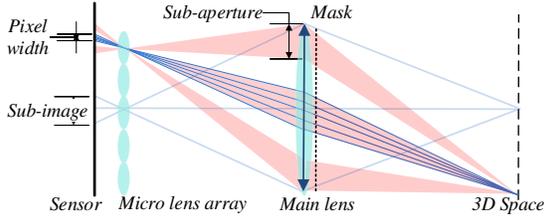


Fig. 1. Imaging model of coded aperture light field.

Generally, the 2D image captured by sensor is the raw data and it is determined by the arrangement of the micro-lens array. We could use camera parameters to decode raw data $\mathbf{I}(x, y)$ to the 4D raw light field $\mathbf{L}^R(s, t, u, v)$ [18]. \mathbf{L}^R is an acquired low resolution light field. The rays from the same angle are imaging at the same offset from the center of the sub-image. The general conversion is written as Eqn. (1).

$$\mathbf{L}^R(s, t, u, v) = \mathbf{I}(\mathbf{c}_s(s, t) + u, \mathbf{c}_t(s, t) + v) \quad (1)$$

where $\mathbf{c}_s(s, t)$ and $\mathbf{c}_t(s, t)$ are the center coordinates of each sub-image. In order to facilitate the description, we straighten the position (s, t) and angular (u, v) into 1D vector respectively. Thus, the light field \mathbf{L}^R can be expressed by the form of ray matrix as follows,

$$\mathbf{L}^R = \begin{bmatrix} l_{11}^r & l_{12}^r & \cdots & l_{1m}^r \\ l_{21}^r & l_{22}^r & \cdots & l_{2m}^r \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1}^r & l_{p2}^r & \cdots & l_{pm}^r \end{bmatrix}_{p \times m} \quad (2)$$

where l_{ij}^r denotes a light ray in the collected light field \mathbf{L}^R , i and j are the spatial and angular indexes of light ray respectively. Each column refers to the ray sets from one sub-aperture. The number of column is equivalent to the angular resolution m . Each row denotes m angular ray sets that pass through the same micro-lens. The number of row is equivalent to the spatial resolution p . Consequently, the resolution of the sensor is pm .

3.2. Coded Aperture Light Field Imaging Model

For each sub-aperture, there are many different angular light rays, but these light rays cannot be resolved by the sensor. We further subdivide it into n blocks by the coded mask. The mask before the main lens is used to control the portion of incident light that passes through the blocks, as shown in Fig.1. When the mask is full on state, the imaging of all angular rays from one point is an integral process in Eqn.(3). Eqn.(4) is a simplification of the Eqn.(3).

$$\begin{bmatrix} l_{i1}^r & l_{i2}^r & \cdots & l_{im}^r \end{bmatrix} = \begin{bmatrix} \begin{pmatrix} l_{i11} \\ l_{i12} \\ \vdots \\ l_{i1n} \end{pmatrix}^T & \begin{pmatrix} l_{i21} \\ l_{i22} \\ \vdots \\ l_{i2n} \end{pmatrix}^T & \cdots & \begin{pmatrix} l_{im1} \\ l_{im2} \\ \vdots \\ l_{imn} \end{pmatrix}^T \end{bmatrix} \begin{bmatrix} \mathbf{1} & & & \\ & \mathbf{1} & & \\ & & \ddots & \\ & & & \mathbf{1} \end{bmatrix} \quad (3)$$

$$\mathbf{L}_i^R = \mathbf{L}_i \text{diag}(\mathbf{1}, \mathbf{1}, \cdots, \mathbf{1}) \quad (4)$$

where $i = 1, 2, \dots, p$, and i denotes the i -th sub-aperture. $\mathbf{1}$ is a $n \times 1$ vector in which each element is $\frac{1}{n}$. The $mn \times 1$ vector \mathbf{L}_i^R is a row of the matrix \mathbf{L}^R , which represents the vector of i -th micro lens image and consists of m angular rays from the i -th position in the collected light field. The $mn \times 1$ vector \mathbf{L}_i contains all angular rays from this position in the light field. Each sub-vector in \mathbf{L}_i represents a ray set of the j -th sub-aperture and the element l_{ijk} denotes the k -th block rays in this sub-aperture. n is the resolution of the coded mask which determines the magnification of the angular resolution, and it is limited by the hardware parameters, such as the resolution of mask and the size of aperture, so it is not infinite.

