# Super Resolution for Light Field Images

Anna Broome
Department of Electrical Engineering
Stanford University
abroome@stanford.edu

## Abstract

*Light field cameras are an interesting technology that has proven useful for applications including post-capture refocusing and single exposure 3D scene capture. However, their popularity has been limited due in part to the trade-off between angular and spatial resolution that is present in light field images. To address this trade-off, this paper explores how super resolution methods may be applied to light field images. By exploiting the aliasing of the scene present in the light field sub-aperture images, these images can be combined into a single view higher resolution image using traditional super resolution techniques. In this paper, gradient descent is used to reconstruct high resolution images based on a light field image formation model. Results are presented for reconstruction of simulated light field data.*

## 1. Introduction

Recently, there has been growing interest in the use of light field cameras for both consumer and scientific applications. These cameras are based on the principal of integral photography, first developed by Lippmann in 1908 [1]. To understand the difference between integral photography and conventional photography, we must first describe how light travels in all directions through all points in space. The behavior of light in this manner is described by the plenoptic function, a 5D function that describes the radiance of all light rays, as parametrized by a position $x, y, z$, and two angles $\phi$ and $\theta$ [2]. Because radiance remains constant along a given direction, assuming no blocking objects or materials, there is redundancy in one dimension of the plenoptic function [2] and we need only concern ourselves with the 4D representation of the light field. A conventional camera integrates the radiance information of a scene over an aperture, which results in the integration (and thus loss) of angular information. On the other hand, integral imaging is based off the idea that light rays incident at different angles can be multiplexed onto different sensor locations, preserving angular information from the scene.

Light field cameras capture this multiplexed information, typically using a microlens array, and have emerged as promising devices, enabling digital refocusing, 3D imaging from a single exposure, and enhanced depth of field retrievals [2]. However, light field cameras suffer significantly degraded performance compared to conventional cameras in terms of spatial resolution. This reduced spatial resolution is due to the inherent spatio-angular resolution tradeoff present in all captured light fields. The total resolution of a light field image is dictated by the camera sensor size, as well as the size of the microlens array. Light field imaging has the potential to be transformative in the fields of imaging and photography, but as of yet the inherent limitations in resolution have reduced its adoption and popularity.

To address these limitations, there has been interest in using super resolution techniques to improve the angular and spatial resolution of light field images in a variety of applications. In order to apply super resolution techniques, we must have multiple low resolution images that are aliased to carry different information of the same scene, not simply shifted and interpolated versions of the same image [3]. This is exactly what we get when using light field cameras with microlens arrays in front of the camera sensor. Thus, it is appropriate to apply the same super resolution techniques to light field images that are applied to series of images from conventional cameras.

The main contributions of this paper are to provide an overview of current research regarding super resolution of light field images, to detail the image formation model for such images, and to provide preliminary results of an implementation of super resolution of light field images based on the gradient descent method. Section 2 discusses further background of light field cameras and related work specific to super resolution of their images. Section 3 details the image formation model and methodology used in the super resolution process. Section 4 introduces preliminary results using gradient descent based super resolution applied to light field images. Section 5 concludes and discusses opportunities for future work and follow-on research in this area.

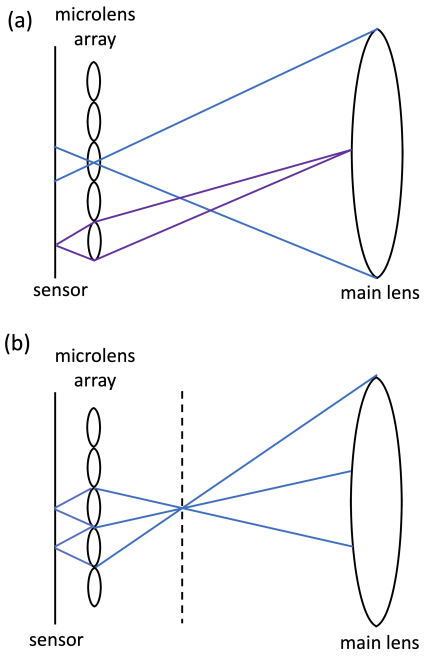



Figure 1. (a) Plenoptic 1.0 design. (b) Plenoptic 2.0 design. Adapted from [5].