The function of mask is equivalent to assigning a weight a_{jk} to each block in a sub-aperture, a_{jk} represents the transmittance of j -th sub-aperture and k -th block. The imaging process of each position is same and independent. Therefore, the general form of high resolution light field model containing all position can be extended as Eqn.(5) and the weights are arranged in Eqn.(6).

$$\mathbf{L}^R = \mathbf{L} \mathbf{A} \quad (5)$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_m \end{bmatrix} \quad \text{with } \mathbf{A}_j = \begin{bmatrix} a_{j1} \\ a_{j2} \\ \vdots \\ a_{jn} \end{bmatrix} \quad (6)$$

where \mathbf{A} is a $mn \times m$ matrix, and \mathbf{A}_j describes a weight set of one sub-aperture. This model is suitable for the plenoptic camera based on the micro-lens array rather than the traditional camera [15]. It is effective for enhancement of high angular resolution by one shot and it will be proved in the following.

4. HIGH ANGULAR RESOLUTION LIGHT FIELD RECONSTRUCTION

4.1. Design of Coded Aperture

In our model without mask, the angular resolution of the captured light field equals to the number of the sub-apertures m . In order to acquire higher angular resolution light field, we encode the light rays passing through each of the sub-apertures. The proposed mask \mathbf{M} has m blocks and each block is subdivided into n units. Specific forms are as follows.

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2 & \dots & \mathbf{M}_{m_0} \\ \mathbf{M}_{m_0+1} & \mathbf{M}_{m_0+2} & \dots & \mathbf{M}_{2m_0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{M}_{m-m_0+1} & \mathbf{M}_{m-m_0+2} & \dots & \mathbf{M}_m \end{bmatrix} \quad (7)$$

with

$$\mathbf{M}_j = \begin{bmatrix} a_{j1} & a_{j2} & \dots & a_{jn_0} \\ a_{j(n_0+1)} & a_{j(n_0+2)} & \dots & a_{j2n_0} \\ \vdots & \vdots & \ddots & \vdots \\ a_{j(n-n_0+1)} & a_{j2} & \dots & a_{jn} \end{bmatrix} \quad (8)$$

where $j = 1, \dots, m$, $m_0 = \sqrt{mn}$, $n_0 = \sqrt{n}$, \mathbf{M}_j is the mask of j -th sub-aperture. a_{jk} is drawn from a Gaussian distribution with 0.5 means, $1/\sqrt{m}$ variance and subjects to $0 < a_{jk} < 1$. There are some possible ways implement the coded mask in real camera, such as LCoS [19]. The measurement matrix \mathbf{A} is contributed with elements of mask \mathbf{M} and the arrangement of optical imaging model, where $\mathbf{A}_j = \text{vec}(\mathbf{M}_j^T)$. \mathbf{A} is extremely sparse and only has non-zero elements on diagonal block of matrix (Eqn.(6)). Candes and Tao [20] has proved that the measurement matrix with Gaussian random distribution can recover a highly accurate estimate signal with high probability.

4.2. High Angular Resolution Reconstruction

As discussed above, \mathbf{L}_i^R corresponds to i -th sub-image which discretely records the under-sampling of light field. \mathbf{L}_i represents all rays from i -th point. According to the optical structure, we can rewrite the \mathbf{L}_i^R and \mathbf{L}_i in the form of 2D matrix,

$$\mathbf{L}_i^R = \begin{bmatrix} l_{i1}^r & \dots & l_{im_0}^r \\ \vdots & \ddots & \vdots \\ l_{i(m-m_0+1)}^r & \dots & l_{im}^r \end{bmatrix} \mathbf{L}_i = \begin{bmatrix} l_{i11} & \dots & l_{i1n} \\ \vdots & \ddots & \vdots \\ l_{im1} & \dots & l_{imn} \end{bmatrix} \quad (9)$$