## 2. Related Work

There are two main types of light field cameras, classified as the plenoptic 1.0 camera and the focused or plenoptic 2.0 camera. Although both classes of camera utilize a microlens array in front of the sensor, their slight differences have important consequences for the application of super resolution. Below, we detail the basic set ups of each camera and the subsequent impact on super resolution.

The plenoptic 1.0 camera embodies the light field camera first pioneered by Ng and the Stanford Graphics Lab [4] that retailed commercially by the company Lytro. It consists of a microlens array between the main lens and the image sensor that serves to multiplex the spatial and angular information of the scene. The microlens array is positioned at the focal plane of the main lens, as in Fig. 1(a). Behind each microlens lie a fixed number of pixels on the sensor. As a result of this geometry, the image formed behind each microlens has lots of angular information about a single spatial point in the scene. The camera's angular resolution is defined by the number of pixels per microlens and its spatial resolution is defined by the number of microlenses in the array, leading to the angular-spatio resolution tradeoff. To render an image from a plenoptic 1.0 camera is equivalent to integrating the pixels underneath each microlens.

The plenoptic 2.0, or focused plenoptic camera, was introduced several years later by Lumsdaine and Georgiev [6] and has subsequently been commercialized by Raytrix. Like the plenoptic 1.0 camera, the 2.0 version consists of a microlens array between the main lens and image sensor. However, in this case the microlens array is positioned such that it focuses onto the image plane of the main lens, as in Fig. 1(b). Now the information from one point in the scene is spread across multiple microlenses and image rendering can be achieved by integrating points within the microlenses that image the same position. In other words, multiple pixels from each microlens are used in the final image and the spatial resolution has been decoupled from the number of microlenses.

The difference in microlens location and focal length between the plenoptic 1.0 and 2.0 cameras means that while super resolution can be applied to each, it is most useful in different cases. For the plenoptic 1.0 camera, light fields capturing scenes with nearby objects are most suitable as candidates for super resolution due to the depth dependence of the scene. Conversely, for plenoptic 2.0 cameras super resolution is most beneficial for objects in the scene located at optical infinity because these objects are guaranteed to be imaged onto the sensor with a non-integer pixel shift (the aliasing required for super resolution techniques).

Several approaches to super resolution of light field images have been explored in the literature, which primarily involve formulating the high resolution image reconstruction as an optimization problem. In [3], super resolution of light field images is demonstrated using a variational Bayesian framework using Lambertian reflectance priors. In order to apply the Bayesian optimization, Bishop et al. first reconstruct the depth map of the scene, which allows them to find the initially unknown misalignment between the different low resolution sub-aperture images of the light field [3]. One of the unique contributions of this work is that the prior used in the optimization, a Markov random field prior, allows for the preservation of local texture in addition to edges within the recovered image [3].

Georgiev and Lumsdaine tackled the challenge of super resolution for light field images, focusing specifically on plenoptic 2.0 cameras and super resolving scenes with objects located at optical infinity [5]. By restricting application of their technique to objects focused at optical infinity in plenoptic 2.0 light fields, the authors are able to achieve super resolution without doing sub-pixel image registration of the sub-aperture images. On one hand, this enables very precise and reliable super resolution, however it is limited in its application due to the requirement of a plenoptic 2.0 camera. In the case where a plenoptic 2.0 camera is not available, it may be possible to modify the camera hardware such that the microlenses are focused on the image plane of the main lens. Figure 2 demonstrates the improvement achieved through super resolution in [5].

In follow on work, Georgiev, Chunev, and Lumsdaine publish a super resolution process that uses the raw light



Figure 2. (Left) Zoom in of bike wheel using traditional light field rendering. (Right) Zoom in of bike wheel after super resolution. From [5].



Figure 3. (Left) Super resolved from demosaiced image. (Right) Super resolved from raw (non-demosaiced) image. From [7].

field data, rather than the already demosaiced data [7]. They argue that this approach produces better results one, because the super resolution step does not suffer from interpolation artifacts due to Bayer demosaicing and two, because after super resolution the color planes are already registered, facilitating combination into a final full color image [7]. Similarly to their work described above, this work does not make use of depth dependent scene information, nor does it utilize a deconvolution kernel. Figure 3 shows the results of their super resolution technique, which has a factor of two improvement in resolution over other super resolution methods [7].

Another interesting and relevant work with super resolution and light field images is that of the *PiCam* from Venkataraman and others [8]. The *PiCam* is a thin monolithic camera array consisting of a 4×4 camera array, where each camera samples a single narrowband spectral channel [8]. Since each camera samples the same scene, aliasing is present in the low resolution images meaning super resolution methods can be applied to improve the final image. For their camera hardware, they determine the maximum possible super resolution factor to be 2.4 [8]. Unlike the plenoptic 2.0 approaches in [5] and [7], the super resolution technique employed in [8] does not require objects to be located at optical infinity and takes into account blur due to the optics and sensor pixels. To begin, the authors perform a flat field correction and find the depth map (i.e. parallax detection) of the scene using a sum of absolute differences cost metric [8]. This depth map is used as an input the super resolution step, which is solved using a maximum-a-posteriori formulation with the *L*2-norm used as the likelihood term and the prior being a locally-adaptive *L*2-norm using a bilateral filter [8]. Results from this process are promising and highlight the potential of a light field camera of a size that could plausibly be integrated into a handheld device such as a cellphone or tablet.

## 3. Methods

The task of super resolution revolves around recovering a high resolution image from multiple lower resolution images with some translational misalignment between them. In the application of super resolution to light field images, each of the low resolution images is one of the sub-aperture images captured in the light field and the translational misalignment arises from a combination of physical offsets between the sub-aperture images as well as the depth dependent parallax of different objects in the scene. As mentioned previously, super resolution requires non-integer, or sub-pixel shifts between the low resolution images so that information from the scene is aliased in each image, rather than simply shifted and interpolated. Below we more formally describe the image formation model and approach to super resolution taken in this paper.

### 3.1. Image Formation Model

The image formation model can be described by (1) where $\mathbf{x}$ is the high resolution image we seek to find and $\mathbf{y}_k$ is the k-th sub-aperture image captured by the light field camera [8]. Because the high resolution image can be synthesized from any number of viewpoints available from the light field data, we must select a reference viewpoint at which we will find $\mathbf{x}$. Here, we select the central viewpoint to be our reference.

$$\mathbf{y}_k = \mathbf{DHW}_k\mathbf{x} + \mathbf{n}_k \tag{1}$$

To go from the estimated high resolution image to the sub-aperture images, we first apply a shift matrix $\mathbf{W}_k$ that warps the high resolution image to the k-th sub-aperture image. This warping operation is defined by the disparity map that relates the central view to each other microlens view. Then, we apply a blur matrix, $\mathbf{H}$, which accounts for the point spread functions of the optics and sensor. Then a downsampling or decimation matrix, $\mathbf{D}$ is applied that describes how the sensor samples the high resolution image. We assume that each sub-aperture image is also affected by additive i.i.d. Gaussian noise, denoted as $\mathbf{n}_k$ in (1).

### 3.2. Gradient Descent Formulation

To formulate the image formation model and task of super resolution as an optimization problem, we rewrite (1) in

the form of the residual $F(x)$,

$$F(x) = \sum(\mathbf{Ax} - \mathbf{b})^2 = \sum_k(\mathbf{DHW}_k\mathbf{x} - \mathbf{y}_k)^2 \quad (2)$$

Our goal is then to minimize this residual, which we do utilizing the gradient descent method. Each gradient descent step is calculated following (3) where $\nabla F(x) = \mathbf{A^T}(\mathbf{Ax} - \mathbf{b})$ and $\alpha$ is the step size.

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \alpha\nabla F(x) \quad (3)$$

The gradient descent update in (3) is run until we visually observe convergence in the log residual. Typically this takes somewhere between 50 and 100 iterations.