According to geometric optics, we find that the light rays of different sub-images do not interfere with each other. Therefore, we can recover high angular resolution sampling from each sub-image one by one and then combine them into

a whole light field. Let $\mathbf{y} = \text{vec}((\mathbf{L}_i^R)^T)$ and $\mathbf{X} = \text{vec}(\mathbf{L}_i^T)$, the cost function is

$$\mathbf{y} = \mathbf{A}^T \mathbf{X} \quad (10)$$

Obviously, Eqn.(10) is an under-determined system. It is generally accepted that the natural image is mostly sparse in the gradient domain. The gradient of \mathbf{X} is,

$$\nabla \mathbf{X} = \mathbf{D}\mathbf{X} = [\mathbf{D}^h; \mathbf{D}^v]\mathbf{X} \quad (11)$$

where \mathbf{D}^h and \mathbf{D}^v are gradient operators along horizontal and vertical directions respectively. \mathbf{D}^h and \mathbf{D}^v are sparse diagonal matrix. When \mathbf{X} is a matrix or vector, the operator has different forms.

With the Lambertian assumption, the light rays which pass through the main lens share similar intensities when they are emitted from one point. Thus, each sub-image is more sparse than natural image in the gradient domain. At the same time, it can be approximated as isotropic. Herein, we further define a transfer $\mathbf{H} = \mathbf{D}^h + \mathbf{D}^v$, which is a full rank matrix. It is noticeable that $\mathbf{H}\mathbf{X}$ is sparse and we assume \mathbf{X} is k -sparse. Let $\mathbf{S} = \mathbf{H}\mathbf{X}$, according to the CS theory, we format the cost function.

$$\begin{aligned} & \min \|\mathbf{S}\|_1 \\ & \text{s.t. } \mathbf{y} = \mathbf{A}^T \mathbf{H}^{-1} \mathbf{S} \end{aligned} \quad (12)$$

where \mathbf{A}^T is a random matrix and \mathbf{H}^{-1} is a deterministic matrix with certain structure. There is incoherence between them. The coefficient \mathbf{S} is very sparse ($k \ll mn$) because of the Lambertian assumption and principle of our imaging model. The number of observations m should be higher than $\mathcal{O}(k \log(mn/k))$, where m is sufficient in our model when we do not divide it into too many blocks in each sub-aperture. We use the log barrier algorithm [21] to solve the cost function in our work.

5. EXPERIMENT RESULT

We have verified the proposed method on both synthetic data and real scene data. The synthetic scene is created by the 3D animation software ‘‘blender’’¹ and two real scene datas are Jelly Bean, Tarot Cards and Crystall Ball². Our evaluation experiment consists of two aspects: the peak signal-to-noise ratio (PSNR) of reconstructed images and the aliasing ratio of refocus images. PSNR is used to measure the quality of light field images after compressed sensing. The aliasing ratio is the quantitative evaluation of angle resolution. In experiments, the noise is drawn from a Gaussian distribution with 0 means and 0.001 variance. Sampling ratio R_S describe the relationship of angular resolution between the collected light field and the reconstructed light field, $R_S = \sqrt{1/n}$.

The result of super angular resolution reconstruction are shown in Fig.2. In these experiments, the angular resolution

¹‘‘<https://www.blender.org/>’’

²‘‘<http://lightfield.stanford.edu/lfs.html>’’

of collected light field is 8×8 and reconstructed light field is 16×16 . (a), (b), (c) is one of the reconstructed multi-view images on three data sets respectively. The first line in (d),(e),(f) are 8 local images of collected light field and the second line are 16 reconstructed results. The reconstruction results have higher angular resolution, smoothly angular changes and the same angle range as collected light field. This facilitates the aliasing removal.