### 3.3. Maximum-a-Posteriori Formulation

Another approach that can be used to solve for $\mathbf{x}$ is the maximum-a-posteriori formulation, which involves an iterative minimization of a convex objective function, $J(x)$. This is the approach used in [8]. The gradient $\nabla J(x)$ is computed as

$$\nabla J(x) = \sum_k \mathbf{W}_k^T\mathbf{H}^T\mathbf{D}^T(\mathbf{DHW}_k\mathbf{x} - \mathbf{y}_k) + \mathbf{G}(\mathbf{x}) \quad (4)$$

where $\mathbf{G}(\mathbf{x})$ is a statistical prior. Common statistical priors include total variation (TV), auto-regressive models, and various Gaussian mixture models. As mentioned earlier, [8] employs a locally adaptive prior described by (5) where $\mathbf{R}(\mathbf{x})$ is a bilateral filter.

$$\mathbf{G}(\mathbf{x}) = \|\mathbf{x} - \mathbf{R}(\mathbf{x})\|_2^2 \quad (5)$$

Although not described here, other optimization methods, such as ADMM and Bayesian frameworks may also be utilized to solve the super resolution problem for light field images.

### 3.4. Dataset

Rather than utilizing images captured from an actual light field camera, such as the Lytro Illum or the Raytrix R8, we instead use a simulated light field dataset. Using a simulated dataset allows us to focus on implementing and enhancing the super resolution algorithms, without dealing with platform specific peculiarities. The dataset we use is the 4D Light Field Benchmark dataset from Heidelberg University and the University of Konstanz [9]. An additional benefit provided by this dataset is that it comes with ground truth depth and disparity maps. We use the ground truth disparity map directly in computing the shift matrix, $\mathbf{W}_k$. Each simulated light field contains 9×9 full color sub-aperture images, each of size 512×512 pixels. The provided disparity maps are of the center view and measure 512×512 pixels in size.

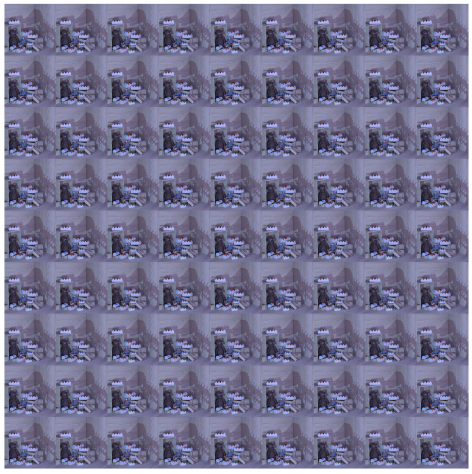


Figure 4. Simulated light field image arranged into sub-aperture views.

## 4. Results

In this section, we present results for super resolution of light field images. In particular, the results presented here use the dataset referenced in Section 3.4 and the gradient descent algorithm described in Section 3.2. Note that due to time constraints, a simplified image formation model is used, in which there is no decimation matrix and in which the noise is assumed to be 0.

An example of the simulated light field is shown in Fig. 4, arranged so that each sub-aperture view can be seen. As noted previously, the simulated light field has 9×9 sub-images, each representing a low resolution sub-aperture microlens image in our image formation model. The disparity map corresponding to the center view is shown in Fig. 5. In this example, the disparity ranges just under ±2 pixels. Other simulated light fields in the dataset have disparities ranging from ±1-3 pixels.

The blur operator $\mathbf{H}$, is constructed using 3×3 Gaussian blur kernel, with standard deviation $\sigma_{\text{blur}}$ and applied to the image in the Fourier domain. Figure 6(a) shows the reconstructed image for $\sigma_{\text{blur}}$ = 0.1. The difference between the reconstruction and target image is shown in Fig. 6(b). This reconstruction has a PSNR of ∼21 dB after 75 gradient descent steps. The PSNR and log of the residual are shown in Fig. 7 as a function of iteration and exhibit the expected trends.

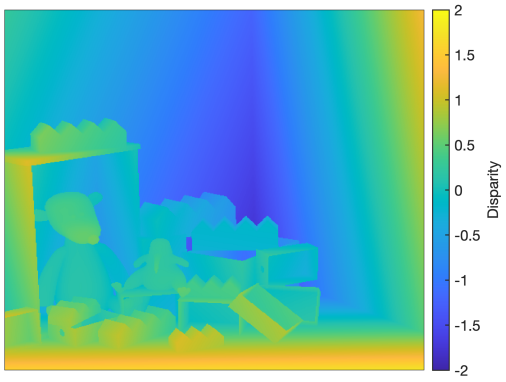



Figure 5. Disparity map for the central view of the light field in Fig. 4.

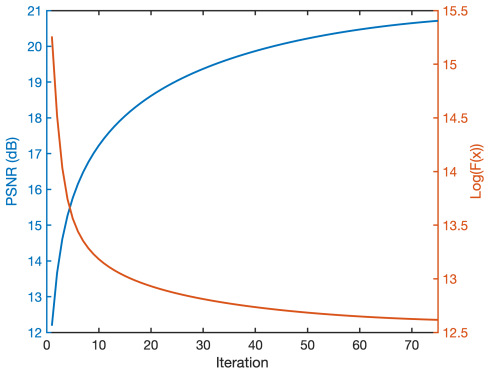



Figure 7. PSNR (blue) and log residual (red) as a function of the gradient descent iteration.

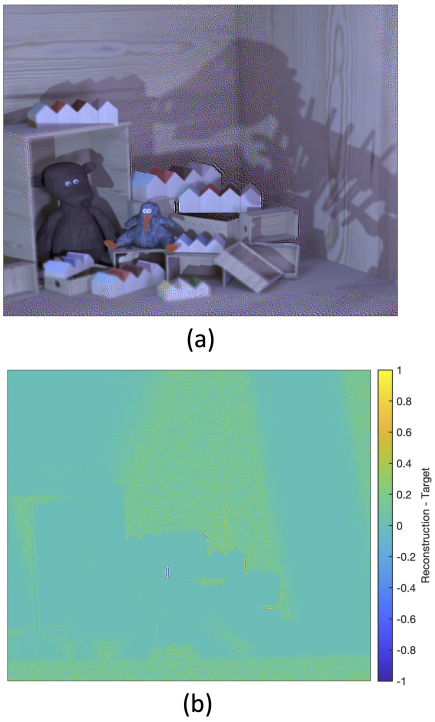



Figure 6. (a) Reconstructed image of children's toys. (b) Difference between reconstructed and target image.

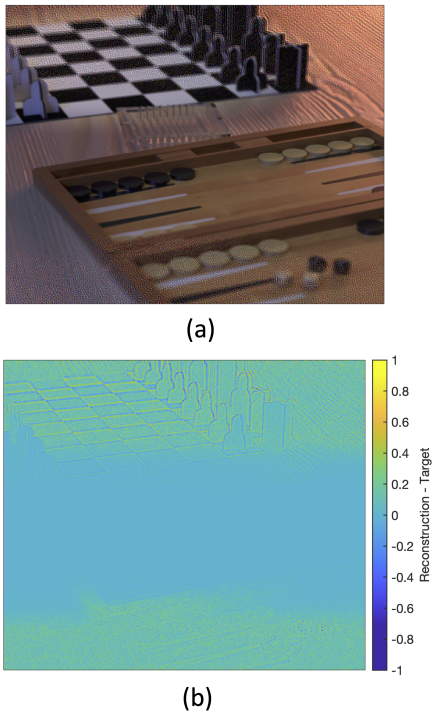



Figure 8. a) Reconstructed image of board games. (b) Difference between reconstructed and target image.

Another example of a super resolved light field is shown in Fig. 8(a). This time the scene is a several board games from much closer perspective than the scene with children's toys. After 75 gradient descent steps, with a $\sigma_{\text{blur}}$ value of 0.1, the final PSNR was $\sim$19 dB. As we see from the difference image in Fig. 8(b), the best reconstruction occurs in the middle horizontal third of the image. It is interesting to consider why this is the case. One potential reason is that this is the part of the scene where the least occlusion occurs. Perhaps when shifting perspectives from left to right along the center line, the disparity map-based warping operator is capturing a true representation of each sub-image, whereas in the upper third of the image where the chess pieces are located it is possible that the warping operation does not accurately generate the sub-aperture images due to object occlusion that changes when adjusting viewpoints. One of the limitations of the current algorithm is that the disparity maps used do not represent these occlusion artifacts. In future work, it will be important to further explore and understand this behavior.

## 5. Future Work and Conclusion

This project was primarily limited by time constraints and there are many interesting directions that future work could expand upon. Several of the potential future work opportunities are described below, though this is not an exclusive list.