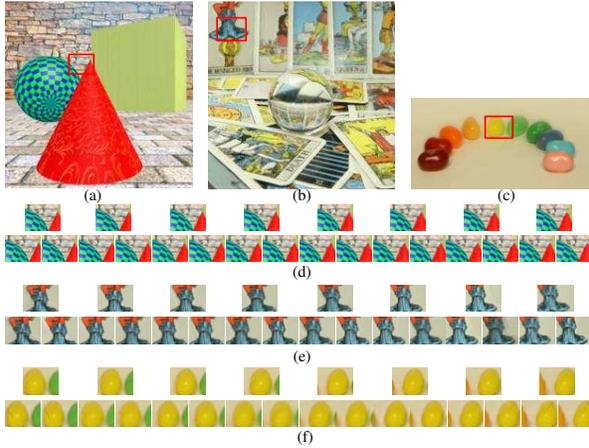


Fig. 2. Reconstructed multiview images. (a) Synthetic data, (b) Tarot Cards and Crystal Ball, (c) Jelly Bean. (d), (e), (f) are local images of collected and reconstructed light field.

Table 1. PSNR results on synthetic and real scene data

Sampling ratio	50%	33%	25%	20%
Synthetic data	34.10	32.76	32.41	29.43
Cards and Ball	Ours	33.98	31.47	30.31
	Wang	25.5	21.5	20

On the aspect of evaluating PSNR, we complete the experiments on synthetic data (Table 1) and the comparative experiments with Wang[16] on two real scene data sets (Table 1 and 2). The spatial resolution of synthetic data and real scene data in table 1 is 400×400 and 60×120 respectively. The PSNRs of reconstructed images by our methods are generally higher than work [16]. In table 2, The original light field we used is down-sampled from Jelly Bean, the resolution of which is $120 \times 240 \times 6 \times 6$. m and mn are the number of collected and recovered angle respectively. The result of our model has high PSNR and it decreases with the decrease of sampling ratio. The result verifies that our model has good reconstruction ability in the gradient continuous region.

Table 2. PSNR result of compared experiment on Jelly Bean

Sampling Ratio	m, mn	Our method	Wang
40.8%	6, 36	36.55	31.36
50%	9, 36	38.44	33.39
58%	12, 36	40.87	34.72

In the experiments of aliasing reduction evaluation, we use the aliasing detection method [4] to test aliasing ratio R_A

at three angular resolution. $R_A = n_a/p$, where n_a is the number of pixels whose aliasing intensity is higher than the threshold. Threshold is 40 in this experiment. Fig.3 shows the aliasing detection results and Table 3 is the quantitative representation of Fig.3. The 8×8 and 16×16 angular resolution light field are all reconstructed by 4×4 angular resolution light field. The focal plane of these three data sets are in rear, front and middle. Every result images denote the lightness of the aliasing at each position. The aliasing ratio decreased obviously with the increase of angular resolution. The reconstruction model can guarantee to recover enough and accurate angular information to eliminate aliasing. Under 50% angular sampling ratio, reconstruct $240 \times 120 \times 16 \times 16 \times 3$ light field needs 1.5 hours, of which we use eight threads of parallel processing.

Table 3. Aliasing ratio with different angular resolutions

Resolution	4×4	8×8	16×16
Synthetic data	21.4%	12%	9.5%
Jelly Beans	24.6%	4.2%	1.3%
Cards and Ball	62.5%	32.9%	18.1%

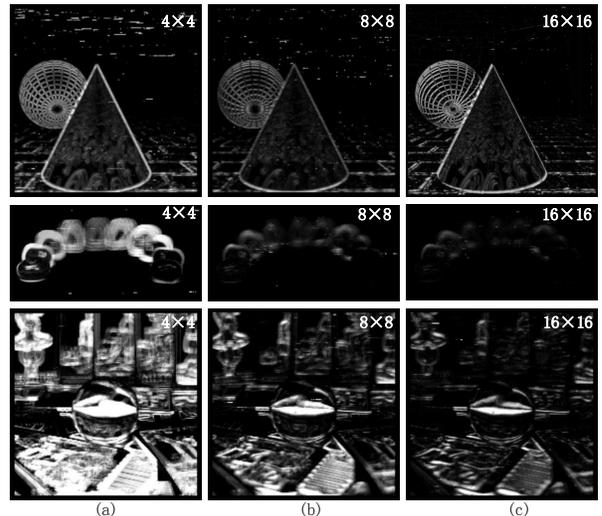


Fig. 3. Aliasing detection of refocus images in the same focal plane. The images from left to right are the detected results from 4×4 , 8×8 and 16×16 angular resolution light fields.