One clear direction for future work is to implement the gradient descent method using the full image formation model described by (1). That is, to include the sampling operator $\mathbf{D}$, as well as the Gaussian noise $\mathbf{n}_k$. Fully implementing this model would be an important step towards having confidence in the algorithm's performance in more realistic scenarios. Along these lines, the next logical step would then be to implement the gradient descent super resolution algorithm on real light field images, as opposed to simulated data. Many light field datasets available online including the Stanford Lytro Light Field Archive [10], the matching Lytro Illum and Raytrix R29 datasets from [11], and several others.

Another interesting direction for future research in this field is to apply different super resolution methods to light field images and determine which method performs the best and/or can be applied most universally. This paper made note of the maximum-a-posteriori algorithm in Sec. 3.3, which is another optimization method that has previously been applied to super resolution of light field images. Other potential methods include ADMM, maximum likelihood, and various Bayesian frameworks, such as the one described in [3]. In conjunction with exploring different methods, developing faster super resolution algorithms that can be run in real or near real-time would also be an attractive area for future research. Real-time super resolution would likely greatly increase the marketability of light field cameras in the consumer domain and could potentially enable new scientific imaging applications.

As discussed near the beginning of this paper, there are key differences in the hardware designs of plenoptic 1.0 and plenoptic 2.0 cameras. Accordingly, it would be useful to analyze and compare the efficacy of different super resolution techniques on plenoptic 1.0 versus plenoptic 2.0 cameras. It is possible that due to differences in the image formation, certain algorithms may be better suited to one set of hardware than the other. It would also be interesting to explore whether the type of light field camera or the content of the scene (depth, amount and location of occluded objects, etc.) has more of an effect on the success of a given super resolution method.

Overall, this paper has aimed to introduce the principles of light field cameras and address the angulo-spatial resolution tradeoff that is inherent to the images they produce. By utilizing super resolution on the sub-aperture light field images, one can increase the spatial resolution of a given light field view by $\sim$2-3 times. While not extremely high, this improvement may open the door for greater usage and deployment of light field cameras. Super resolution techniques such as gradient descent may be use, however further research is needed to evaluate all potential super resolution techniques and determine which, if any, is most appropriate for light field images, and under what conditions.

## Acknowledgements

The author would like to thank Professor Wetzstein for his mentorship and assistance with this project.

## References

[1] G. Lippmann, "Epreuves reversibles photographies integrals," *Comptes-Rendus Academie des Sciences*, vol. 146, pp. 446–451, 1908.

[2] M. Levoy, "Light Fields and Computational Imaging," *IEEE Computer Society*, pp. 46–55, 2006.

[3] T. E. Bishop, S. Zanetti, and P. Favaro, "Light Field Superresolution," *2009 IEEE Int. Conf. on Computational Photography*, pp. 1–9, 2009.

[4] R. Ng, "Digital Light Field Photography," Ph.D. dissertation, Stanford University, 2006.

[5] T. Georgiev and A. Lumsdaine, "Superresolution with Plenoptic Camera 2.0," Adobe Technical Report, Tech. Rep., 2009.

[6] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," *2009 IEEE Int. Conf. on Computational Photography*, pp. 1–8, 2009.

[7] T. Georgiev, G. Chunev, and A. Lumsdaine, "Superresolution with the focused plenoptic camera," *Proc. Vol. 7873, Computational Imaging IX*, 2011.

[8] K. Venkataraman *et al.*, "PiCam : an ultra-thin high performance monolithic camera array," *ACM Trans. on Graphics*, vol. 32, no. 6, p. 166, 2013.

[9] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldluecke, "A Dataset and Evaluation Methodology for Depth Estimation on 4D Light Fields," *Computer Vision, ACCV 2016*, pp. 19–34, 2017.

[10] A. S. Raj, M. Lowney, R. Shah, and G. Wetzstein, "Stanford Lytro Light Field Archive," "http://lightfields.stanford.edu/LF2016.html".

[11] W. Ahmad, L. Palmieri, R. Koch, and M. Sjostrom, "Matching light field datasets from plenoptic cameras 1.0 and 2.0," *2018 3DTV Conference*, pp. 1–4, 2018.