6. CONCLUSION

In this paper, we propose a novel plenoptic camera model and the corresponding optimization method to obtain high angle resolution light field. We have validated the model on both synthetic and real light field. The reconstructed light field images have lower aliasing ratio and higher PSNR. Experimental results prove the effectiveness of our model and method. More importantly, the light field was obtained with single shot. Therefore, the proposed method can be extended to applications that require real-time shooting, such as high angular resolution light field video.

7. REFERENCES

- [1] M. Levoy and P. Hanrahan, "Light field rendering," in *ACM SIGGRAPH*, 1996, pp. 31–42.
- [2] S. K. Nayar, "Computational cameras: Approaches, benefits and limits," *Columbia University Computer Science Technical Reports*, 2011.
- [3] A. Levin, W. T. Freeman, and F. Durand, "Understanding camera trade-offsthrough a bayesian analysis of light field projections," in *IEEE ICCV*, 2008, pp. 88–101.
- [4] Z. Xiao, Q. Wang, G. Zhou, et al., "Aliasing detection and reduction in plenoptic imaging," in *IEEE CVPR*, 2014, pp. 3326–3333.
- [5] G. Wetzstein, I. Ihrke, and W. Heidrich, "On plenoptic multiplexing and reconstruction," *International Journal of Computer Vision*, vol. 101, no. 2, pp. 384–400, 2013.
- [6] S. Wanner and B. Goldluecke, *Spatial and Angular Variational Super-Resolution of 4D Light Fields*, Springer Berlin Heidelberg, 2012.
- [7] A. Levin and F. Durand, "Linear view synthesis using a dimensionality gap light field prior," *IEEE CVPR*, vol. 26, pp. 1831–1838, 2010.
- [8] Z. Zhang, Y. Liu, and Q. Dai, "Light field from micro-baseline image pair," *IEEE CVPR*, pp. 3800–3809, 2015.
- [9] K. Marwah, G. Wetzstein, Y. Bando, et al., "Compressive light field photography using overcomplete dictionaries and optimized projections," in *ACM Transactions on Graphics*, 2013, vol. 32, pp. 101–102.
- [10] Z. Xu, J. Ke, and E. Y. Lam, "High-resolution lightfield photography using two masks," *Optics Express*, vol. 20, no. 10, pp. 10971–83, 2012.
- [11] S. Tambe, A. Veeraraghavan, and A. Agrawal, "Towards motion aware light field video for dynamic scenes," in *IEEE ICCV*, 2013, pp. 1009–1016.
- [12] A. Ashok and M. A. Neifeld, "Compressive light-field imaging," *International Society for Optics and Photonics*, vol. 7690, no. 1, pp. 281–290, 2010.
- [13] C. K. Liang, T. H. Lin, B. Y. Wong, et al., "Programmable aperture photography: multiplexed light field acquisition," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 15–19, 2008.
- [14] T. Georgiev, K. C. Zheng, B. Curless, et al., "Spatio-angular resolution tradeoffs in integral photography," in *Eurographics Symposium on Rendering Techniques, Nicosia, Cyprus*, 2006, pp. 263–272.
- [15] S. D. Babacan, R. Ansorge, M. Luessi, et al., "Compressive light field sensing," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4746–57, 2012.
- [16] Y. Wang, L. Wang, D. Kong, and B. Yin, "High-resolution light field capture with coded aperture," *IEEE Transactions on Image Processing*, vol. 24, pp. 1, 2015.
- [17] R. Ng, M. Levoy, M. Brdif, et al., "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, pp. 1–11, 2005.
- [18] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," *IEEE CVPR*, pp. 1027–1034, 2013.
- [19] H. Nagahara, C. Zhou, T. Watanabe, et al., "Programmable aperture camera using lcos," *ECCV*, pp. 337–350, 2010.
- [20] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [21] E. Candes and J. Romberg, "11-magic: Recovery of sparse signals via convex programming," pp. 1–19, 2005.